DESIGNING GESTURAL INTERFACES FOR LIVE SOUND DIFFUSION

Bridget Johnson
[1] New Zealand School of Music
Victoria University of Wellington

Jim Murphy
New Zealand School of Music
Victoria University of Wellington

Ajay Kapur [1,2]
[2] California Institute of the Arts
School of Music

ABSTRACT

This paper presents a new paradigm in interface design for diffusion performance. It looks at past and current standings in the diffusion paradigm and introduces new custom-built controllers for the performance of space. The spatialisation techniques and algorithms that have been implemented into such systems are discussed, along with a focus on how they encourage a wide array of potential sonic trajectories available to the performance artist. It is hoped that the techniques and developments described in this paper afford performers and composers an enhanced level of creative expression and encourage a conscious engagement with space for performers and audiences alike.

1. INTRODUCTION

After a brief introduction to the field of diffusion and significant developments that have taken place in the last twenty years, two new interfaces are introduced. Discussions of the tools implemented to build the interfaces and case studies on each interface with their specific motivations, capabilities and performance examples are provided with a focus on the increased spatial expressivity the new interfaces afford the performer.

For over half a century, electro-acoustic musicians have looked to the spatial characteristics of their pieces as an element to be manipulated and performed in the concert setting. This paradigm of the diffusion performance has evolved since its inception, undergoing many changes in technology. Until recently, these developments have largely been concerned with the software controlling the spatialisation. However, the last 10 years have seen an increasing desire to develop the performance interfaces used for diffusion, encouraged by the field of new interfaces for musical expression (NIME). In contribution to the ongoing developments in diffusion, this paper represents the authors' contribution to the field through describing two new interfaces.

Sound diffusion in the 1970s was characterised by large-scale travelling speaker orchestras been built by institutions in Europe and Great Britain, the most notable of which are the Gmebaphone [2], BEAST [6] and the Acousmonium [3]. These early systems allowed the performer to control the amplitude of individual or groups of speakers with a mixing desk. The system itself provided a coloration of the piece through its specific inclusion and spatial positioning of certain speakers. As these systems developed and amplitude panning was refined, the performer was given heightened control of the spatial positioning of sounds and therefore an increase in potential expressivity in performance. However, these developments were largely restricted to the software controlling the spatial movement of sounds, meaning that, while performers could pre-program or trigger certain sonic trajectories, they were not able to intuitively perform them live.

Systems like the Gmebaphone (later known as the Cybernephone) included an ability to be controlled remotely over a network, featuring pre-programmable sonic trajectories and two control screens for graphical user feedback; however, these systems limited performance interaction to mouse and keyboard computing and gain control through fader banks on a mixing desk. BEAST acknowledged the limitations of the mixing desk: Harrison [5] argues that it was designed for multiple inputs to be mapped to minimal outputs, the opposite of which is desired in sound diffusion. To counteract this problematic interface, the BEAST team developed a new diffusion interface which is essentially an inversed mixing desk, allowing from two to eight channel inputs to be individually controlled through a large number of speaker outputs. While this customised mixing desk interface affords the diffusion artist greater control and configurability of spatial movements and positions, Mooney states that "the biggest criticism [of the system] is relatively poor interface ergonomics, stemming from the fact that diffusion must be executed on a one-fader-to-one-loudspeaker (or group of loudspeakers) basis" [12].

In more recent years, electronic performance artists have begun to use existing interfaces as performance tools. Interfaces popular in the NIME community, such as the WiiMote and GameTrak, allow the performer to map transparent performative gestures to any number of musical parameters. It is possible for diffusion artists to adopt these NIME-developed techniques, instead mapping the data from the game device into more sophisticated spatialisation algorithms. While the gestural nature of such user interfaces can give the artist a more dramatic spatialisation performance than a mixing desk, they are still limited by the restrictions of the interface. It is these limitations that we hope to rectify by the development of custom interfaces for diffusion performance.

2. TOOLS

This section describes two new interfaces designed as gestural interfaces for diffusion performance. The first, *tactile.space*, is a user configurable multi-touch interface; the second, Chronus, has been designed as part of a collection of interfaces to complement and control the Bacchus sound sculpture, discussed in more detail below.

