GRAINSTICK: A COLLABORATIVE, INTERACTIVE SOUND INSTALLATION

Grace Leslie
University of California, San Diego
La Jolla, CA, USA

Diemo Schwarz, Olivier Warusfel, Frederic Bevilacqua, Bruno Zamborlin, Pierre Jodlowski, Norbert Schnell
IRCAM–CNRS
Paris, France

ABSTRACT

The interactive installation GrainStick was developed as a collaboration between the composer Pierre Jodlowski and the European project Sound And Music For Everyone Everyday Everywhere Everyway (SAME) at Ircam, Paris. The interaction design of GrainStick presents a new development in multimodal interfaces and multichannel sound by allowing users control of their auditory scene through gesture analysis performed on infrared camera motion tracking and accelerometer data. Spatial rendering was performed on a wavefield synthesis system; we discuss the advantages held by this technology in collaborative and interactive applications of this kind.

1. INTRODUCTION

The goal of the European Project SAME is the development of context-aware technologies that bring interaction and collaboration to the experience of audio media. The interactive sound installation GrainStick is a recent contribution by Ircam that is intended as an example of how wavefield synthesis (WFS) and human-computer interaction technologies can be applied to the creation of a musical experience by an expert composer. Past interactive productions at Ircam have incorporated spatial audio technology, notably with binaural scene rendering within the framework of the LISTEN project [7], and WFS in the DADA exhibition at the Centre Pompidou art museum in Paris. GrainStick extends the previous use of WFS by incorporating scene rendering which is collaboratively controlled by multiple users' gestures.

In the GrainStick installation, one or two participants balance a virtual tube in which grains of sound can be made rolling to one end or the other, or shaken up. Depending on the inclination of the tube (given by the two controllers), the grains roll faster or slower, and stretching the distance influences their sound. Additional surrounding soundscapes can be controlled by the tilt of the controllers, and percussive movements trigger incidental sounds, or change the sound of the grains, if they are synchronised.

2. THE SETUP

The GrainStick installation is comprised of three main components, each of which is a combination of hardware and custom software: motion tracking and gesture recognition, sound synthesis, and audio scene rendering. A combination of hardware is used for motion tracking: a six-camera ARtrack motion tracking camera system[1] records the absolute position of a user's marker in space (presenting a specific spatial pattern of reflective balls), which is attached to a Nintendo Wiimote controller that records the acceleration of the user's gestures in each of three dimensions. For the purpose of the SAME project, the initial setup used the Nokia N95 smartphone, but for the public performance Wiimotes were used for robustness and availability. The WFS rendering is performed on a SonicEmotion system[2], comprised of multiple networked computers, using custom filters designed at Ircam. An additional 5 loudspeakers are placed to the sides and back of the room to give an adequate surround impression. The hardware setup and software architecture are depicted in figures 1 and 2, respectively.

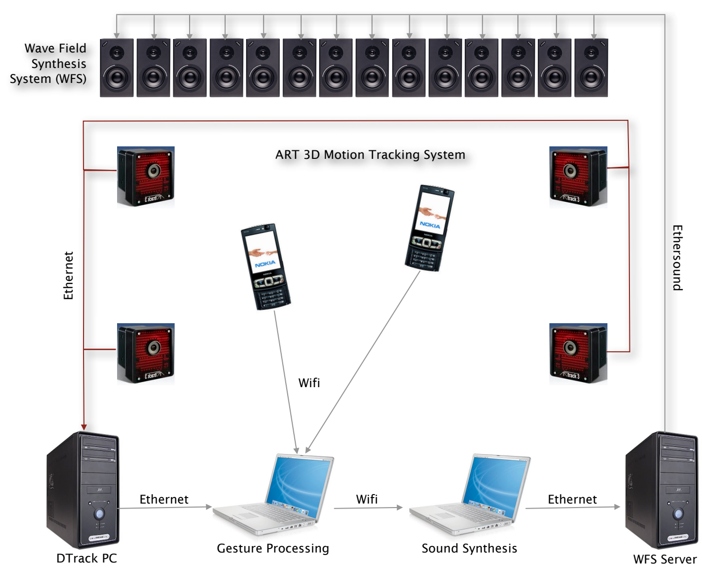

Figure 1. Diagram of the installation hardware setup with the communication protocols between components.

