Musical Toys: Interactive Audio for Non-musicians.

Oded BEN-TAL
Lecturer Multimedia Technology and Design, School of Engineering and Design, Brunel University
Kingston Lane
Uxbridge, UK, UB8 3PH
Oded.Bental@brunel.ac.uk

Abstract

This paper describes the design and implementation of interactive sound installations using open source software. Aimed at encouraging an audience of non-musicians to explore electronic sounds, the features identified as the key elements in their design were robustness and responsiveness: an interface that is reliable, intuitive and easy to learn, and yields rich and interesting sounds. The main method of achieving this aim was using a single sensor as an input device and mapping the input data in multiple, concurrent ways to sound generation and manipulation parameters.

Keywords

Interaction, installations, multi-media, pure-data.

1 Introduction

In 2007 the author was awarded a grant from the Leverhulme Trust to be an artist in residence at the School of Engineering of Brunel University for one year. The Leverhulme Trust's Artist in Residence programme is designed to enable the development of creative collaborations by bringing artists into higher education institutions where creative arts is not already part of the curriculum, such as an engineering school.

The main focus of this artistic residency was to develop interactive sound installations in the common areas of the school buildings. Working with students and staff at the school, utilising their expertise and the resources available at the school, these installations would, on the one hand, communicate some of the school's research in innovative and entertaining ways and, on the other hand, stimulate members of the school community to explore new ideas and explore sound in ways most have no opportunity to try.

The author is a classically trained composer working in both the acoustic and electronic realms, with additional interests in research and the interaction between research activities and creative outputs. For example, previous work included a research/composition project on data sonification and the use of methods developed for sonification as a basis for composition[1].

The initial phase of the residency involved an investigation of the means and methods for developing interactive installations that will contribute to the school environment. This involved developing collaborative relationship with staff members to learn about current research activities in the school as well as learning to use new hardware and software. The first installation was deployed in November of 2007 and 3 additional ones are planned during this academic year. In addition, as a result of this residency, the author became involved in an interactive multimedia performance integrating dance improvisation, video and animation, fashion design, and electronic music and experimenting with wireless sensors to allow the dancers to interact with their digital environment. The piece was premiered in December 2007 at the Laban Centre in London.

Both the installations and the music for the aforementioned performance were created using open source tools mainly pure-data(Pd)[2] and common lisp music(clm)[3]. This paper will describe the development and design process with regard to the aim of the residency and the limitation of the technology and its intended users, the use of open source software in audio-visual context, some lessons learned during the process, and end with plans for further work the author plans to undertake in this field.

2 Environmental constraints

This section will consider the context of the interactive audio installations produced during the period of residency – in terms of the technology available, the physical placement of the installations, and the anticipated user for which the systems were designed.

2.1 Human factors

Most student and staff at the school have little or no musical training and their normal involvement with music is, like most people in our society, fairly passive: selecting the type of music they listen to, how loud to play it and, possibly, the order of tracks. The aim of the installations made available to members of the school community is two fold: (1) offer them an opportunity to listen to different sounds and (2) introduce them to new ways of interacting with sound.

The reasonable assumption was that most users would interact with the installations for 3-10 minutes, as they pause on their way to/from classes or meetings. It was also clear that most users would not stop to read instructions about operating the system. Thus it was important to design an interface that would be easy to learn in a very short time through actually using it, but at the same time the resulting sound should be rich and complex enough to be interesting and engaging. It became clear that following a musical instrument paradigm for these installation would not work – musical instruments allow for very precise and nuanced control of audio production but the price is long training period needed to master the instrument.

Instead these installations would need a simple input mechanism coupled with more complex mappings from input to audio parameters. This would mean sacrificing detailed, nuanced control over sound production in order to achieve a system that generates rich and complex sounds from a relatively impoverished input. It is, of course, not difficult to generate complex sounds from any input, but an essential aspect of interactive systems is that the user can explore the connections between the input and output and learn to manipulate the output by changing the input. In other words the key concept is to design a system that is *responsive*.