2.1. Table top surfaces

Diffusion performance interface *tactile.space*, discussed in more detail in [8], was designed to run on BrickTable, a multi-touch tabletop surface designed by music technologists Owen Vallis and Jordan Hochenbaum [7]. The BrickTable uses the reacTIVision framework [10] as originally developed for the reacTable [11]. Large scale multi-touch surfaces have much potential as a gestural performance interface: it is for this reason that the system was explored as a possibility for a diffusion performance interface.

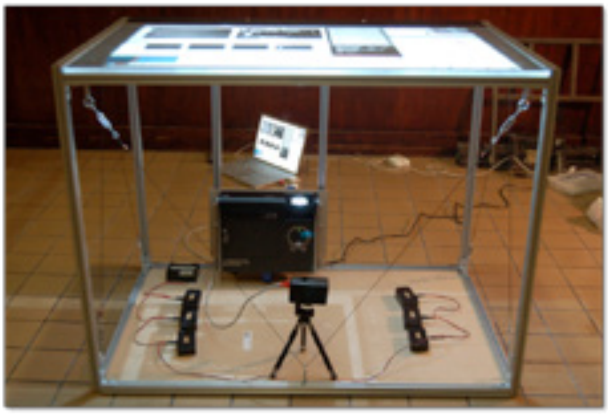

Figure 1. The BrickTable [7]

The original musical applications for touch tables, including those on the reACTable and AudioPad [13], relied on the placing of physical objects, called fiducials, on the table's surface. These objects were then tracked by computer vision software and their positions and movements mapped to musical parameters. In 2009, the reacTIVision tracking software was given the added capability of finger tracking [11], no longer limiting the potential tracking data to that of the amount of fiducials available. A major advantage of finger tracking over fiducial tracking is that it allows user interaction with virtual application specific objects that can be built into the graphical user interface (GUI), with far-reaching implications for gesture recognition as in other multi-touch devices. Such techniques exhibit great potential for the development of multi-touch diffusion performance interfaces.

2.2. Arduino Encoder-Based Interfaces

In designing *Chronus*, a new interface for the specific purpose of gestural and intuitive diffusion performance, the decision was made to pursue a rotary encoder-based design. The rotary encoder is similar to a standard rotary potentiometer (knob) in its appearance but differs in one very specific way that makes it far more suited for diffusion: it can rotate continuously in a circular motion past the 360 degree point. If a standard rotary potentiometer was mapped directly from its point to a position in space, the potential sonic trajectories directly controlled by the potentiometer would be limited to with clockwise or counterclockwise movement up to 360 degrees; therefore, continuous circular motions would not be possible. Rotary encoders remove this limitation, allowing for continuous circular spatial control: sound sources can be spun continuously around a space.

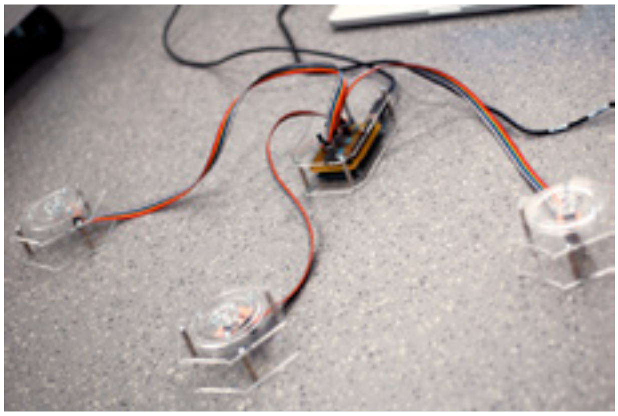

Figure 2. Rotary Encoder Performance Interface

The position of the encoder is read by an Arduino Mega microcontroller. The ATMEGA 2560-equipped Arduino Mega was chosen for its multiple external interrupts, allowing many rotary encoders to easily be simultaneously decoded. Additionally the Arduino platform was chosen due to the availability of musical interface-specific firmware libraries [4]. The custom Arduino firmware sends serial data to Processing, which in turn can send either OSC [15] or MIDI messages to any audio application. The mapping of this data is discussed in section 4.2.

3. SPATIALISATION ALGORITHMS

The traditional setup for a diffusion performance involved the mapping of each fader on a mixing desk directly to the gain of a particular speaker. As speaker orchestras grew in size, the number of speakers quickly outweighed the number of faders; therefore, speakers were divided into pairs or groups with one fader controlling the gain for the entire group. The possible spatial trajectories are limited by the physical restrictions of the performer and the specific grouping of the speakers. The subsequent sections outline the spatialisation algorithms that have been implemented into these new diffusion interfaces with an emphasis on their potential to give a performer an increased expressivity in their diffusion.

