Creative Control of Granular Synthesis Using Fluid Simulation & Motion Tracking

Andrew Bluff & Andrew Johnston
Creativity & Cognition Studios
University of Technology Sydney
PO Box 123
Broadway, NSW, Australia
bluffy@rollerchimp.com / andrew.johnston@uts.edu.au

ABSTRACT

This paper describes the development of an audio-visual performance system which applies 'reality based interaction' techniques. The real-time gestures and sounds of a musician playing an acoustic instrument are tracked and translated into forces which act on a fluid simulation. The simulation is visualised and also sonified using granular synthesis. Several strategies for linking live performance, fluid behaviour and generated sounds and visuals are discussed.

Categories and Subject Descriptors

H.5.5 [**Sound and Music Computing**]: Methodologies and techniques; J.5 [**Arts and Humanities**]: Performing arts

Keywords

Interaction, granular synthesis, audio-visual performance

1. INTRODUCTION

This paper describes an audio-visual performance instrument based on real-time fluid simulation. The aim of the work is to explore how the use of basic motion-tracking techniques can be made more compelling and engaging for performers and audiences through the use of 'reality-based interaction' techniques which support intuitive, embodied interaction.

Motion tracking systems have become increasingly prevalent in creative performance and more broadly in human-computer interaction in recent years. These motion tracking systems open up exciting possibilities for dance and physical theatre and numerous works have been created which explore their use in live performance. Landmark works include David Rokeby's *Very Nervous System* [13], *Glow* (2006) and *Mortal Engine* (2008) by Melbourne-based dance company Chunky Move and numerous works by Troika Ranch (USA) [9] and Palindrome (Germany) which feature sophisticated live motion tracking systems linked to computer systems that produce interactive graphics in real-time.

Much of the second author's prior work has focused on the use of simulated physical models as 'mediators' between live acoustic sounds and computer generated sounds and visuals [4, 7, 6]. In these works, the sounds produced by musicians on their acoustic instruments is mapped to forces which act on a physical model. Using their instruments, performers are able to poke and prod the models which move in response and generate their own sounds.

This basic approach is a kind of 'reality-based interaction' (RBI). RBI is a term proposed by Jacob et al [3] to describe the emergence of an approach to interaction design which aims to take advantage of users' existing understanding of how to interact with the non-digital world. Researchers and artists developing live performance systems have argued that reality-based interfaces based on physical simulations offer a number of advantages. Cadoz et al [2, 1] argue that physical models provide more intuitive, direct control over sound synthesis because they are intuitively understandable but also exhibit rich and complex behaviours which produce engaging and interesting sounds. Momeni and Henry [10] propose the use of physical models as an intermediate mapping layer arguing that they effectively support 'rich and intuitive exploration' of audio-visual synthesis and provide an 'intrinsic link' between sound and visuals. Studies of professional musicians using reality-based interfaces based on mass-spring models have found evidence to support this [4].

Systems such as the ReacTable [8], among many others, have successfully explored the use of tangible objects in creative interfaces. However, the use of actual physical objects in these interfaces is in some senses a limitation, because they lack the 'malleability' of digital objects, which are, "easy to create, modify, replicate and distribute" [11]. The use of physical simulations in the user interface is an attempt to gain the malleability provided by digital objects while retaining as many benefits of physical interaction as possible.

1.1 Past Work

The system we describe here has evolved considerably over several years. Initially, the system was developed for use in the Stalker Theatre dance/physical theatre work *Encoded*, which premiered in late 2012. The *Encoded* system is primarily an interactive *visual* system which is presented with pre-recorded audio in performance. The system itself was based on an interactive real-time fluid simulation. Using

MOCO'14, June 16-17 2014, Paris, France.
ACM 978-1-4503-2814-2/14/06 ...$15.00.
http://dx.doi.org/10.1145/2617995.2618022.

infra-red motion tracking systems, performers directly interact with simulated 'fluids', stirring them and creating sophisticated effects. Because the parameters of the fluid simulation can be manipulated in real-time, the fluid can change viscosity, colour, etc as the performer moves, providing potential for rich creative dialogue to occur.

Encoded toured to South Korea and the Netherlands. The flexibility of the interactive systems led them to be used in a more recent production *Pixel Mountain*, which was presented in Gwacheon and Hi Seoul Festivals in 2013. Highlights of these shows can be seen at `http://vimeo.com/55150853` and `https://vimeo.com/76746676`.

