DEVELOPING INTERACTIVE STRATEGIES FOR THREE DIMENSIONAL SOUND SPATIALIZATION

CHRISTOPHER HUMPHRIES

A thesis submitted to the University of Huddersfield

in fulfillment of the requirements for the degree of

Master of Arts by Research

The University of Huddersfield

January 2010

Abstract

This thesis, and accompanying materials, is a research project aimed at experimenting with various alternative methods for facilitating three-dimensional sound spatialization, with a focus on gestural controllers. The aim of the project was to investigate and evaluate alternative ways to attempt to narrow the gap between technology and creativity by developing a more intuitive controller than the standard keyboard and mouse configuration, thereby allowing for more of a direct correlation between the ideas and the result.

During the research period, the author created four hardware interfaces that could potentially fulfill this brief, as well as a number of patches in Max/MSP to implement and control them. These interfaces were then tested by a small number of independent persons, and the results assessed and evaluated. The tests raised some important points to be taken into consideration for future development of gestural controllers.
As an accompaniment to this practical work, this thesis describes the progression of the research including challenges faced and how they were overcome, as well as an evaluation of the practical aspects of gestural control in general.

Acknowledgements

Professor Michael Clarke and Mr Mark Bokowiec for guiding me through the research and generally supporting my work.

Dr. Pierre-Alexandre Tremblay for first introducing me to sound diffusion and spatialization, and for allowing me the use of his equipment to experiment.

Elizabeth Dobson, Mike Fisher, Sam Freeman, Richard Glover, Adam Jansch, Dr Scott McLaughlin and Jonathan Rich for giving up their time to test the interfaces.

I would also like to thank Mark Paine and the staff at Headington School, Oxford for allowing me to continue and complete this project whilst in their employ.

Contents

List of Figures

Word Count: 10,347

‘A computer terminal is not some clunky old television with a typewriter in front of it. It is an interface where the mind and body can connect with the universe and move bits of it about.’ – Douglas Adams, *Mostly Harmless* (1992)

1. Introduction

The purpose of this project is to research methods of 'performing' sound spatialization, and to ascertain whether there is a more efficient, ergonomic, or intuitive way to move sound around a three dimensional space than those at composers' disposal today.

The motivation for this research arose from the author's own experience of composing in the multi-channel domain, through dissatisfaction with the means available to move sound around a space, namely either a diffusion setup using a bank of faders, or the standard keyboard and mouse combination that the vast majority of computers now use. It seemed that there must be a more direct method to moving a sound around a space than having to know which speaker was assigned to which channel, what the value was in dB for one speaker in relation to another or having to get the necessary logarithmic crossfade just right using a pair of faders, just to pan between two speakers. Arguably this makes technology a barrier between composer and composition rather than an aid, which is not its intention. In order to address this problem it was decided to investigate alternative methods for spatializing sound based around gestural movement, which hopefully would allow composers to bridge that gap between idea and realization by 'performing' their spatialization in the studio.

This project was concerned with finding various alternative methods for controlling sound spatialization in three dimensions. Sometimes this meant using existing commercially available hardware, and other times custom elements had to be created. The purpose is to assess strategies for controlling spatialization and testing alternatives by comparing interfaces.

Though due to the scale and budget of the 12 month project it has not been possible to create a final product of any interfaces, working prototypes have been developed and assessed to test hypotheses and draw conclusions about new ways to approach the task. A DVD has been provided with this submission containing a video demonstration of the software and devices in use (Appendix 10.1).

2. Literature Review

It is only in the last ten to fifteen years that compositions for, and performances of, multi-channel electroacoustic music have become more common, and even more recently for three dimensional works. However, one of the first to investigate multichannel movement was Pierre Schaeffer [1910-1995], whose early concerts were presented using four static channels, with a fifth 'live' channel that was sent to the four speakers and moved in real time via a joystick. Also pioneering composition in the multi-channel domain was Karlheinz Stockhausen [1928-2007] when he performed his four-channel piece *Gesang der Jünglinge* (1955-56), which was billed as the first electronic music for four tracks[1]. Soon after, he wrote *Kontakte* (1958-60), which incorporated a speaker placed on a rotating table which he recorded using four microphones on a multitrack tape machine. Stockhausen was then able to project these around the performance space, creating realtime sound spatialization[2]. Years later he would again push the boundaries again in his work *Oktophonie* (GER) (1990). The 69 minute piece was written for a 'cube' of 8 sets of speakers, positioned around the audience who would be sat on the bottom plane of the cube. Although it was a fixed media (tape) piece, a large amount of the spatialization would be performed live in the studio by Stockhausen himself. He incorporated a large variety of techniques

[1] 'Stockhausen introducing "Kontakte" at The Royal College of Music in Stockholm 12th May 2001', *Stockhausen in Stockholm May 2001*, http://home.swipnet.se/sonoloco7/stockhausen/stockhausenmay2001.html (Accessed 28th January 2010)