3.1. Stereo Pairing

Stereo pairing is the most traditional spatialisation technique implemented. This mode divides the speaker configuration into vertical pairs (as shown in Figure 3) and allows the performer to control the pair of speakers to be used and the stereo spread between said pair.

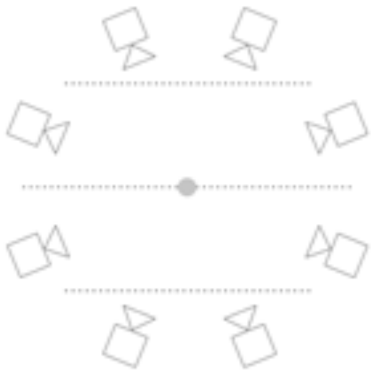

Figure 3. Division of Stereo Pairs

While this technique limits the possible sonic trajectories that may be performed live, it does maintain a great deal of flexibility in speaker configurations, as it has no real reliance on equidistant speakers. The system is designed to fade in and out of each speaker pair as a sound moves through the space. Though the default setting for fade time is 500ms, the performer may configure this to their needs.

3.2. Vector Base Amplitude Panning

Vector Base Amplitude Pairing (VBAP) is a spatialisation technique that was introduced by Ville Pulkki in the 1990s [14]. VBAP extends what had already been achieved in frontal stereo phantom source positioning into a 2-dimensional ring of speakers, thus allowing the creation of a phantom image at any point in a pantophonic array. Unlike in the Stereo Pairing spatialisation technique, each speaker in the array is part of two pairs of speakers, one with each of its adjacent speakers. Pulkki called this pair-wise panning. Once the appropriate pair of speakers for creating a phantom source in the desired location has been deciphered, any pan pot algorithm may be used to control the gain factors for the speakers creating the image.

This technique allows for an accurate creation of phantom images in a pantophonic array, but does not allow for any sense of spatial depth; as such, all sounds are perceived at the edge of the speaker array.

3.3. Source Spreading

The Vector Base Amplitude Panning technique of spatialisation allows the creation of discrete phantom source positions with in the spatial field, but doesn't allow for size or spread of that position, nor does the Stereo Pairing technique. In an attempt to increase expressivity in diffusion performance, the ability to control spatial spread of a source was developed. In the *tactile.space* multi-touch user interface this is done by placing a second finger inside an audio object and separating the two fingers in an arch shape around the representation of the sweet spot.

By spreading a source position, the performer may create partial of full immersion within the sound field. The position of each of the edges of the source is deciphered by the same techniques discussed in section 3.2; subsequently, any speakers that fall between the two extremes of the arc are set to a gain factor of 0.8, creating a wall of sound of the desired size. If the arc shape reaches a width of 345 degrees, it is assumed that the desired effect is a full immersion in the sound field and the gain factors for all speakers are updated accordingly.

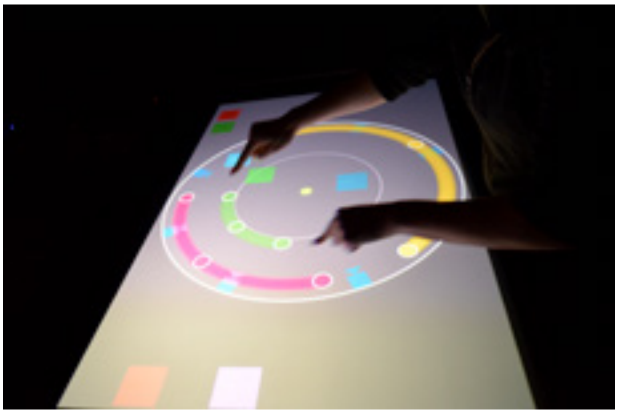

Figure 4. User Controlling The Spread of A Sound on *tactile.space*

In the example above, shown in Figure 4, the user may update the size or spread of an object by repositioning the arc's edge by moving the white circle at either end of the arc. The middle circle allows the user to move the spatial position of the arc whilst maintaining its established spread. The addition of source spreading capabilities and dynamic updating of the source width to source positioning techniques gives the diffusion performance artist an increased level of creative options in their performance and a greater variety of potential sonic trajectories.