Grainstick uses three stages of software designed in MAX/MSP, using the FTM&CO. extension library [5] [3] for its real-time optimised data structures and powerful sound,

[1] http://www.ar-tracking.de
[2] http://www.sonicemotion.com
[3] http://ftm.ircam.fr

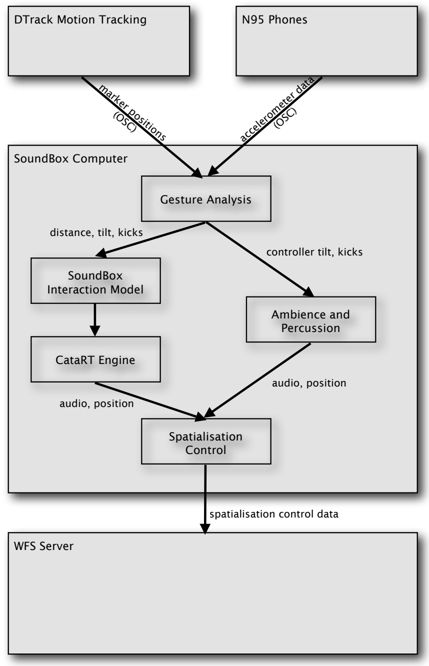

Figure 2. Software architecture dataflow.

gesture, and matrix processing tools [1]. The inter-module communication and the mapping from gesture data to sound parameters is realised using JAMOMA[4] [4].

2.1. Gesture Analysis

The gesture analysis portion uses the raw accelerometer data from the Nintendo Wiimotes to derive a smoothed tilt value.

It also identifies percussive "kick" gestures in any of the 6 possible directions, as shown in figure 3. In order to achieve this, the 3D acceleration data are split into 6 axes considering both positive and negative accelerations. Typically, a "kick" gesture appears as a positive peak. As it is generally directly followed by a negative peak (or the inverse), a "short" gate (around 100 ms) has been set to avoid the second "twin" peak. As a kick in a given axis also provokes small peaks in the others axes appearing with a longer delay, a "long" gate (around 400 ms) was set to avoid these "crosstalk" peaks.

The motion tracking data is used to derive the inclination and length of the line connecting the two controllers (their distance), and their azimuth (the angular direction they are pointing to).

2.2. Sound Synthesis

The installation is composed of four sound elements offering various levels of control and interaction to a participant:

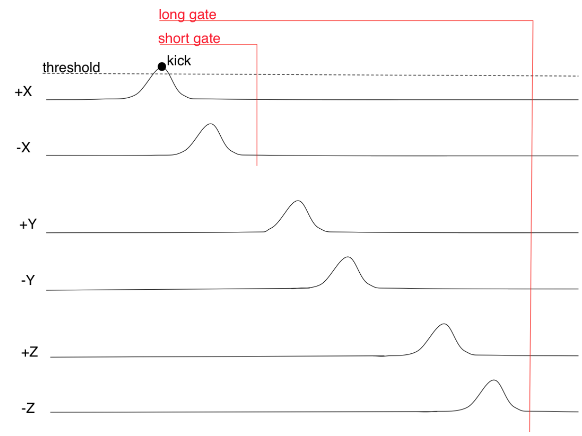

Figure 3. Kick detection.

Percussive Sounds When an individual controller is shaken downwards, different randomly chosen percussive sounds are triggered and rendered as focal sources on the WFS system at a location corresponding to the angle at which the user had pointed the controller, as if the percussive instruments follow the user's hand.

Articulated Ambience Sounds A layer of minimalistic incidental sound events create an animated but subliminal atmosphere. The volume of this layer of ambience sounds is controlled by the distance between the controllers.

Ambient Texture Layer An ambient sound texture, developed using the SOGS smooth overlap–add granular synthesis module[5] were spatialized along an eight-channel array consisting of three virtual loudspeakers rendered using the WFS system and five loudspeakers positioned at the sides and back of the room.

Ambiences are chosen randomly at every three grain sound set changes (see below). Their amplitude and frequency modulation are controlled by the individual Wiimotes' tilt, and the frequency of a peaknotch filter is controlled by the distance between the controllers.

