

Conducting Sound in Space

Wayne Siegel
Royal Academy of Music, Aarhus
Skovgårdsgade 2c
DK-8000 Aarhus C
ws@musikons.dk
www.waynesiegel.dk

ABSTRACT

This paper discusses control of multichannel sound diffusion by means of motion-tracking hardware and software within the context of a live performance. The idea developed from the author's previous use of motion-tracking technology in his own artistic practice as a composer and performer. Various motion tracking systems were considered, experiments were conducted with three sound diffusion setups at three venues and a new composition for solo performer and motion-tracking system took form.

Author Keywords

Sound diffusion, motion tracking, computer vision

ACM Classification

H.5.2 [Information Interfaces and Presentation] User Interfaces—Input devices and strategies H.5.5 [Information Interfaces and Presentation] Sound and Music Computing.

1. INTRODUCTION

The goal of this project was to explore the potential of electronic music that combines production, performance and diffusion into a single integrated creative process. In pursuing this goal I hoped to develop intuitive ways of controlling sound in space that might be relevant not only for my own practice as a composer and performer but also for other artists facing the challenges of creating spatial electronic music within the context of live performance.

2. BACKGROUND

Since 1995 I have explored the use of motion-tracking technology as a means of controlling musical parameters in live performance. This exploration began with the DIEM Digital Dance project, which focused on tracking the motion of dancers, allowing them to control musical elements. Custom motion-tracking hardware using flex sensors was developed and two interactive dance performance works were created: Movement Study and Sisters. In 2008 I continued my work with interactive dance using other types of hardware, including camera-based technology and accelerometers, in collaboration with dancers and choreographers in what was called The Pandora Project [1]. The camera-based technology that I used consisted of digital cameras mounted in front of and above the stage for tracking the movement of the dancers. I developed interactive software using the cv.jit (computer vision) library [2] in the Max/MSP programming environment. For testing mapping between movement and sound I used my laptop computer with its built-in camera.

3. TWO HANDS (NOT CLAPPING)

In the process of testing and experimenting with this setup I found myself waving my hands in front of my laptop, controlling the sounds intended to be controlled by the dancers. It dawned on me that this activity was both enjoyable and musically interesting with obvious parallels to historical electronic music interfaces such as the Theremin (1920) and The Hands (1984) [3]. I decided to use this system in a new composition, which led to *Two Hands (not clapping)* for solo performer and motion-tracking performance system, a work commissioned by the Dark Music Days Festival and premiered in Reykjavik in 2010.

The mapping used for this work was quite direct. The image from the webcam is divided into a matrix of twelve rectangles. The amount of movement in each rectangle is calculated in software by comparing each video frame with the previous video frame. Using this mapping algorithm I can control twelve sounds individually and dynamically: the more I move in each rectangle the more sound is heard. Multiple sounds can be controlled by moving in more multiple rectangles. The closer my hands are to the webcam, the more they fill the image, which means that by moving my hands closer to the webcam I can control all twelve sounds at once, while more subtle control of individual sounds can be achieved by moving my hands farther from the webcam.

3.1 Conducting (not dancing)

One important difference between working with dancers and working as a performer was that I could make greater demands on controlling the music and concern myself less with the visual content of the performance. Dance is a visual medium and any aspect of interaction that requires specific gestures or movements will limit a dancer's freedom of movement and might interfere with the visual performance. For my solo work, sound was the main focus. Visual content is certainly an important part of experiencing a live musical performance, but the main focus is on sound, not on visual appearance [1]. By comparison, gesticulations of a conductor do not constitute a visual performance but are rather means to the end of making the orchestra perform in a certain way.

3.2 No Water, No moon

In 2011 I was invited to work with the sound system at the Royal Library in Copenhagen. The building, known as the Black Diamond, includes a large public space with a glass facade overlooking the harbor. Permanently installed in this space is a powerful 12-channel sound system with four large speakers (Meyer UP1) on each of the three levels and two subwoofers on the second level.