[2] Stockhausen, K., *Kontakte*, (London, 1968)

including static mixing (a mono source sent to four faders to position on a plane[3], and then group faders to crossfade between them) to circular movement (controlling each plane's individual movement with a joystick). He also combined the two to allow movement in spirals, controlling movement within a plane using a joystick (or pre-recorded using MIDI) and fading between planes horizontally, vertically, or even diagonally using additional faders. In order to achieve these highly complex movements, he had to use custom modified controllers. For example, the QUEG (QUadrophoner Effekt Generator - a four channel spatialization controller) and Yamaha DMP-7 mixer playing sequences controlled by a Fadermaster (MIDI Command Controller)[4]. This hardware may seem crude by our standards today, but at the time these were revolutionary controllers, and were the precursors to most of what we use as interactive sound interfaces in modern times. Also, Pierre Schaeffer's early concerts were presented to an audience using four static channels, with a fifth 'live' channel that was sent to the four speakers and moved in real time via a joystick. Twenty years later, John Chowning [b. 1934] was also experimenting with spatialization and movement of sound around a space. For his work *Turenas* (1972), he created a front end control for Stanford's Music 10 language that allowed the user to create a sonic path for FM synthesized sounds in a quadraphonic spatial environment[5]. The program allowed the user to plot graphically a route for the sound around the space, and would calculate the necessary volume, reverb, and even Doppler shift for each speaker to provide the illusion of that position. Though this was pre-programmed rather than done in real-time, this is doing what programs like IRCAM's Spatialisateur do today (see page 14).

[3] A 'plane' of sound is a way of describing a two-dimensional wall of sound that is typically then moved around a three-dimensional space. It may be vertical, horizontal, or diagonal in any direction.
[4] Clarke, J.M. and Manning, P., *The influence of technology on the composition of Stockhausen's Octophonie, with particular reference to the issues of spatialization in a three-dimensional listening environment.* Organised Sound, 13 (3). pp. 177-187

Talking about *Turenas* in conversation with Curtis Roads in 1982, Chowning explains his software:

> "So what I did was write a program that incorporated a distance cue, an angular cue, and a velocity, in such a way that a composer could use it gesturally."[5]

Though it may not have been real-time or using hardware, it seems there has long been a desire for gestural control of spatialization.

One of the first uses of gestural movement to 'perform' sound was in 1984 when Michael Waisvisz [1949-2008] created an electronic instrument, or more a multi-function controller, called 'The Hands'[6]. As the name suggests, the device is in two parts, and is worn as gloves. There are various sensors mounted on the device that allow the user to control elements of the sound source (either recorded from a built-in microphone or externally sourced) using buttons as well as proximity and gyroscopic sensors. The effect of this is that the performer can control multiple aspects of a sound gesturally, and can interact much more organically with the sound than with a computer. This produces a dramatic performance aspect for the audience, but the gestures used are only effects or general shifts, and are rarely used for precise actions. Another use of gestural movement controlling audio manipulation can be found in the form of The 'Bodycoder' system. This interface, developed by Mark Bokowiec and Julie Wilson-Bokowiec is described by its creators as:

> 'a flexible sensor array worn on the body of a performer that sends data generated by movement to an MSP environment via radio. Movement data can be mapped in a variety of different ways to the live processing and manipulation of sound.'[7]

Similar to 'The Hands', the performer can use gestural movements to control actively

[5] Roads, C., 'John Chowning On Composition', in *Composers and the Computer* (Los Altos, CA, 1985) pp. 17-25

[6] Waisvisz, M., '1984 - 1989 The Hands (first version)' http://www.crackle.org/The%20Hands%201984.htm (Accessed 8th January 2010)

[7] Bokowiec, M., 'Bodycoder Interactive Performance' http://www.bodycoder.com/bodycoderprogram.html (Accessed 10th January 2010)

the source material (again, synthesized or recorded during the performance) in real time.

If this aspect of performance and interaction with sound can be successfully merged with satisfactory three-dimensional spatialization then hopefully this will help composers work more creatively in the multi-channel domain.

In January 2008 I attended a concert given by Canadian electroacoustic composer Gilles Gobeil. It was the first time I had attended a live diffusion performance in a multi-channel environment, and was thoroughly absorbed by the experience. However, at the time I was sitting in front of the desk, and did not see what was happening to create the sound movement I heard. It wasn't until I attended a masterclass the following day and tried the system for myself that I appreciated the sheer amount of performance that went into the concert. I resolved then to create a more physical, gestural, controller for spatialization so that the performer's actions could be seen and appreciated by the audience, in order to add an extra dimension to the event. Over the course of the next 3 months I created 'Diffuserguide', a practical system for simple gestural diffusion in a live performance[8]. Using orientation and bend sensors mounted on the body, the system allowed two functions within a circular horizontal 8-channel array. Under one function the user could change the level of a channel by reaching out and physically grabbing the sound and pulling it towards / pushing it away from themselves, and the other allowed the movement of two separate sounds independently around the space only by movement of the body.