1.2 Incorporating Audio

While the systems developed for *Encoded* and *Pixel Mountain* were effective, we wanted to explore how interaction with the physical simulation could be used to create a system which integrated both audio and visuals. Our initial attempts to bring audio into the system focused on linking the velocity of fluid in sections of the simulation to parameters to band-pass filters which were applied to pre-recorded audio files. Using this technique a wide range of sounds could reliably be produced and affected by the movements of performers. However, the fact that movement in particular regions of the fluid simulation always produced sounds with particular frequencies (eg. high sounds to left of screen, low sounds to the right, etc) quickly became tedious.

We next applied a more sophisticated approach which made use of concatenative synthesis techniques [12]. For this system, a pre-recorded audio file was broken up into short segments, which were analysed to produce a set of perceptual descriptors of their timbres. These audio segments were then grouped together by timbal similarity and associated with sections of the fluid simulation. Because similar sounds are close together, the effect is that neighbouring regions of the fluid generate similar sounds. These systems are described in detail in [5].

2. SYSTEM DESCRIPTION

In order to focus on the development of the audio features of the system, for this stage of development we decided to work with a single musician. To convey a sense of the performer being 'immersed' in an interactive, audio-visual fluid, the musician plays behind a scrim. We use LED stage lights to illuminate the performer so that she is just visible, and the output from the system is front-projected on the scrim. This setup also has the advantage that the performer is able to directly interact with the visuals produced by the system while still facing the audience.

The overall technical structure of the system, primarily written in C++ using OpenFrameworks[1], and Pure Data,[2] can be seen in figure 1. The movements of the performer are captured via a Point Grey FireFly camera[3] fitted with a visible light filter (Lee #87), which blocks visible light while allowing infra-red light to pass through. An infra-red LED is attached to the performer's instrument so that their movements can be tracked independently of stage lights and projections.

[1]`http://www.openframeworks.cc/`
[2]`http://puredata.info/`
[3]`http://www.ptgrey.com/`

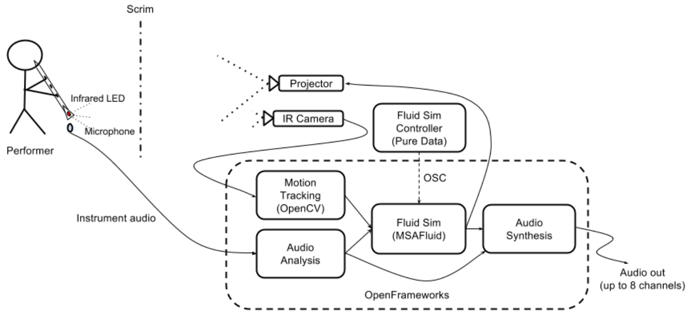

Figure 1: An overview of the system.

The camera images are passed through an OpenCV[4]-based motion tracking system which tracks the movement of the infra-red LED attached to the performer's instrument. The pixel movements identified by the motion tracking system were linked to a real-time fluid simulation, a heavily modified version of `MSAFluid` by Mehmet Atken.[5] The movements of the performer effectively 'stirs' the fluid and may also add particles and/or colour to it. The fluid simulation is extremely flexible and there are numerous parameters which can be adjusted during performance. These include parameters which affect the response of the fluid simulation to performers' movements (viscosity, for example) as well as settings which affect how the fluid is visualised (eg. colour or black and white, particles or lines, etc). All settings are set from a Pure Data patch which sends Open Sound Control (OSC) [14] control messages to the fluid simulation.

2.1 Fluid Manipulation

The physical motion of the performer and the sounds they produce are the primary ways to manipulate the fluid. However, semi-autonomous force 'emitters' can also be added to the simulation. These emitters can be programmed to repel or attract the fluid (and therefore the particles suspended in the fluid). They are moved around the fluid either through mouse interaction, Pure Data parametric control or by attaching them to the performer's movements by way of OpenCV blob tracking.

In recent development, virtual emitters can also be connected to real world microphones, both physically by way of blob tracking an infrared LED that is attached to the microphone or instrument and sonically through the input of sound into the fluid. The volume of sound passed into the real world microphone creates a corresponding force into the fluid and, more importantly, injects sound particles into the fluid. These particles contain a small sound recording (or grain) from the microphone of up to one second in duration. These emitters can be thought of as virtual speakers as they transform the sonic energy from the real world microphone into a physical force and sound which is digitally injected into the virtual fluid system.