By coincidence, *Two Hands (not clapping)* used 12 independent audio channels or voices mixed down to stereo output. It seemed natural that the 12 voices could be routed directly to the 12 loudspeakers in the Black Diamond to create a 12-channel version. In testing this setup I found that the result was fascinating. When I moved my hands higher, the sounds activated were routed to the upper speakers. When I moved my hands lower, the sounds activated were routed to the lower speakers. By moving my hands I could very



Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.

intuitively control not only which sounds I wanted to hear but also which of the 12 speakers I wanted to hear. The experience of performing Two Hands on this 12-channel sound system inspired me to create a new, site-specific 12-channel work entitled *No Water, No Moon*, which was commissioned by the Danish Composers Union to commemorate its centennial anniversary and premiered at the Black Diamond on May 4th, 2013.



Figure 1, Wayne Siegel performing at the Black Diamond, Copenhagen

4. MOTION-TRACKING HARDWARE

The use of computer vision to control sound had become an intuitive means of musical expression for me. I wanted to expand this live composition environment to include live control of sound diffusion. Up to this point I had routed each of the 12 channels of my setup to one of twelve speakers. I wanted to be able to control sound diffusion or live panning using motion-tracking technology. My criteria for choosing hardware and software to control sound diffusion were 1) the system must be intuitive and fairly easy to learn how to use and 2) the system must not inhibit or interfere with body movement already being used to control sound. I experimented with three different types of motion-tracking hardware for controlling sound diffusion.

4.1 Computer Vision using cv.jit

Created by Jean-Marc Pelletier, cv.jit is an object library for Jitter that includes tools to assist users in tasks such as image segmentation, shape and gesture recognition, and motion tracking [2]. As mentioned, I have used some of these objects extensively in working with camera-based motion tracking in interactive dance and in my solo works that use motion tracking.

I considered using these same techniques to control sound diffusion but decided against this at an early stage. My main concern was that using the same motion tracking techniques to control both sound production and sound diffusion during a performance would make it difficult or impossible to control these two aspects of the performance independently.

4.2 Leap Motion

Leap Motion™ is a commercially available hardware interface designed for gesture tracking for use as an alternative to a mouse or touch screen. The device is placed on a table in front of the computer and uses built-in infrared LEDs and two cameras concealed behind its glass casing. The gesture tracking is software based. Information about exactly how the software works is not publicly available. The software can track both hands when held above the unit including discrete finger position in a skeletal image. Several object libraries are available to allow integration with Max/MSP, including Masayuki Akamatsu's aka.leapmotion and IRCAM's skeletal tracking software [4], which is based on aka.leapmotion. After some testing I was discouraged by the discrepancy between what my fingers were doing and the screen image of the test software. I also felt that that using the unit required a great degree of dependence on visual feedback. Reading product reviews [5] and Han's & Gold's informative paper on the subject [6] left me no less discouraged. I decided not to use the Leap Motion.

4.3 Hot Hand

The Hot Hand™ USB is a commercially available MIDI-controller manufactured by Source Audio [7]. The controller consists of two units: 1) a 3-axis (X, Y, Z) accelerometer embedded in a plastic finger ring with a built-in bluetooth transmitter and a built-in battery and 2) a separate receiver unit, designed to be connected to the USB-port of a computer. This type of accelerometer is commonly used in various controllers, including smartphones. For my purposes, the main advantage of the Hot Hand over other accelerometers is that it is well integrated into a ring, completely wireless and easily configurable with any music software.

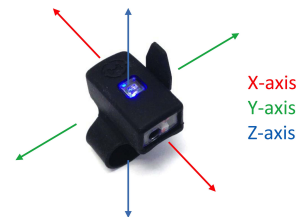


Figure 2, Hot Hand™ USB MIDI controller

4.4 A word on gesture recognition

Motion-tracking hardware can be used in connection with gesture recognition [8]. Both cv.jit and Leap Motion include tools for gesture recognition. The MuBu software platform developed at IRCAM can be used to record sensor data for use in gesture recognition [9]. I chose not to work with gesture recognition for this project because I was interested in mapping motion-tracking data directly to sound diffusion parameters.

4.5 Choice of hardware: two Hot Hands

After testing these three systems I decided to use two Hot Hand controllers, one worn on the middle finger of each hand. The use of accelerometers did not interfere directly with the computer vision tracking already in use and I found the interface to be stable and intuitive.