In principle the system was good, though due to the nature of the sensors (for example the orient sensor relied on detection of the earth's magnetic field and so was prone to

[8] Humphries, C., *DiffuserGuide*, Interactive Sound Design 3 (BA (Hons) Music Technology, 2008)

interference), there were unmanageable elements that would have been problematic in the rigours of live performance.

In February 2009 a conference was held at Leeds Metropolitan University called Sonic Spatial Perspectives, which provided a forum for presentation and discussion of some of the most recent thinking on spatialization. The keynote speaker was Dr. Jonty Harrison, and his team from Birmingham University gave a demonstration of a 40 channel version of their 'BEAST' sound spatialization system. The pieces performed, mostly by their composers, used the versatility of the system to demonstrate a variety of different approaches to using spatialization. A roundtable discussion provided an opportunity to explore issues in this area, some of which proved relevant to this project. I was interested in what the forum members thought of the visual aspects of my research. Presenting this question to the members of the panel, Dr Harrison was opposed to the idea of including a visual aspect to spatialization in a live performance environment. Dr Harrison felt that the responsibility of the sound diffuser was to manipulate the audio and let the sound be the main focus, and that a visual element would change the intended effect of the performance. However, it was concluded as a group that such a system may well have great merit as a system within itself, with pieces, visual or sound art composed or created especially for it, but probably not to change what is known as diffusion today.

This reaction primarily encouraged me to think that I might be able to create this new medium for performance, perhaps as a combination of acting, musical performance, diffusion and performance art. However as this would be a remarkably different project from the one that I set out to research, I decided not to pursue it in this instance. Armed with these reactions, it did make me look at the subject in a different way, and played a big part in the decision of changing the research focus to studio performance, rather than live performance to an audience.

3. Methodology

Many of the types of controllers that have been investigated during the early part of this project can be, and have in the past been, used to capture the gestures or movements of instrumental or dance-based performers, and translated to be used to achieve a musical / spatial change in real time. Whilst this is an interesting and often effective use of the technology, they are most commonly used for wide, imprecise gestures that show simply a change from one state to another (be it sonic, spatial, or effect-based). However, this project is focusing more on what precise values can be measured using these devices, so that the user can control parameters accurately in space rather than having to do so on a computer. This self-imposed limitation does both restrict choice of sensors to use, and also makes it difficult to achieve the project's somewhat ambitious goals on a limited budget, as the most accurate sensors are likely to be the most well made, and so the most expensive. However, a compromise has been achieved in this area, and devices have been created that are prototypes, to show the idea and principle involved, rather than fully functioning devices themselves.

This project started as a personal quest for me to find a more practical method for 3D sound spatialization than is currently offered, but grew into an objective study to ascertain whether others felt the same, and if an improved system could be developed that would be of use to a wider range of users.

As technology has advanced, even in the last 10-15 years, we have increasing access to many more technologies to make multi-channel compositions and installations more popular than ever. However, there are not yet the tools in place to work to the full creative potential of the medium in an intuitive way. There is still a lot of thought that has to go into diffusing sound using faders, and the mouse and keyboard combination to which we are now accustomed is really a two dimensional controller at best. Can we find a better way to control sound movement in three dimensions?

I opted for gestural control as a research area because it seems the most intuitive for composers to use. Most of the time if there's a movement you want a sound to make in a space you can demonstrate it using your hands. If that movement could be captured, interpreted and replicated by the computer, that would hide any significant calculations or non-compositional elements from the user, thereby increasing productivity and allowing the technology to once again be an aid to the compositional process rather than a hindrance.

A decision was taken early on to focus on manipulation of point sources rather than stereo or whole soundfields. This was partly for simplicity, to try and get the interfaces working well with a single channel before adding the complexity of additional channels. It was also felt that using mono sources would allow clearer localization of sounds and therefore be the most useful for evaluating and comparing different interfaces.

The goal of this project was to investigate the control of spatialization, not the algorithms themselves. It was therefore decided to use standard spatialization software for the underlying spatial processing.

The basis of the spatialization algorithm that I used was the 'Spatialisateur' version 3.4 (hereafter referred to as 'spat~'). I chose this software over others primarily because of its flexibility and modular format, but also because it is implemented in the Max/MSP environment, allowing for maximum integration into the rest of the project. Spat~ was developed by IRCAM in collaboration with Espaces Nouveaux in order to 'propose a virtual acoustics processor which allows composers, performers or sound engineers to control the diffusion of sounds in a real or virtual space'[9]. For the purposes of this project, using four spat~ objects within Max/MSP allowed for the spatialization of four sounds independently of each other on a horizontal plane of 8 speakers.