2.2 Technology

The premise of his residency was to use technology available at the school. Several copies of the iCubeX sensor kit from Infusion Systems[4] were purchased by the school for teaching purposes, the availability of these made this sensor kit an obvious option for the interactivity of the proposed installations. Each kit included a selection of basic sensors which would be connected (up to 8 in parallel) to a digitiser converting the input into MIDI messages. An editor available for download, in either windows or MacOS versions, allows the user to program the digitiser according to the sensors connected as well as adjusting sensitivity of measurements and type of MIDI information sent out. At the school these sensors were successfully used under MacOS but not Windows or Linux. This, coupled with the availability of MacPro (intel based) computers, one of which would need to be borrowed to run an installation, led to the choice of Mac as the basis for the project (though some of the development work was done on Fedora Core Linux).

Pd was the natural software choice for implementing the installations, using Gem for the visual side. Over the last 5 years the author has been using Pd for interactive electroacoustic compositions, focusing mainly on ways of using Pd to extract information from the performer and generate a musically appropriate (as defined by the composer) response. This interest in interaction through the use of decision-producing systems proved a useful basis for the projects undertaken during this residency.

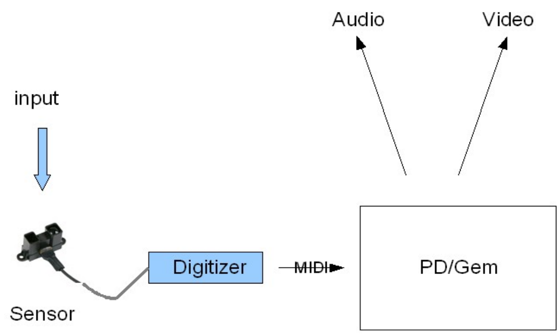

Illustration 1: basic set up

2.3 Physical environment

Aiming to engage students and staff outside of structured activities, these installations would be placed in common areas in the school buildings: an entrance hall, a staircase, a corridor, or a foyer. The need to display them in accessible areas would need to be balanced with the need to secure the electronic equipment from possible theft or damage. The solution devised, in consultation with the technical staff, was to construct a display box locking the computer (plus some of the additional components) inside with the sensors, speakers, and video monitor placed on top (and the whole thing secured with cables to a wall to prevent someone taking the entire box).

The plan was to deploy individual installations for a period of 2-3 weeks each, in different locations around the school. These locations would need to be visible and accessible without becoming an obstruction, not too crowded and noisy preventing a user from engaging with the system, and be located close enough to a power outlet (the last being much less available outside offices than expected).

Illustration 2: picture of first installation

3 Design

Considering the aims and the constraints discussed above, the design of the installations would need to display two key features: robustness and responsiveness. Since the installations would run for 2-3 weeks at a time, they would have to work properly whenever someone tries to interact with them, without the constant need to monitor them or restart either Pd or the computer itself. Thus the system (both hardware and software components) would need to be designed and tested to run unsupervised. The idea of responsiveness means that a user needs to feel in direct control of the system's audio(visual) output. This includes enough predictability in the system, low latency in the response, and an intuitively understood relationship between the controlling action and the response.

A responsive system, in this sense, has more in common with a toy than with a performance instrument. After all most interactive systems developed as a performance tool are optimized for the use of a specific person (usually the improvising musician/programmer) who knows the system well, allowing that person the maximum control over musical parameters needed for the performance. A musical toy, on the other hand, is designed to allow any user some degree of control after a relatively short learning period.

3.1 Sensors

After a period of experimenting with the various sensors available in the basic IcubeX kit, it became clear that the infra-red proximity/distance sensor (*ReachClose*[1]) provided the optimum input: the distance measurements were stable and uniform over a large range (about 10-170 cm.) allowing for a comparatively, rich data stream to be used as input. By comparison the *Touch* sensor did not provide a wide range of readings and is almost an on/off switch, the *taptile* (activated by foot pressure) provided a range of readings but less nuanced manipulation than hand gesture detection with the distance sensor.

The multimedia performance collaboration that the author became involved with during this residency also used Infusion Systems' sensors. These were integrated into the dancer's garments (designed by Michele Danjoux) and connected to the computer through a wireless bluetooth transmitter, creating 'intelligent clothing' through which the dancers interact with their digital environment. Both the *Gforce-2D* (accelerometer) and *flash* (light) sensors were used. While it was possible to get a good range of reading from both sensors the whole system proved very difficult to work with: the wireless connection was not reliable, bringing all components of the system up (sensor, bluemidi, IcubeX, and Pd) and talking with each other took a long time and often did not work at all.