The Hot Hand outputs 3 controller parameters: X, Y and Z coordinates. I began experimenting with a single Hot Hand controller, but found it difficult to map these three parameters independently to sound diffusion parameters. For example, rotating my hand changed at least two parameters simultaneously. For this reason I decided to use two Hot Hands, mapping only one parameter from each: X-axis on my right hand and Z-axis on my left hand. For my right hand, the "neutral" position was holding my hand with my palm facing left in

relation to myself. By rotating my right hand counter-clockwise I could increase controller values, by rotating my hand clockwise I could decrease controller values. For my left hand the neutral position was with the palm facing down. By raising my left hand (palm facing forward) I could increase controller values, by lowering my left hand (middle finger pointing down) I could decrease controller values.

5. SOUND DIFFUSION

The tradition of sound diffusion dates back to the late 1950's and Musique Concrète, a concept originally conceived by Pierre Schaeffer and others working at GRM (Radio France) in Paris and further developed there and elsewhere. Important sound diffusion systems include the Acousmonium, developed at GRM in the early 1970s, BEAST (Birmingham Electroacoustic Sound Theatre) developed by Jonty Harrison at the University of Birmingham [10] and the GMEBaphone, developed at IMEB in Bourges [11]. Many different approaches to sound diffusion have been taken by different composers and institutions over the years, including the creation of a great diversity of speaker setups as well as the use of various types of hardware and software for controlling sound diffusion. [12].

One musical advantage of controlling sound diffusion live is that the performer can create a site-specific spatial interpretation of a work adapted to the actual listening space. This type of diffusion is often performed by a composer or interpreter moving faders on a large mixing console.

5.1 Controlling Sound Diffusion

One of the problems of controlling sound diffusion with multiple faders is the difficulty of controlling many faders independently. We have 10 fingers, but moving all ten of them dynamically in different directions and at different speeds simultaneously is no easy task. To address this problem various software solutions have been developed, two of which will be discussed below. For my purposes, the problem was greatly complicated by the fact that the motion of my hands was already being mapped to sound control.

5.2 Zirkonium

Zirkonium is software developed at ZKM (Zentrum für Kunst und Medientechnologie) in Karlsruhe, Germany, designed for programming trajectories of sound in space in relation to the Klangdom: a 47-channel speaker array permanently installed at ZKM [13]. Zirkonium software allows a composer to create an independent multichannel panning track for each audio track. The system was designed for programming off line and used for fixed media playback. Some parameters can also be controlled in a live situation using external controllers.

5.3 Spat

Spat (Spatialisateur in French) is a group of software tools developed at IRCAM in Paris and designed for spatialization of sound signals in real-time intended for musical creation, postproduction, and live performances [14]. Spat is suitable for creating virtual placement of sounds in a virtual acoustic environment. For example, using ambisonics, surround sound or a binaural configuration, sounds can be projected to virtual positions and distances in relation to the listener. Spat can also be used for live sound diffusion in a large space with a multichannel setup. Spat can be controlled in a live situation by means of external hardware.

Choice of software

Both Spat and Zirkonium were constructed within the Max/MSP programming environment. I found both to be powerful and sophisticated diffusion tools. Other more specific diffusion tools have also been designed within this environment. For various reasons I found it practical not to use either of these but instead to create my own pan-

ning objects and integrate them into the Max/MSP performance patch that I was already using.

The concept that I chose was a simple one that I call rotational panning. Instead of thinking in terms of panning individual sound sources between speakers, I imagined the whole room rotating left and right, or back and forth. All 12 voices rotated as a group. Each of the twelve channels or voices of my setup were routed to one of twelve fixed speakers. Values transmitted by the two Hot Hand controllers were mapped to panning functions. When I rotated my right hand clockwise all of the speaker positions would rotate to the right, as if I was floating in a fixed position while the whole room rotated clockwise. When I raised my left hand all of the speaker positions would rotate backwards, as if the whole room was rotating backwards.

6. TWO HANDS ON

Intuitive control of sound diffusion is a complex issue. It can be difficult to imagine, realize or even perceive multiple audio sources moving in various patterns and at various speeds at the same time. Controlling complex spatial movement in a live situation can be a great challenge.

I had an opportunity to experiment with live diffusion in three very different spaces. My approach was experimental and site-specific. I viewed the multichannel sound systems embedded in these three spaces not as vehicles for linear sound reproduction but rather as acoustic environments, each with its own unique characteristics.