However, to create movement in three dimensions, a variation of this patch was used, as adapted by Prof. Michael Clarke in 2006 for his 'Enmeshed' series of works[10]. This used the spat~ object several times for the spatialization, but added a height control to pan effectively the horizontal spatialization vertically between three planes of speakers, offering the use of 24 channels of audio, and adding the third dimension to the sound source. This adaptation was once again adapted for this project to be used for 16 channels in two octophonic planes. It is worth noting that since the research element of this project has been completed, a more recent version (4.0) of spat has been released, which may or may not be compatible with the software created for this project.

Much of the practical and experimental research was undertaken in the newly built SRIF-funded room, SPIRAL (Spatialization and Interactive Research Lab) at the University of Huddersfield. SPIRAL is a purpose built sound research studio containing 25 Genelec speakers, set out in 3 horizontal octophonic arrays (with an

[9] 'The Spatialisateur Project Overview', *IRCAM online*, http://support.ircam.fr/forum-ol-doc/spat/3.0/spat-3-intro/co/overview.html (Accessed 8th November 2009)
[10] Clarke, J.M., 'Enmeshed: live in 3D fog~', In *Proceedings of the International Computer Music Conference 2006,* pp. 13-16. (New Orleans, Louisiana: International Computer Music Association, 2006)

additional loudspeaker also suspended centrally from the ceiling), designed for 'research into spatial and interactive music composition, production, and sensoral control'[11].

As there are three horizontal circles of eight speakers, the software could have been adapted to incorporate the middle row also. However, it was decided that for the purposes of these tests two levels of audio would be sufficient because of the way the human ear determines position of sound, the vertical element is one of the least sensitive[12]. So whereas the change from one plane to two would be quite discernable, the change from two to three would be much less perceptible.

The idea of the project would be that a number of alternative interfaces were created as a result of experimentation, which would then be adapted and set up to function using a common parent patch. In order to evaluate this research and bring interfaces to some of its potential users, it was decided that once the research was complete to invite a number of people who may have an interest in the technology to put the practical elements through their paces and assess performance.

4. Research & Experimentation

The purpose of this project is to try and ascertain if there is a more efficient, ergonomic and user-friendly way to control the spatialization of sound in three dimensions. The aim was to cut out as many technical steps as possible from the composition process, allowing for the maximum creativity to productivity ratio. Therefore whatever device was created had to be simple, both to setup and whilst in use.

---

[11] Mac, P., 'Right Round, Baby – An Immersive Multi-Channel Experiment At Huddersfield Uni' *Audio Media*, ed. Marshall, L. (June 2009, Issue 223, p. 16)

[12] Roffler, S.K. and Butler, R.A., *Factors that influence the localization of sound in the vertical plane*, Journal of the Acoustical Society of America, 43(6): (pp. 1255-59) (1968)

In order to achieve this, the process must utilise the simplest form of communicating a movement, i.e. a gesture. If a performer wishes to move a sound from speaker X to speaker Y over a period of time, there is no clearer way of expressing that motion than by actually pointing. This principle led to the first focus to be on the idea of motion capture technology; using cameras to detect sensors worn on the body of the user, with the physical position then being correlated to the same position in the virtual environment and positioning the sound in that space. If that virtual environment could be re-mapped back into the physical world (in sound) by careful calibration of the speakers, then the user would effectively be able to move and manipulate a sound in real time using their hands.

In hindsight, the initial goals were ambitious; to track three sensors on the body, one on each hand and one on the head for reference. After some investigation into motion capture techniques, it emerged that the standard method for motion capture is by using powerful infra-red emitters and reflectors on the body, combined with infra-red cameras to record and track the motion[13]. It was decided that this was not feasible as the three independent points needing to be tracked would frequently be overlapping and there would be no way of differentiating between them. Also putting crosses around the room would be wildly impractical and against the fundamental goal of simplicity. To solve this problem I opted instead to use something further up the spectrum: visible light. The theory was that if there were three differently coloured Light Emitting Diodes (LEDs), and they were different enough (the primary colors, red green and blue), then the unwanted frequencies could be digitally filtered out from each RGB feed and track a single LED. There was brief experimentation with a

[13] 'Human Mocap', *Xsens*, http://www.xsens.com/en/company-pages/company/human-mocap (Accessed 31st December 2009)

cheap, homemade version of this, as shown by Johnny Chung Lee formerly of Carnegie Mellon University (now a Microsoft researcher)[14] with his Wiimote Whiteboard demonstration. Though it was decided not to use it for this portion of the project, the method did form the basis of another idea later on.

Also, in order to track the path of a point in three dimensions, at least three cameras are needed. If Camera A is placed in front, Camera B to the side and Camera C above (Figure 4.1), then each axis of movement is covered twice (A: X&Y, B: Z&Y, C: X&Z) where X is horizontal movement from the user's point of view, Y is vertical and Z is forward and backward (Figure 4.2).