3.4. Distance Encoding

A common problem among the majority of spatialisation techniques including those discussed in this paper is their ineptitude in creating a perception of spatial depth in a pantophonic speaker array. In an attempt to provide the performer with more expressivity in the diffusion, we have implemented a distance-encoding algorithm into these systems. The distance encoding system is based on an implementation of the *1/r* law of atmospheric absorption. As sound waves travel through the air, they encounter friction caused by contact with air particles. Therefore, the amplitude of a sound decreases with any distance traveled [1]. Further, the spectral content is lost at varying rates depending on the amount of distance traveled; the higher frequencies drop off first, with a subsequent decrease in the cutoff frequency accelerating with further distance traveled. In order to simulate this phenomenon, spectral filtering is applied to the waveform of the incoming sound source based on the perceptual distance desired. The final signal process applied to simulate a change in distance perception of the space is a small amount of reverberation, mimicking the natural reverberation that occurs with a sound that has traveled a further distance. Listeners hear the reflection of the space relatively closer to the original sound source causing confusion between what is the original sound source and the reflected sound that is perceived as resonance. This signal processing is implemented in Max/MSP[1].

As these laws are based on sounding objects in a free field environment and not their reconstruction on a speaker-based system, the best one can hope for is an illusion of perception of spatial depth. While the distance encoding was conceived as a way to allow the performer to place sounds within the edge of the speaker array and the sweet spot, it is important to consider that human distance localization is much more accurate in relative rather than precise localization. Therefore, while it may seem pertinent to assume the maximum distance for a piece should be the distance from the speakers to the sweet spot, it was found by performers that setting this distance for aesthetic rather than physical reasons allowed greater expressivity in performance. Some performers chose a small perceptual space and opted for subtle and finite changes within the space, while others went for greater width to give the perceptual depth range potential for more variation within their piece. The maximum distance for the distance encoding system may be set by the performer at any distance between 1 and 20 meters. Each meter of perceptual space is represented on the GUI by a faint grey circle allowing the user visual feedback for distance location as shown in Figure 5.

4. PERFORMANCE CASE STUDIES

The following case studies describe two different spatialisation performance interfaces that have been designed and developed by the authors. The two interfaces were developed for differing contexts in order to meet the needs of performers and aesthetic desires for specific pieces. Both interfaces were designed with a goal to increase the expressive and performative qualities of a diffusion performance and incorporate the techniques discussed in section 3.

4.1. tactile.space

tactile.space is a multi-touch performance interface for sound diffusion. It was developed as a generic interface that could be configured to the needs of a variety of performers and composers and could be used in a concert setting in the place of a mixing desk (the most common diffusion user interface). One of the major design considerations was to create an interface that was highly intuitive and very easy to use, so the diffusion artist would not need to undergo extensive training on the interface. It was hoped that such an interface might also increase the gestural performance elements of diffusion practice and allow the performer a heightened level of expressivity. As such, the interface was developed to be easily configured to each performer's aesthetic desires. Some user-defined settings include the number of speakers and their configuration, the number of audio input channels and their type (live or audio file), the maximum distance desired for the distance encoding functions, and various fade rates.

Since its development in 2012, *tactile.space* has been used in a number of concert settings. One of these settings was a traditionally-inspired diffusion concert featuring eight performer/composers diffusing their own acousmatic works. Each performer was able to configure the interface in their own way depending on the aesthetic needs of their piece.

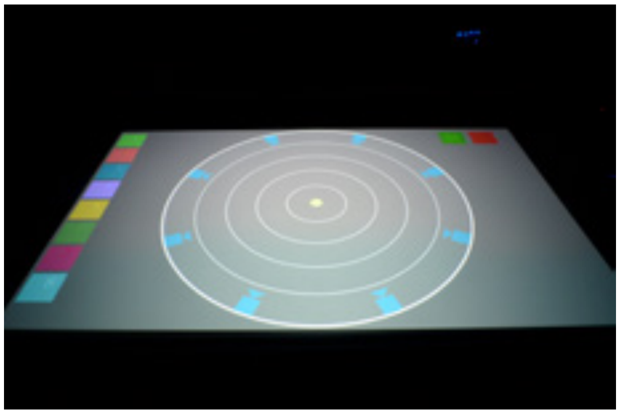

Figure 5. The start up GUI of *tactile.space*

The performers split their fixed media compositions into a number of audio stems; some based on frequency bands, others grouping specific sonic gestures or textures.