Just as there are virtual speakers to inject audible sound from a microphone into the virtual fluid system, there are also virtual microphones which are placed into the fluid and output sound when connected to real world speakers. Fig-

[4]`http://opencv.org/`
[5]`http://www.memo.tv/msafluid/`

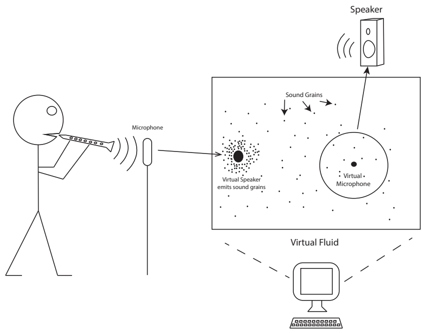

Figure 2: The transformation of audio into virtual sound grains and then back out to speakers via the virtual microphones.

ure 2 shows the signal flow from real world audible sound into the looping sound grains suspended in virtual fluid and back again.

2.2 Audio Synthesis

The audio synthesis technique employed by the work is quite simple, but the multitude of independent configuration options for the virtual microphones, virtual speakers, fluid and sound grains facilitates very complex sonic interactions. Each sound grain is effectively a tiny sampler that records sound from a physical microphone and continually loops this recording. Standard granular synthesis parameters such as playback speed, grain size (loop duration), volume and playback direction are all directly controllable via the Pure Data interface. Any of these granular parameters can also be linked directly to the physical speed of the corresponding sound particle in the fluid. This allows the performer(s) to indirectly control each sound grain individually, simply by moving about the space and stirring the fluid.

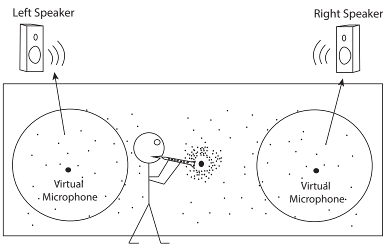

Figure 3: A standard stereo output configuration.

Rather than outputting all of the sound grains simultaneously, the addition of virtual microphones has opened up different interaction modes and spatialisation possibilities. The proximity of each sound grain to the virtual omni-directional microphone appropriately scales the volume of each individual sound grain. The radius and x/y position of each virtual microphone can be controlled dynamically, and each virtual microphone can be directly connected to an audio output channel and physical speaker. One obvious configuration is to place one virtual microphone to the left of the virtual canvas and one to the right of the canvas corresponding to the location of a standard stereo speaker setup (See Figure 3). This provides an intuitive format for the observer where the location of the audible sound grains matches the projected visual location of the corresponding virtual particle. Extending this strategy, any number of speakers may be attached to the system to easily allow multi-channel spatialisation of the virtual fluid.

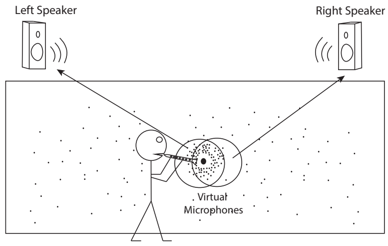

Figure 4: A stereo pair of microphones are attached to the performer's movement with optical blob tracking.

A more engaging mode of sonic interaction was uncovered when we we attached the location of these virtual microphones to the physical location of the performer using optical blob tracking. In this configuration, only the sound grains that are close to the performer's instrument are actually audible and the performer is able to effectively play with any of the sound grains on the large projected canvas. Performers are able to inject their own sonic material into the canvas at a multitude of disparate locations and then replay these unique sonic interactions at a later stage by simply moving to the physical origin of the recording. These localised recordings can also be re-mixed by physically stirring the fluid around to merge the localised grains. By placing two separate virtual microphones a small distance apart but locking them both to the performer with blob tracking, a pleasing stereo spatialisation was created. This tracked stereo configuration (Figure 4) can be both responsive and engaging for the performer while remaining relatively intuitive to the greater audience.

We also have the ability to make each sound particle flock to a location in the fluid based on it's individual sound properties, essentially turning the fluid controller into a giant concatenative synthesis interface. Pitch is spread across the canvas horizontally and volume is arranged vertically. Although pitch is commonly associated with verticality, the horizontal layout was easier to physically traverse for the performer and the result is somewhat analogous to a giant virtual piano keyboard.

2.3 Visuals

An informative and intuitive visual feedback can greatly reduce the learning curve of a new interface for musical expression and facilitate virtuosic mastery by advanced users. [8]

As this work was initially designed to visually augment physical movement in *Encoded*, the system already contained a rich visual palette. The efforts have subsequently been focused on creating a solid and intuitive link between the sonic granular synthesis and the existing movement visualisations. In our initial attempt at converting the piece into a fluid controlled musical interface, the motion of the fluid itself was sonified whereas the particles floating in the fluid create the visuals. This slight dislocation between the representation and mechanics of sound and vision created a confusing disconnect between the two modalities. In this iteration we have both sonified and visualised the floating particles directly which has resulted in a significantly more intuitive interface.