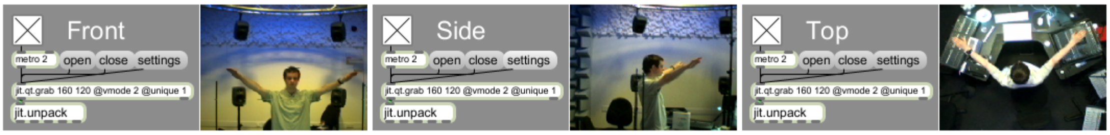


Figure 4.1: Part of the patch that shows the camera inputs from each location.

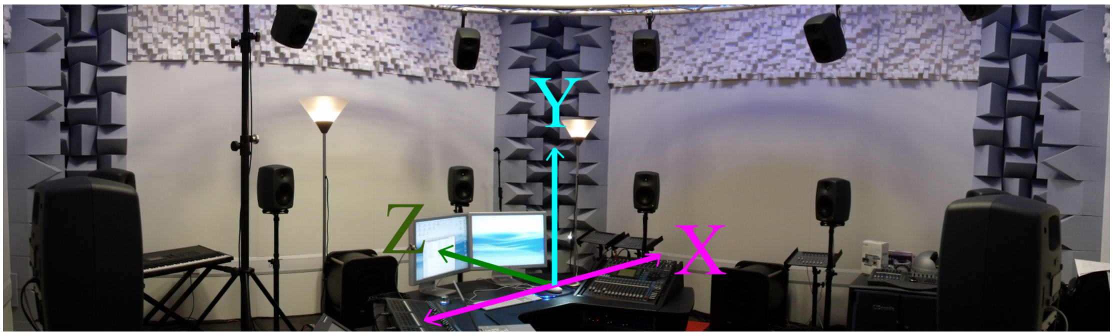


Figure 4.2: Location of axes in relation to the SPIRAL layout

After some initial testing the system had potential, so three cameras were installed and a patch was created in Max/MSP that would interpret the data. Apple's iSight cameras were chosen as the capture devices, partly because they have excellent resolution quality for the price and because they connect using a firewire interface, meaning that feeding three live feeds into the computer would be relatively straightforward. Also,

[14] Chung Lee, J., 'Johnny Chung Lee – Human Computer Interaction Research', http://johnnylee.net (Accessed 24th November 2009)

the use of commercially available 'off-the-shelf' hardware would make this system easier for others pursuing similar research to replicate. For the tracking algorithm, Jean-Marc Pelletier's 'cv.jit' collection of video interpretation objects for Max was used, trying to find a compromise between distance and detection accuracy.

The process for tracking is this: Take the three inputs from the cameras and split each one into its component elements (RGB) using 'jit.unpack'. Each of these is then converted into a binary image using 'jit.op', thereby isolating each color into a black and white, positive and negative simplified feed. The user can at this point adjust several arguments to obtain the most clear separation between the sources. Next, these outputs are fed into contrast detection algorithms, 'cv.jit.features', that shows a red cross at the point of most significance in the image, updating in real time. At this point the user presses a button (either on screen or using a mounted switch or foot pedal) to confirm that the tracker has selected the most salient point. This then sends the position of the point through to a tracking algorithm 'cv.jit.track' that recognizes patterns in movement of video feeds and will follow the point as it moves. Using 'jit.iter', this information can then be converted into standard integers for conversion and manipulation within Max. Here this involves viewing the points proportionally on a multislider and sending the resultant 9 scaled X and Y outputs to another patch.

If there are three camera feeds, each producing red green and blue and each of these giving X and Y values, there will be 18 values changing in real time from which to average out and calculate the XYZ values (and hence exact position in space) of all three points. By putting X and Y into a 'cartopol' object, a value can be created for both angle and distance from the center of the circle, which can then be sent to spat~ to reproduce the spatialization in very nearly real time. (Figure 4.3).