On start up, the performer is presented with a graphical user interface that has a visual representation of each of their audio stems and the spatial field (shown in Figure 5). The user may then touch and drag the audio objects displayed on the right of the BrickTable into the speaker array in their desired spatial location; the gain factors and distance functions will be calculated to create a phantom source image in that location.

tactile.space has also been used as a diffusion tool for the live electronics piece *nebular*, which was performed at the 2012 New Zealand Electro-acoustic Music Symposium[2]. *Nebular* features Blake Johnston's eZither [9] as the input device; *tactile.space* receives 8 channels of live input from the eZither, each having undergone a separate audio effects process. The collaborative piece allows for a number of interactive relationships to be developed throughout. The improvised nature of the piece means both the performers have a dynamic relationship with each other and the spatial field. The eZither is able to react to sonic trajectories in real time, while *tactile.space* can react to gestural and textural sonic events. The visual element and gestural nature of *tactile.space* provides musicians collaborating with diffusion artists the opportunity to see and predict potential sonic trajectories that the diffusion artist may be developing in the performance. These spatial relationships, which in traditional diffusion setups may have been opaque or difficult to achieve, are brought forward with *tactile.space*, creating

[1] http://cycling74.com/

[2] The piece may be viewed at http://vimeo.com/45230938

transparency in performance for the collaborator and audience alike.

4.2. Chronus

Unlike *tactile.space*, Chronus was a piece-specific interface. It was built for diffusion performance with the authors' sound sculpture Bacchus: a desire for portability and aesthetic continuity drove the development of Chronus. The piece was performed in November 2012 at the University of California Irvine.

Chronus features three rotary encoders, as described in section 2.2. Two of the encoders map directly to the angle of the source position to be deciphered into speaker gains through the VBAP spatialisation technique. The third encoder controls the radius for the distance perception. The angle position data is mapped to two separate channels of incoming audio data from the Bacchus sculptures and the radius data responds to both signals unifying the sculptures.

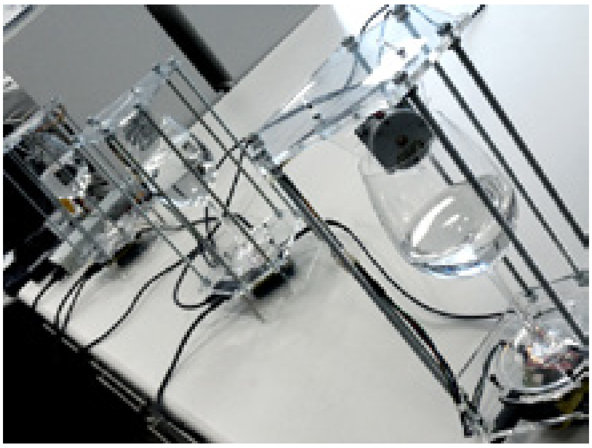

Figure 6. The Bacchus sound sculptures, whose design informed the appearance of the Chronus interface in Figure 2.

The aesthetic design of the Chronus interface was intended to resemble the Bacchus sculptures and control interface as closely as possible so the piece would be conceived as a coherent visual whole. Bacchus, shown in Figure 6, is a sculpture focusing on microsounds produced by mechanically-plucked glass objects, and is built with a variety of transparent materials; these transparent materials were further used in the Chronus interface. During performances involving the Chronus interface and the Bacchus sound sculptures, the speed of the motorized plucking mechanisms of the Bacchus sculptures are controlled by a custom potentiometer-equipped Arduino-based MIDI interface. Like the Bacchus sculptures and the Chronus spatialisation interface, the motor control interface is built of transparent materials: audiences viewing performances involving these three elements are presented with visually cohesive sound generation and modification apparatus.

The sonic output of Bacchus is transmitted via a microphone to Chronus's software and is subsequently diffused through a multichannel speaker array.

The expressive capabilities of Chronus encourage a similar interaction with space from both performers as in *tactile.space*. With the increased intuitive relationship between performative gestural and sonic trajectory, both performers may easily read, and therefore predict and react to, each other's intentions. While this phenomenon is arguably one that comes from two performers practicing and understanding each other's musical intentions and styles, this is often hindered with a diffusion artist's lack of accessibility to suitable rehearsal spaces and the abstraction of their standard user interface. By advancing the performance interface for spatialisation the diffusion artist is able to use the space expressively and have their collaborator do the same.