My first experiments in working with live control of diffusion took place at ZKM in September 2015.

6.1 The Klangdom at ZKM

The Klangdom at ZKM is a small concert hall equipped with a digital mixer and 47 independent speakers. The setup is made up of four rings of speakers. Channels 1- 14 constitute an outer/lower ring, channels 15-28 constitute a slightly higher ring, channels 29-36 constitute a more centered, higher ring, channels 37-42 constitute an even more centered, still higher ring, channel 43 is at zenith and channels 44-47 are subwoofers (one in each corner) [13].



Figure 3, The Klangdom at ZKM, Karlsruhe

My laptop computer was connected to the mixer in the Klangdom via a MADI interface, allowing direct access to all 47 channels from my Max/MSP patch. Much to my delight, this setup was up and running perfectly in less than an hour.

At ZKM I tested my concept of rotational panning using two different configurations. The first configuration can be called 12-12 routing and used only 12 (out of 43 possible) discrete speakers. Rotational panning consisted of changing panning positions between the 12. The second configuration can be called 12-42 routing and used a total of 42 speakers, plus subwoofers (only the zenith speaker was not in use). With 12-42 routing each of the 12 voices was by default routed to a single speaker, but the voice could be panned to six other speakers.

This allowed each voice to be panned to any of a total of 7 speakers (original position, left front, right front, left center, right center, left rear, right rear).

Two of my existing works were used for testing at the Klangdom: *Outside-In* and *No Water, No Moon*.

6.1.1 *Outside-In*

Outside-In is a permanent, site-specific sound installation that I created for the Black Diamond. The installation is heard in this public space 3-4 minutes every day at 1:00 PM. Audio is played from a computer connected to an audio interface with 12 discrete outputs: one routed to each speaker. The work consists of about 100 sections, each 70 seconds in length. An algorithm based on Markov chains determines the section order and amount of overlap between sections. I set *Outside-In* to play continuously so that I could concentrate on using the two Hot Hand controllers for rotational panning of the 12 output channels in the Klangdom. I found that 12-42 routing worked quite well. I could create a feeling of spatial motion with simple and intuitive hand movement. The 12-12 routing configuration worked fine as well, but tended to sound less coherent. Panning with only 12 speakers was not as seamless and subtle as panning with 42 speakers.

6.1.2 *No Water, No Moon*

The next step was to try to perform *No Water, No Moon* live using motion-tracking via a webcam to control sounds and two Hot Hand controllers to control rotational panning. A few challenges immediately became apparent. Both the Hot Hands and the camera-based system react to movement. At first it was difficult to control them independently. After practicing I discovered that it was possible to integrate the motion required to control rotational panning with the motion required to control sound. I stumbled across a few tricks while practicing: 1) I could take my hands “off camera” (for example off to the sides) for a moment to use the Hot Hand controller without affecting the camera-based controller, 2) I could move my whole hand at a fixed angle (for example palm facing down) to affect the camera-based controller without affecting the Hot Hand controller, 3) I could forget about the Hot Hand controllers on one or both hands, making rotational panning less controlled but not necessarily less interesting. After spending some time experimenting I was convinced that this setup was both artistically and technically viable, although it would require some hours of practice on my part. I felt that the panning effects that I could achieve with the Hot Hand controller could be worked into the piece and that the use of live diffusion could influence the performance in a positive way without interfering with the character of the performance. All in all my first impression was extremely positive and practicing *No Water, No Moon* in the Klangdom was a good experience.

6.2 The Black Diamond, Copenhagen

After my visit to ZKM I had an opportunity to conduct further experiments at the Black Diamond as composer in residence at the Royal Library. The permanent 12-channel sound system at the Black Diamond has 12 main speakers and 2 subwoofers hidden in the ceilings on three levels of the main foyer or atrium. The speaker setup is asymmetrical in correspondence with the asymmetrical architecture. The 12-channel setup consists of three trapezoids on three different floors or levels. Distances between the speakers in each trapezoid range from about 10-14 meters. Ceiling height is about 5 meters, with the second level speakers about 11 meters above the ground floor and the third level speakers about 17 meters above the floor. When performing I stand on a bridge on the second level overlooking the harbor (figure 1).