Each individual sound grain is colour coded according to simple sonic properties of volume and pitch. This adds further depth to the projected visuals and also provides the performer and audience with useful information as to which section of the fluid contains which types of sounds.

3. DISCUSSION & CONCLUSION

In this paper we have described the development and refinement of a creative audio-visual synthesis system in which live audio, computer generated visuals and synthesised sounds are unified through the metaphor of fluid.

In this approach, live acoustic sounds and physical gestures are both treated as 'forces' which are applied to fluid. All sounds and gestures made by the performer affect the fluid, and all sounds and visuals produced by the computer are modulated by the movement of the fluid. The fluid therefore acts as a complex mediator, which, in some ways has 'a life of its own', but is not a disembodied autonomous abstract agent, which lacks a connection with our physical, embodied experience. We can see that the behaviour of the fluid is a direct response to our gestures and sounds, but we could not necessarily have predicted its precise response. This, we argue, is an effective way to enable creative engagement in performance.

Putting simulated physical systems at the core of the interaction has interesting effects on how the software is developed. Instead of combining different widgets into a 'user interface' which we arrange to maximise ease of use and efficiency, we are instead 'tinkering in the garage', creating a strange kind of pseudo-physical object which which has visual, sonic and behavioural characteristics which are engaging for performers and compelling for audiences. That is, we are not constructing an interface to create sounds and images, but instead constructing a kind of sculpture which has compelling sonic and visual properties.

This is an area we are motivated to further explore. The next step for us will involve bringing the two halves of our work together, including both dancers and acoustic musicians in live performances where both sounds and movements are visualised and sonified via interaction with fluid simulation.

4. ACKNOWLEDGMENTS

This work was funded by a UTS Faculty of Engineering and Information Technology Industry and Innovation Grant. Our thanks to Linda Walsh who premiered the work described here. Thanks also to Lukasz Karluk for his work on the optical-flow tracking and to the developers of OpenFrameworks and Pure Data.

5. REFERENCES

[1] C. Cadoz. The physical model as metaphor for musical creation: "pico..TERA", a piece entirely generated by physical model. In *Proceedings of the 2002 International Computer Music Conference*, pages 305–312, 2002.

[2] C. Cadoz, A. Luciani, and J. L. Florens. Responsive input devices and sound synthesis by simulation of instrumental mechanisms: The CORDIS system. *Computer Music Journal*, 8(3):60–73, 1984.

[3] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 201–210, New York, NY, USA, 2008. ACM.

[4] A. Johnston. *Interfaces for Musical Expression Based on Simulated Physical Models*. PhD thesis, University of Technology Sydney, 2009.

[5] A. Johnston. Fluid simulation as full body audio-visual instrument. In *Proceedings of New Interfaces for Musical Expression (NIME)*, pages 132–135, 2013.

[6] A. Johnston, L. Candy, and E. Edmonds. Designing and evaluating virtual musical instruments: facilitating conversational user interaction. *Design Studies*, 29(6):556–571, 2008.

[7] A. Johnston, L. Candy, and E. Edmonds. Designing for conversational interaction. In *Proceedings of New Interfaces for Musical Expression (NIME)*, 2009.

[8] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, pages 139–146, New York, NY, USA, 2007. ACM.

[9] L. S. Kepner. Dance and digital media: Troika Ranch and the art of technology. *Digital Creativity*, 8(1):11–19, 1997.

[10] A. Momeni and C. Henry. Dynamic independent mapping layers for concurrent control of audio and video synthesis. *Computer Music Journal*, 30(1):49–66, 2006.

[11] I. Poupyrev, T. Nashida, and M. Okabe. Actuation and tangible user interfaces: the vaucanson duck, robots, and shape displays. In B. Ullmer and A. Schmidt, editors, *Tangible and Embedded Interaction*, pages 205–212. ACM, 2007.

[12] D. Schwarz. A system for data-driven concatenative sound synthesis. In *Digital Audio Effects (DAFx)*, pages 97–102, Verona, Italy, December 2000.

[13] T. Winkler. Creating interactive dance with the very nervous system. In *Proceedings of Connecticut College Symposium on Arts and Technology*, 1997.

[14] M. Wright and A. Freed. Open sound control: A new protocol for communicating with sound synthesizers. In *International Computer Music Conference*, pages 101–104, Thessaloniki, Hellas, 1997. International Computer Music Association.