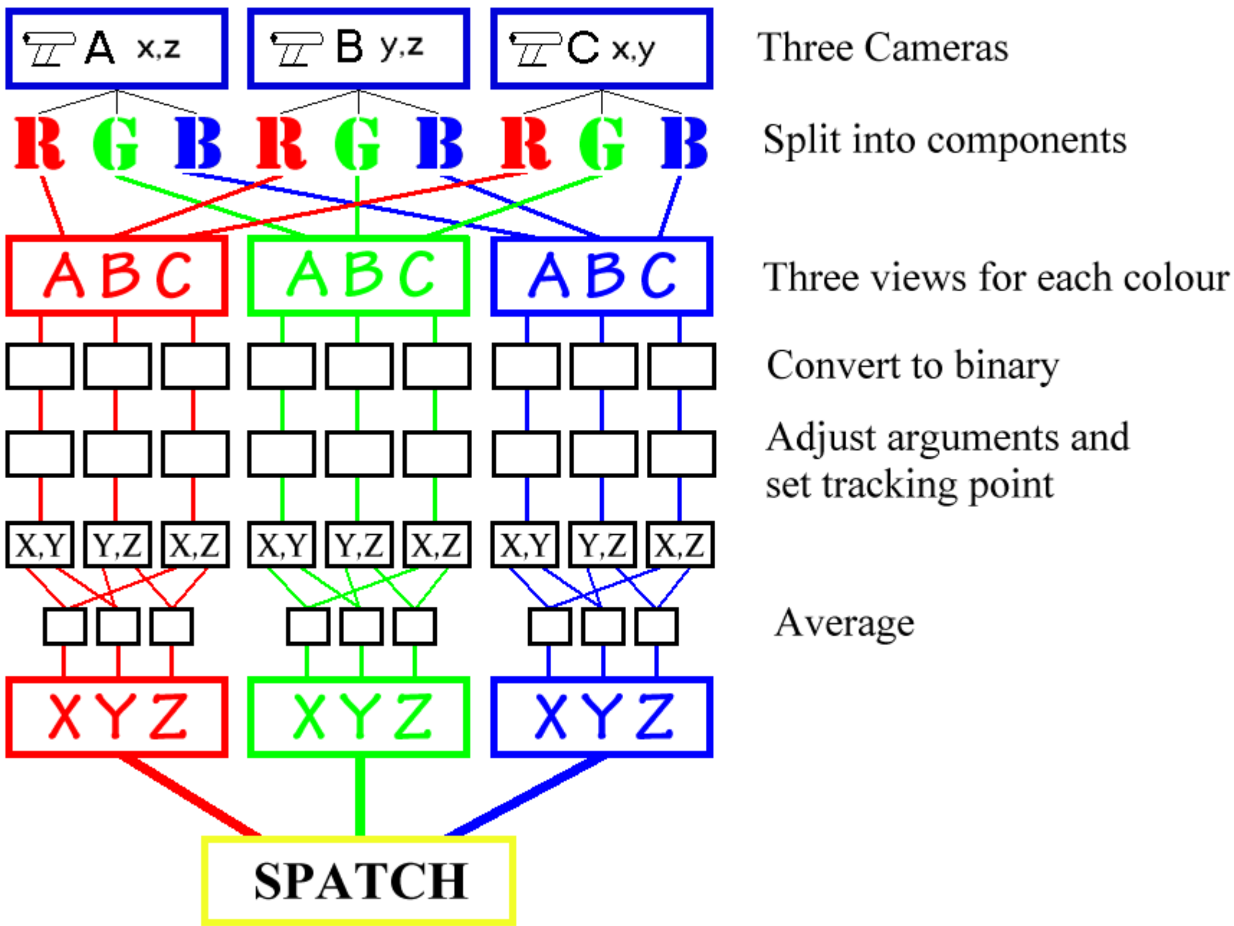


Figure 4.3: Flow chart showing how three sets of XYZ co-ordinates are determined.

Through use of and experimentation with this system, certain issues became apparent. The difficulty in using visible light is that there are many distractions that can confuse the detection and tracking systems, even the glow from the monitor in front of the user. One attempted solution was simply to turn off the lights to make the LEDs more dominant in the room, but being commercial webcams, the iSight cameras have an irremovable built in setting for adaptation to dark conditions, which diminishes the quality and therefore the effectiveness of the tracking. The LEDs, when pointed at a certain angle to the cameras at close range tracked very well, but when moved away from this optimum angle would disappear, or when pointed directly would flood the camera with light, both of which would confuse the tracking software. This issue with brightness of the source varying with movement was prominent in all situations, whether the source was an LED or reflected light. After much experimentation and variation, it was decided that the external distractions were still too influential on the tracking system to be truly reliable enough for the purposes of this project. If tested as a potential solution for spatialization, it would be likely to frustrate users much more than aid them in composition and this would not be a fair test of the principle of gestural control. However, future research with alternative sensors in a different environment may yet create a reliable system with increased accuracy; the idea still has potential.

In terms of implementation, incorporating gesture into spatialization can be looked at in three ways:

i) from within,

ii) in proportional space,

iii) from outside (in terms of where the user is in relation to the physical sonic space he/she is controlling).

Experiments with direct proportionality of movement (the sound being replicated in exactly the same place as the hand) showed this to be impractical, as it would require a lot of movement in the physical space to produce few perceptible results. Also, during the performance the user would only ever hear the sound within the radius of their arms as the sound would be moving with them. In all cases the user would see a direct correlation between their movements and the movement of the sound source, but in different proportions. Being ‘inside’ the space allows for small gestures to give large results, which is useful for performing fast or large movements. Being ‘outside’ the space facilitates wider gestures, perhaps some of those that are impractical to perform in a bigger way, like spinning sounds or spirals. Alternatively the user could be allowed to navigate through the space by using a different controller and representing its position completely virtually on the screen.