I tested live rotational panning of *Outside-In* using 12-12 routing described in section 5.1. This configuration was easily adapted to this 12-channel sound system. In fact I found that it worked better here than it did in the Klangdom: the panning seemed smoother. This is probably due to the large size and lively acoustics of the space. The acoustics tend to blur panning, making movement between speakers

less obvious, even with a lot of panning motion. Because of the large size of the Black Diamond it is in fact difficult to pinpoint exactly which speaker a sound is projecting from.

I found that the intended rotational effect was not readily perceptible. Instead I had a general sense of sounds moving without being able to pinpoint the exact location of a particular sound at a particular moment or the precise path of the panning motion. I also found that the original 12-channel mix had great influence on how clear or blurred the live diffusion was perceived.

I tested *No Water, No Moon* with the same setup. Again the panning seemed smooth and less obvious than I had imagined. The type of sounds being played had great influence on how rotational panning was perceived: for example the panning of sounds with sharp transients in the higher frequency spectrum was more perceptible than panning static, drone-like sounds in the lower frequency spectrum. Again the rotation of the room was not immediately obvious. My experience was probably influenced by the fact that *No Water, No Moon*, was conceived and composed for performance in the Black Diamond. The 12-channel sound system is part of the concept of the piece. Although I did not feel that live rotational panning detracted from the performance and experience of the work, I was not completely convinced that it added a new dimension, either. But using rotational panning in the Black Diamond was fascinating and did incubate new ideas for future work.

6.3 Symphony Hall, Aarhus

Finally I experimented using a 12-channel speaker setup at Symphony Hall in Aarhus, a hall with a seating capacity of 1,200 that was acoustically designed for symphonic music (Figure 4). There is no permanent sound diffusion system in the hall, so I was at liberty to place the 12 speakers wherever I wished. I chose a flattened setup, using only two levels: 1) stage/ground and 2) balcony (the balcony surrounds the entire hall including behind the stage, the sides and the rear). Speakers were placed on stage (narrow stereo pair plus subwoofers) above/behind the stage (wide stereo pair), two on each side of the audience on the ground floor, two on the side balconies and two on the rear balcony.



Figure 4, Symphony Hall, Aarhus

The concept of rotational panning was not easily adapted to this hall, mainly because of the elongated and flat speaker setup. I concluded that a new method of group panning needed to be designed and implemented for this hall. I was also not satisfied with the speaker arrangement that I had chosen. I felt that it would be necessary to develop a new rotational panning system designed specifically for the Symphony Hall in Aarhus and create a new speaker setup layout with more speakers in front of the audience and fewer on the sides.

6.4 Ritual

Based on my experiments I found that using two accelerometers to control rotational panning could be combined with camera-based motion tracking to provide a flexible and intuitive interface for controlling live diffusion. I ultimately chose this configuration for a new work for solo performer and motion-tracking system.

This work, entitled *Ritual*, employs both camera-based motion tracking using the webcam of a laptop computer as well as a pair of accelerometers, one worn on each hand. The webcam controls sound in two different ways: 1) altering the amplitude envelope of continuous looped samples and 2) triggering single samples when movement in any given zone increases beyond a fixed threshold. The accelerometers are used to control live rotational panning using only one control parameter from each accelerometer. The webcam image is divided into a matrix of 12 rectangles. Each rectangle is linked to a sound. The performer controls the sounds by moving his hands within the 12 rectangles. Two accelerometers (Hot Hands®) are also worn by the performer. These are used to control panning and placement of the sounds in the hall using the 12-channel sound diffusion system. Live control of sound diffusion is an important element in the work.

Ritual consists of 18 sections or scenes. Each scene has its own sounds. Changes of scenes can be either abrupt or gradual with overlapping transitions between scenes. The mood of the piece is contemplative, evoking associations to an unknown ritual. A beta version of the piece was tested at Aarhus Symphony Hall in February, 2016. The work received its premiere performance on April 30, 2016 at the Black Diamond in Copenhagen as part of the composer's affiliation with the Royal Library in Copenhagen as composer in residence 2015-2017. The author was invited to perform *Ritual* at the Black Diamond on May 18th, 2017 at the closing concert of the NIME 2017 Conference in Copenhagen.