The Grainstick The control metaphor is related to the Rainstick percussion instrument, a closed wooden tube with rice or other grains inside, that emit a rain-like sound when they are made to roll down to one side, with density and amplitude depending on the tilt of the tube. This metaphor is transferred to a virtual tube that is determined by the two controllers, i.e. the relative height of the controllers determine the tilt of the virtual tube. When the controllers are held at equal height, the grains are still and no sound is heard. If the tilt is increased, the grains start moving towards the downward side and emit their sounds. If the tilt is very steep, a dense avalanche of grains happens. When all grains

[4] http://jamoma.org

[5] http://forumnet.ircam.fr/708.html

are on one end of the tube, the participant has to inverse the tilt in order to make the grains move again towards the other end. These grains are rendered as directional, focal sources on the WFS system. The rendered grains travel along the full length of the WFS array as they tumble from one end to the other in the virtual instrument, integrating synthesis control with spatialization control.

When both controllers are shaken upwards, all grains "fly up" and sound all together. When both controllers are shaken downwards synchronously, the grains are exchanged for a different set, like water drops, coins, wood pieces, slabs of clay, or more unusual sound grains like guitar, voice, trombone, bleeps. Additionally to the tilt, the distance between the two controllers also controls parameters of the sound, like their pitch and randomness of pitch, according to different presets chosen at random with the sound change.

2.2.1. Realisation of the Rainstick Metaphor

The rainstick metaphor is realized using the CATART[6] engine [6] for real-time interactive corpus-based concatenative synthesis. This approach makes it possible to navigate through a two- or more-dimensional projection of the descriptor space of a sound corpus in real-time, effectively extending granular synthesis by content-based direct access to specific sound characteristics.

We use this concept here to organise the about 40 grains per sound set by their "intensity" along the length of the virtual rainstick. This means that different grains sound at different fixed positions along the rainstick. This departure from the physical metaphor aids the orientation of the participant and assures sonic variety.

Triggering of the grains is controlled by a play position along the grain axis. When the play position passes over a grain, it is played at its position in space. The rainstick inclination controls the speed and direction of the play position movement. From a certain threshold of steepness onwards, the density of grains is augmented by triggering more and more neighboring grains at a faster rate, leading to an avalanche of grains.

The overall effect is a natural-feeling and sounding recreation of small grains spilling from one side of a container to another. For small inclinations, the participant's control is very fine and grains can be triggered one by one. Increasing the inclination augments the rate and density of grains leading to a large dynamic range in the resulting sound.

2.3. Interaction Design

The GrainStick design is flexible in its interaction potential in that it allows individual or multiple people to interact with the system. One user can choose to use both controllers to hit percussive instruments, or to tip the virtual instrument, spreading sound grains and percussive musical sounds across the rendered sound field. Otherwise, two users can each hold a controller, in which case they must consciously collaborate in the control of the virtual instrument, e.g. to control the inclination of the Grainstick, and must synchronise their shaking movements in order to change sound sets, otherwise it will not sound.

The four sound elements provide different levels of interactivity and sense of control, ranging from the very clear percussion gesture, over the easily discoverable one-to-one mapping of tilt to ambient texture modulation, over the one-to-many mapping of distance to several sound parameters, to the more complex rainstick interaction, that invites the participants to explore and discover the interactions with the sonic environment.

[6] http://imtr.ircam.fr

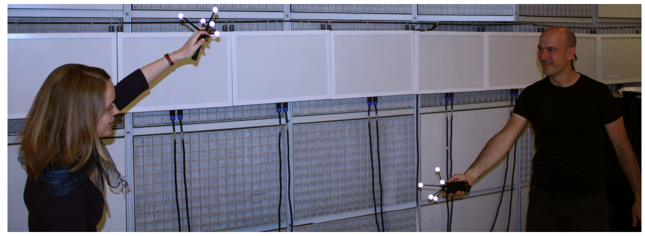

Figure 4. The two-person scenario turns the control of the spatialized virtual instrument into a highly collaborative task in which an exact virtual source position impression adds a distinct advantage.

3. SPATIAL RENDERING IN GRAINSTICK

Traditional rendering techniques such as stereophony and its extensions assume a fixed listener position and create the impression of location of virtual sources by recreating signal levels and relative delays at the speakers as they would sound to the assumed listener. Similarly, the binaural method of spatialization applies filtering to the sound source to be rendered in such a way as to recreate the listener's body's filtering of a sound coming from a particular direction, thus creating a spatial impression. However, the position of the listener's head must be known at all times in order to properly render the spatial scene. In contrast, the wavefield technology used for the Grainstick installation encouraged listener movement and common spatial audio experience regardless of room position–this in turn greatly influenced the interaction design.

3.1. WFS technology and Ircam's setup

In WFS many speakers, often small and closely spaced, are driven individually to recreate a wavefront emanating from a virtual source as it would pass through the array of loudspeakers at a particular distance.