Once the research was completed and potential users were to be brought in to evaluate the research, it was felt that the best way to achieve useful results was to give them a choice of interfaces to test, demonstrating the use of some of these environments. Ultimately then, four interfaces were developed, that would work with a largely uniform spatialization patch, the ‘Spatch’.

5. The Software

The main part of the ‘Spatch’ is the housing for most of the important sub-patchers and is the control panel for most of the patch. It contains the audio inputs, sensor inputs, automation (in and out), audio outputs and the spatialization patches themselves are also all contained within.
It was decided early on that it would be prudent to create every element of the programming for this project in a modular way, that is to keep all elements of the patch separate, in different files, linked together by sends and receives or by inlets and outlets. This choice was made for several reasons. Firstly because it would make the patches easier to troubleshoot, being able to narrow down problems based on which sub-patches worked, rather than trying to work though one large patch. Also, being a project that is based around one set of processes but with different interfaces, it would be much easier to control if there were some elements that were constant and could be added in by using objects rather than copying and pasting (and so probably creating accidental subtle differences between interfaces). Thirdly, the hope is now that the project is completed, that work will continue to develop further one or more of the

interfaces or areas that have been studied and parts of the resultant patches can be used again. Even if this is not the case, there may be some utilities that could be useful for future research in SPIRAL at the University of Huddersfield.

5.1. Visual Representation

Another of the key goals for this project was to create a means to represent visually where sound was in the room, rather than relying solely on the ears to convey this information. This is because due to the nature of human hearing, some sounds or frequencies are much harder to determine the spatial position of than others. A piece of OpenGL programming originally created by Zachary Seldess[15], was adapted for this purpose. He had created a simple 6 plane cube in a space, with the option for manoeuvring the space using certain keyboard commands. Taking this as a starting point, the planes became virtual walls, blocks were added to represent furniture (as fixed point references in the room), as well as the addition of four movable balls to represent the sounds, which were then mapped to the outputs from each of the relevant interfaces to show precisely where in the 3D environment the sounds had been placed. Altering the virtual camera view parameters to be controlled by an external device allowed for more intuitive navigation. After experimentation with a number of interfaces, including joysticks, button-based devices and haptic controllers, one stood out, a device created by 3D Connexion called the Space Navigator. As the name might imply, this seems an ideal controller in this context - something that enables 3D navigation in space. The device is essentially a joystick, but with more degrees of freedom. It can pan in 4 directions, tilt in the same 4 directions (but shows a separate value), a 'zoom' up and down, a twist function, as well as two buttons. This

[15] Seldess, Z., 'OpenGL 3D Flight simulators' *Zachary Seldess.com*, http://www.zacharyseldess.com/z.glNav.html (Accessed 12th December 2009)

multitude of functions seemed to be a perfect way to navigate through the virtual environment. This is one of the project's greater successes, both in terms of practicality and in terms of achieving the objective. The next step in this part of the research would be to place this virtual environment into a pair of video glasses (so as to immerse completely the user in the environment) and attach a sensor that detects the user's head movement and compensates for this, thereby allowing the user to look around a virtual environment that is exactly proportional to the space that the user is in, but digitally indicating the visual positions of the sounds they are spatializing, seemingly in the air around them.

Because all of the project's interfaces would be restricted to manipulating one sound at a time, some element of automation was required to make the interfaces realistically usable and efficient. Max/MSP offers several ways of performing this function within the program itself like seq and mtr, but these are basic and limited and have inherent limitations that would hinder the desired simplicity element. Neither has a timeline view for automation, which firstly makes synchronization with the audio very difficult and secondly does not have an easy way to show non-realtime viewing of what has been automated, nor do they allow for any sort of breakpoint editing. In the same way as the spat~ is being used for spatialization algorithms, it makes sense to use existing software that is designed for this purpose, in this case Steinberg's Nuendo sequencing package. Using Cycling '74's audio routing system extension Soundflower, it is possible to have the sound source originate from Nuendo. This in itself is a significant advantage as in theory an entire project could be playing, with effects, EQ, editing etc. within the program, as long as the outputs are routed to 4 separate channels of Soundflower. Also though, it means that the data from the sensors, once converted into integers, can be sent out as midi data, fed back into

Nuendo and be recorded in time with the audio. This means that the user can view and edit the spatialization data which is directly adjacent to the audio and then save it along with the project. This automation data can then also be played back along with the project, at the same time as recording new data in. The controls ‘play’, ‘stop’ and ‘record’ for automation can also be mapped into Max, so that the user will never have to switch between programs whilst working. In order to keep the automation data as closely linked to the audio as possible, there has been included a simple VST plugin (built using Max v.4.5 for plugin export compatibility[16]) containing labeled parameters for Azimuth, Distance and Height. This will not change the sound in any way, but having them as parameters means that they can be allocated via the ‘Remote Control’ function of Nuendo to be allocated for control by an external device (in this case Max/MSP and the chosen interface).