5. CONCLUSIONS AND FUTURE WORK

Currently, a second iteration of *tactile.space* is being developed. This version is being targeted to mobile computing platforms, as well as including further gesture recognition to allow more complex sonic trajectories to be achieved and including new spatialisation algorithms. The Chronus interface has inspired further designs of spatialisation tools for performance with sound sculptures and interactive installation pieces. It is hoped that more of these types of interfaces will be developed soon. The single-point-source nature of the encoder-based Chronus has, up to this point, not allowed for any variation of the width of phantom sources, current research includes looking for a way to allow this without compromising the intuitive and gestural simplicity of the interface.

These two new interfaces represent a new trend in diffusion performance wherein the focus is not just on a spread of sound through space, but also the means by which the spatialisation occurs and is controlled by the performer. The interfaces encourage the performance of the spatial element of any given piece to be just as dynamic and engaging to the audience as is any other element. Whilst utilizing the same techniques and spatialisation algorithms, the two interfaces had different practice concerns, therefore the aesthetic design of each interface is individualized. A user study evaluating the success of *tacile.space* was completed, the findings are available in Johnson (2013) [8] .

6. REFERENCES

[1] Blauert, J. *Spatial Hearing The Psychophysics of Human Sound Localization*, Revised Edition, English Translation. Cambridge Massachusetts: MIT Press, 1997.

[2] Clozier, C. "The Gmebaphone Concept and Cybernephone Instrument," *Computer Music Journal MIT Press*, vol. 25, no. 4, pp. 81–90, 2001.

[3] Desantos, S. Roads, C. and Bayle, F. "Acousmatic Morphology: An Interview with Francois Bayle,"

Computer Music Journal MIT Press, vol. 21, no. 3, pp. 11–19, 1997.

[4] Diakopoulos, D. and Kapur, A. “HIDUINO: A firmware for building driverless USB-MIDI devices using the Arduino microcontroller,” in *Proceedings of New Interfaces For Musical Expression*, Oslo, Norway, 2011.

[5] Harrison, J. “Sound, space, sculpture: some thoughts on the ‘what’, ‘how’ and ‘why’ of sound diffusion,” *Organised Sound, Cambridge University Press*, vol. 3, no. 2, pp. 117–127, 1999.

[6] Harrison, J. and Wilson, S. “Rethinking the BEAST: Recent developments in multichannel composition at Birmingham ElectroAcoustic Sound Theatre,” *Organised Sound, Cambridge University Press*, vol. 15, no. 3, pp. 239–250, 2010.

[7] Hochenbaum, J. and Vallis, O. “BrickTable: A Musical Tangible Multi-Touch Interface,” in *Proceedings of Berlin Open Conference 09*, Berlin, Germany, 2009.

[8] Johnson, B. “Multi-Touch Interfaces For Phantom Source Positioning In Live Sound Diffusion”, in *Proceedings of New Interfaces For Musical Expression*, Seoul, Korea, 2013.

[9] Johnston, B. “Ezither: Extended Techniques For Customised Digital Bowed String Instrument,” in *Australiasian Computer Music Conference*, Brisbane, Australia, 2012.

[10] Jordà, S. Kaltenbrunner, M. Geiger, G. and Bencina, R. “THE REACTABLE,” in *Proceedings of New Interfaces For Musical Expression*, Vancouver, Canada, 2005.

[11] Kaltenbrunner, M. “reacTIVision and TUIO: A Tangible Tabletop Toolkit,” in *Proceedings of ITS’ 09*, Banff, Canada, 2009.

[12] Mooney, J. “Sound Diffusion Systems for the Live Performance of Electroacoustic Music,” Ph.D, University of Shefield, Sheffield, United Kingdom, 2005.

[13] Patten, J. Recht, B. and Ishii, H. “AudioPad: A Tag Based Interface For Musical Performace,” in *Proceedings of New Interfaces For Musical Expression*, Dublin, Ireland, 2002.

[14] Pulkki, V. “Virtual source positioning using vector base amplitude panning,” *Journal of Audio Engineering Society*, vol. 45, no. 6, pp. 456–466, 1997.

[15] Wright, M. Freed, A. and Momeni, A. “Open Sound Control: State of the Art 2003,” in *Proceedings of New Interfaces For Musical Expression*, Montreal, Canada, 2003, pp. 153–159.