The system used at Ircam for the GrainStick Installation was designed by Sonic Emotion, and used optimized filter designs by Ircam which enabled the creation of focal sources, and the ability to control radiation pattern and apparent directivity of sources [2]. It involves the use of three computers which are responsible for 16 audio inputs and outputs each, for a total of 48 controlled drivers. The drivers are spaced 15 cm apart on 12 multiactuator panels in a metal housing, as seen in figure 4. The length of the entire array is 7.2 meters. Position and directivity messages are sent via UDP using software specific to this installation.

3.2. Spatial Interaction in Grainstick

The fact that the apparent position of virtual sound sources remain stable despite movement on the listener's part has strong implications for the virtual spatial design of interactive applications using this technology. On the philosophical level, designs can be made allocentric, that is, designed around a common coordinate system and stable reference point, rather than egocentric, or designed based on a coordinate system using the subject's head as a reference point.

A virtual sound scene is no longer designed to surround the listener, but rather is rendered as many distinct focal sources that together create a complex scene that can be more or less navigated by the listener through his or her natural movements in the rendering space. Similarly, the allocentric point of view opens up new possibilities for creating highly convincing virtual scenes that can be navigated by more than one person.

In the case of the collaborative scenario of the GrainStick installation, the actual synthesis of the instrument was controlled by the users based on these shared localization cues, since the position along the array of the grains indicated the amount of grains that remained moving in one direction, and thus the point at which the direction could or should be reversed. Thus a collaborative effort is needed in order to initiate and control the evolution of the synthesized sound, and the spatialization effect is a crucial part of this process.

4. FUTURE RESEARCH DIRECTIONS

Questionnaires were filled out by participants during the GrainStick exhibition, part of the 2009 Agora festival at Ircam. The questionnaires asked participants to evaluate the demos displayed by Ircam and the other institutional partners involved in the SAME project. Feedback indicated a positive response and high ratings for impressions of immersiveness for the installation. 16 out of 39 survey responders said the Grainstick gave them a more immersive experience then they find in their daily listening, and 19 felt they listened differently to the installation because they were given control over the synthesis.

Pilot subject tests were performed in Fall 2009 which used the Grainstick as a set of embodied musical stimuli for a test of musical memory. These preliminary results support the importance of spatial and modulation effects for sound recognition in interactive musical settings. Next steps for tests include allowing movement around the space for active participants, and looking at the effect of visual observation on passive listeners.

An expanded version of the Grainstick installation is being developed at the moment that will be produced in June 2010 at the *Cité des Sciences* in Paris. It is planned to enlarge the interaction space to a circular area of 8m in diameter, introduce zones in that space with different sets of grains between which the participant's position interpolates [3], a temporal organisation into scenes and transitions, and video projection.

5. ACKNOWLEDGEMENTS

The authors would like to thank Joseph Sanson and Markus Noisternig for their help with the WFS installation at Ircam.

GrainStick was created and funded under the framework of the EU project SAME (Sound and Music for Everyone Everywhere Everyday Everyway).

6. REFERENCES

[1] F. Bevilacqua, R. Muller, and N. Schnell, “MnM: a Max/MSP mapping toolbox,” in *NIME*, Vancouver, 2005.

[2] E. Corteel, “Synthesis of directional sources using wave field synthesis, possibilities and limitations,” *EURASIP Journal on Advances in Signal Processing, special issue on Spatial Sound and Virtual Acoustics*, 2007.

[3] A. Einbond and D. Schwarz, “Spatializing timbre with corpus-based concatenative synthesis,” in *Proc. ICMC*, New York, 2010.

[4] T. Place, T. Lossius, A. R. Jensenius, and N. Peters, “Flexible control of composite parameters in Max/MSP,” in *Proc. ICMC*, Belfast, UK, Aug. 2008.

[5] N. Schnell, R. Borghesi, D. Schwarz, F. Bevilacqua, and R. Müller, “FTM—Complex Data Structures for Max,” in *Proc. ICMC*, Barcelona, 2005.

[6] D. Schwarz, R. Cahen, and S. Britton, “Principles and applications of interactive corpus-based concatenative synthesis,” in *JIM*, GMEA, Albi, France, Mar. 2008.

[7] O. Warusfel and G. Eckel, “LISTEN – augmenting everyday environments through interactive soundscapes,” in *IEEE Workshop on VR for public consumption*, Chicago, 2004.