7. CONCLUSION

Working with various types of motion tracking for controlling sound diffusion has provided me with insight and inspiration in relation to creating spatial electronic music within the context of live performance. After intensive experimentation using three sound systems in three spaces I decided to use two different motion-trackings systems: a camera-based system to control sound and an accelerometer-based system to control sound diffusion.

The simplicity of this mapping has made the interface fairly easy for me to learn to use. In spite of this simplicity, I have found that complex sound textures can be created by combining and mixing 12 voices and that subtle and musically relevant multi-channel sound diffusion can be controlled during a performance. The potential of creating varied sonic textures and multi-channel panning by means of a few simple parameters controlled by motion tracking hardware and software continues to fascinate me, and this fascination has inspired me to create a new work that explores the idea of conducting sound in space.

8. ACKNOWLEDGMENTS

I would like to thank the Royal Academy of Music in Aarhus, the Royal Library in Copenhagen and ZKM in Karlsruhe for supporting this project. I would also like to thank the Danish Arts Foundation and the KUV fund of the Danish Ministry of Culture for financial support.

9. REFERENCES

- [1] W. Siegel, "Dancing the Music", The Oxford Handbook of Computer Music, Oxford University Press, 2009.
- [2] M. Pelletier, Computer Vision for Jitter. Accessed 15 December 2015. [Online]. Available: <http://jmpelletier.com/cvjit/>.

- [3] M. Waisvisz, "The Hands: A Set of Remote MIDI-controllers", Proceedings of the 1985 International Computer Music Conference, Vancouver, Canada, 1985, pp. 313-318.
- [4] Leap Motion skeletal tracking in Max. Accessed 15 December 2015. [Online]. Available: <http://ismm.ircam.fr/leapmotion/>
- [5] R. Metz, Look Before You Leap Motion. MIT Technology Review. Accessed 15 December 2015. [Online]. Available: <http://www.technologyreview.com/news/517331/look-before-you-leap-motion/>
- [6] J. Han and N. Gold, "Lessons Learned in Exploring the Leap Motion Sensor for Gesture-based Instrument Design," Proceedings of the International Conference on New Interfaces for Musical Expression, London, 2014, pp. 371-374.
- [7] J. Helfer, Review: Source Audio Hot Hand USB wireless controller, Accessed 15 December 2015, [Online]. Available: <http://djworx.com/review-hot-hand-wireless-controller/>
- [8] A. Camurri, G. Volpe, S. Menocci, E. Rocca, and I. Vallone. "Expressive gesture and multimodal interactive systems." Proceedings of AISB 2004 Convention: Motion, Emotion and Cognition. Leeds, 2004, pp. 15-21.
- [9] Jules Françoise, Norbert Schnell, Riccardo Borghesi, and Frédéric Bevilacqua. "Probabilistic Models for Designing Motion and Sound Relationships," Proceedings of the 2014 International Conference on New Interfaces for Musical Expression, London, UK, 2014, pp. 287-292.
- [10] J. Harrison, "Sound, space, sculpture: some thoughts on the 'what', 'how' and 'why' of sound diffusion," Organised Sound, vol. 3, no. 2, pp. 117-127, 1998.
- [11] C. Clozier, "The Gmebaphone Concept and the Cybernéphone Instrument," Computer Music Journal, vol. 25, no. 4, pp. 81-90, 2001.
- [12] S. Emmerson, Living Electronic Music, Ashgate Publishing, 2007.
- [13] C. Ramakrishnan, J. Großmann and L. Brümmer, "The ZKM Klangdom", Proceedings of the 2006 International Conference on New Interfaces for Musical Expression, Paris, France, 2006, pp. 140-143.
- [14] T. Carpentier, M. Noisternig and O. Warusfel, "Twenty Years of Ircam Spat: Looking Back, Looking Forward," Proceedings of the 2015 International Computer Music Conference, Denton, Texas, USA, 2015, pp. 270-277.

10. APPENDIX

Further information about the author can be found here:
<http://waynesiegel.dk/>

Further information about works mentioned in this paper including video examples can be found at the following web locations:

Two Hands (not clapping)
http://waynesiegel.dk/?page_id=773

No Water, No Moon
http://waynesiegel.dk/?page_id=1263

Ritual
http://waynesiegel.dk/?page_id=2679