As the work is still ongoing, the author is still experimenting with using various sensor configurations but the initial three installations already configured use the distance sensor as the vehicle for interaction through hand movements. It was a conscious choice to focus on interaction through one input device. Adding more sensors either in parallel (where each sensor affects a different aspect of sound) or in interaction (where one sensor's data modifies the behaviour of another) considerably increases complexity making the system's behaviour that much harder to learn in a short time – which is essential for these installations to be successful.

Data from the sensor is converted, by the digitiser, to MIDI messages. The specifics are set through the editor allowing to change sensitivity, thresholds, channels, and type of message. These messages are sent to Pd using the [ctlin] object and mapped to various sound generation and manipulation parameters.

[1] For a description of the different types of sensors: infusionsystems.com/catalog/index.php/cPath/24

3.2 Mappings

Inside the Pd patches used to run the applications the input from the sensor is a stream of numbers 0-127. Using just one input stream, yet aiming to create an installation that is both interesting and responsive meant mapping this input stream in a variety of ways, often in parallel, into sound parameters (and visual manipulators when a visual component is used). The mapping strategies employed fall into several general categories:

- Mapping of full range (linear as well as non-linear e.g. using log or cosine functions).
- dividing the range into subsections each affecting a different parameter.
- calculating change: either range of change or a "time" derivative of input stream.
- trigger points or event detection.

Mapping the full range of the sensor data to a parameter is the simplest and most straightforward mapping. This method was used mostly to map measured distance to the amplitude of the sound. In addition to linear mapping, both log and cosine functions (e.g. mapping the 0-127 range on to half a cycle – zero at both ends and maximum at the middle) were used.

Dividing the full range into sections and using each section to control a different parameter is, essentially, a simulation of multiple, *alternate* sensors in one input device (alternate because only one range can be active at any given moment). For example the lower half of the range could be mapped to modulation frequency, and the upper half to modulation index of an fm-synthesis instrument. The fact that only one parameter can be changed at any given moment makes the learning of the connection between the hand movement and the sonic result simpler, while the range of possibilities available is still larger than removing one of these parameters from the user's control completely.

Another method used was to calculate change, either magnitude of change or taking the derivative of the input sequence, and use this information to control the output. Calculating the variance of the input (over the last n readings) was used to measure magnitude of change. Calculating the derivative was done with the differential pulse code modulation (DPCM[5]) method. Assuming that the sensor readings are at constant rate it is simple to estimate the time derivative of the input signal based on the last 4 numbers.

Trigger points or event detection were used when a binary rather than continuous change was needed. Examples include resetting the installation if no input is detected, or using a magnitude change above a certain threshold to advance slide in a visual sequence.

The decision to use one sensor required using different mappings in parallel in order to create rich and interesting sounds. For example the full input range would be mapped to amplitude, the derivative would increase/decrease the frequency, and the range would be divided into regions each mapped to the bandwidth coefficient of a set of filters. Arriving at the specific combinations deemed best was a trial and error process guided by the basic aim of creating a responsive system: robust and predictable enough to be easy and intuitive to learn while also producing sounds that are interesting enough to engage the user.

3.2.1 Change vis. position

Using change in the input, whether magnitude of change or derivative of input, to manipulate parameters proved the most useful mapping strategy. Measuring change in the hand position is an estimation of the velocity of the movement, when this was mapped to salient sonic parameters the result was a more responsive system. Such a system seemed to exhibit a stronger, and much more intuitive, correlation between the hand gestures and the sounds produced.

The importance of haptic feedback in learning interactive systems is well documented[6], feedback which was absent in these installations. It is possible that internal monitoring of muscle activation sequences, which correlated to speed and direction of movement, provide a better feedback mechanism than absolute position, facilitating the learning of gestures even in the absence of haptic feedback[2].

Another possibility is that an inherent correlation between music and motion is part of the experience of listening to music. Bharucha, Curtis, and Parvoo[9] identify several kinds of internal experiences of motion that are important for the experiential nature of music, and Leman and Camurri[10] discuss the relationship between

---

[2]Providing solid evidence for this conjecture is beyond the scope of this paper or the expertise of the author, and is based entirely on the author's own observations. Nevertheless there is experimental evidence that processing of hand motion is superior to hand position. For example: Vindras et al[7] suggest that distance and direction rather than absolute spatial position give a better account of the way aimed hand movements are planned, while Bevan et al[8] found that for hand joint movement, angular distance is parsed more accurately than the absolute angle.

gesture and musical expressiveness. Thus the perceived correlation between the hand gestures measured by the sensor and their sonic result could be a reflection of the way we experience music as listeners.

4 Suna No Onna

Brunel's School of Performing Arts has a strong and dynamic digital performance group: working with interactive technology in multimedia performance settings. The author was invited by Professor Johannes Birringer (chair in drama and performance technologies at Brunel) to take part in an interactive multimedia performance based on the novel, and film 'Woman in the Dunes'. The work combines dance, interactive video and animation, fashion design, and electronic music created by an ensemble of artists from diverse creative backgrounds. The development of this performance followed a collaborative process where the different elements (movement, 3D animation, video, music, etc.) were developed in parallel with continuous feedback from the group, as did the dramatic structure of the piece which evolved over the course of the project with the film's narrative providing a reference point.

Part of the aim of this project was exploring interaction between the dancers and their digital environment in a performance setting. Wireless sensors were integrated into the design of the clothing to create 'intelligent garments' which would serve as a link between the dancers – their appearance, their movements, their expressiveness – and their suroundings.

Initially Pd provided the basis for the interaction: data from the sensors was used to control the play speed of short videos and 3D animations (created outside Pd) allowing the dancer to slowdown and speedup the motion on the screens, as well as freeze at any moment. This allowed for the emergence of a dialogue between the dancer and the visual projection. Eventually the visual team switched to a different graphic programming environment which they were more familiar with (Isadora[11]). As part of the development of the piece we explored some ways of integrating sensor inputs into the music for the piece. But we abandoned these due to the unreliability of the setup (see section 3.1) and the time limitation of the production.

The music for the performance was done using two open source tools: Pd and cm/clm (common music and common lisp music). Pd patches generating music in real-time are mixed with clm-generated soundfiles triggered in coordination with the performance.

Working as part of group production, and in a rather complex project (from both the artistic and technological points of view), poses some serious challenges to the participants. In one of these, namely developing composition strategies in collaborative settings, Pd proved a useful tool.

4.1 Composition strategies.

Composing music for an hour-long performance is in itself a new experience for the author. The improvised, collaborative nature of the creation process added to the challenge for this author (for whom the composition process often still involves pencil and paper). As the various elements of the performance are developed (set design, video, dance as well as scenes, transitions, and dramatic path) the individual contributors need to balance the artistic (and technical) needs of their particular medium with those of other participants and the (often unknown) context of the specific item currently developed.

From a musical perspective this process creates multiple challenges:

- creating musical section whose eventual duration would be determined (how?) at some future point.
- Writing music that is both interesting in itself, balances with the attentional demands of additional media on the viewer (without necessarily knowing these other components), and integrate into the expressive purpose of the moment
- Composing music with limited knowledge of the context: when will this section appear in the final product? What comes before? After?

Addressing these issues necessitated adapting previous working methods and developing different ones as well as learning new composition strategies.

One method used was the concept of sonic sketching of material. The author's composition process usually involves sketching out ideas and outlines at the initial stages. Creating sonic sketches marking the basic music material, without the long process of fine tuning and refining, allowed for a relatively rapid and efficient mechanism to try various ideas with the rest of the group and pick the ones that work well for the piece as a whole.

Another method adapted from previous work is creating music with 'spaces' in it. The idea is to compose a layer of music where 'holes' are left to be filled later by other layers. This idea was used by the author in some compositions combining instruments with electronic components – one of these would be created before the other (at least for a section) leaving space for the other, as yet undetermined, layer. Similarly the music composed for Suna tries to leave room for additional components created by others in this instance.

Composing music of unknown length was the most difficult initial obstacle, because of the author's musical thinking and methods of composing. One solution was to use Pd patches that generate a musical texture. These textures were created using quasi-random processes in both the timing of events and the specific parameters of each event (within carefully constraint limits). Another solution was to create an algorithmic melody generator in Pd. A set of Markov-like processes is used to generate synthesis parameters: frequency, amplitude, duration, vibrato, etc. In both cases, random textures and melody generator, an optional end condition was programmed which meant that the sound could either continue indefinitely, run until stopped, or terminate automatically when certain conditions are met.

Having created independent sound modules the rehearsal process involved mixing these with each other and the various visual components (themselves evolving over the period) to try to arrive together at the best possible result. However, the static nature of the sounds produced by each module made transitions, musical transformations, and dynamic processes difficult to achieve. Controlling patches, allowing gradual change of parameters such as amplitude or density, made it possible to generate crescendi, accellerandi etc, but generating a sufficiently convincing 3-minute section where musical ideas are developed proved too difficult to create in Pd.

The combination of Common Music and Common Lisp Music (cm/clm) provides an integrated music composition environment allowing detailed control of musical processes through the algorithmic generation of synthesis parameters. Thus much of the music used for the performance is in the form of fixed soundfiles, created with cm/clm, and only triggered within Pd, while Pd modules, generating music of indetermined length, provide the full music in only short sections of the piece and is mixed with additional material in other sections. Such modules also serve as a linking material between some of the section – providing the flexibility needed for a largely improvised performance.

5 Conclusion

Working, as an artist, within an engineering school provided a fantastic learning opportunity for the author: new dsp techniques, new hardware and software, new modes of work, some of these will, in turn, feed into the author's future work in both composition and research. One such future project currently under development involves trying to utilize various motion tracking methods (sensors, video, motion capture) to map movement to sound synthesis focusing on the notion of creating a *responsive* system – integrating some of the ideas developed for the installation work into a performance system.

As for the use of open source software – all the author's in this project was successfully done using Pd and clm (both open source developments). What clm provided (for the author) that was missing from Pd was the integration of sound synthesis with a higher level of algorithmic control over musical evolution (using common music). Some algorithmic control was implemented in Pd (e.g. melody generating module based on markov-like processes) but this area seems to be underdeveloped in Pd.

The fact that the team abandoned Pd/Gem for the purpose of running the video side of the performance is, probably, the result of difficulties learning Pd while also addressing the technical difficulties with the sensor system, and developing interactive concepts connecting garments, movement, and animation. Learning the visual side of Pd (Gem, PDP) should have been a much higher priority at the early stages of the work in order to facilitate the creative work between the dancers and the digital environment.

6 Acknowledgements

The work described in this paper was supported by the Leverhulme Trust's Artist in Residence programme. The author would also like to thank colleagues at the ECE department at Brunel and especially Professor Malcolm Irving, Dr. Vanja Garaj, and Stephen Cockett for their help with this project.

References

[1] Ben-Tal O. and Berger J. 2004. Creative Aspects of Sonification. *Leonardo* Vol. 37/3 pp. 229-233

[2] http://puredata.info

[3] http://ccrma.stanford.edu/software/clm/

[4] infusionsystems.com

[5] Sayood K. 2000. *Introduction to data compression (Second edition)*, Academic Press, Publisher: Morgan Kauffmann, pp. 316-320

[6] Essl G. and O'Modhrain S. 2006. An enactive approach to the design of new tangible musical instruments. *Organised Sound* Vol. 11/3 pp. 273-284

[7] Vindras et al 1998. Pointing Errors Reflect Biases in the Perception of the Initial Hand Position. *Journal of Neurophysiology* Vol. 79/6, pp. 3290-3294

[8] Bevan et al 1994. Proprioceptive coordination of movement sequences: discrimination of joint angle versus angular distance. *Journal of Neurophysiology*, Vol. 71/5, pp. 1862-1872

[9] Bharucha, J.J. Curtis, M. and Parvoo K. 2006. Varieties of musical experience. *Cognition*, 100, pp. 157-165.

[10] Leman, M. and Camurri A. 2006. Understanding musical expressiveness using an interactive multimedia platform. *Musicae Scientiae*, Special issue on Interdisciplinary Musicology, pp. 211-214

[11] designed by Mark Coniglio, available from www.troikatronix.com/isadora.html