In the current setup there is some preparation to be done to allow all of this to happen, but I am confident that it is possible to automate a large part of the setup procedure and make it a truly realistic option for automation of spatialization in the future.

Also in Spatch there is the main volume control, audio on/off switch, a link to the output controls (routing to the speaker outputs), as well as graphically laid out level meters for each speaker. For ease of opening, the spat~ objects are integrated into Spatch (even though it is in its own patcher), so reducing setup time.

Because sound is routed through Soundflower, the audio inputs are also incorporated directly into the spat~ patch.

As illustrated in Figure 5.1, when the sensor information comes into Max, it is sent first to Spatch, which then routes it either directly to spat~, or to be recorded as automation data in Nuendo (which then goes out and returns to go into spat~ to perform the spatialization). It also gets sent to the graphic display where it is scaled

[16] Dobrian, C. et al., ‘MSP Reference Manual’ (Cycling ’74, 2001) p. 396

and controls the movement of the balls. The output of the spat~ gets sent out of 16 outputs to the speakers, as well as to the meters for reference. As mentioned before, these 16 speakers are arranged in two octophonic circles at two heights, which creates the 3D element of the spatialization.

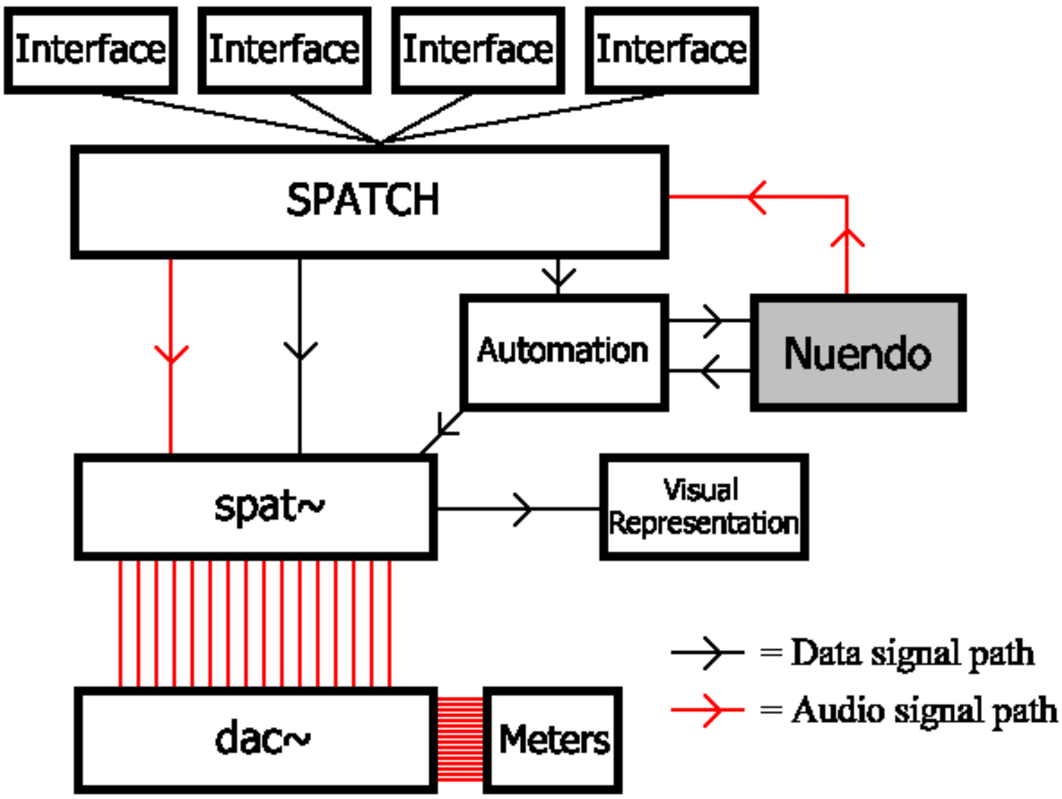


Figure 5.1: A flow chart of how the various patches connect to each other.

6. The Interfaces:

In Spatch there is the facility for four sounds to be controlled and spatialized separately and simultaneously. These correspond to four channels output from the sequencer and are routed through to four separate 'spat~' objects. The control for the spatialization of these sounds originates from the individual interface patches, which are all self contained so that only one can be in use at one time. The outputs from these patches however, are all the same. For most there are three outputs (for Azimuth, Distance and Height) along with a number between one and four to select which sound to control. Alternatively there could be an output for each parameter of each sound (e.g. Sens1Az, Sens2Az, Sens1Dist, Sens2Dist). These are sent to 'Spatch' which routes them through to automation, visual representation, or the 'spat~' objects.

A range of devices were used to test the different approaches researched. In some cases these were standard commercial devices, but in other cases in order to achieve the desired approach, prototypes needed to be built.

6.1. P5 Glove

Through researching and trialing spatialization using camera tracking, the conclusion was reached that it did not fulfill the criteria expected of it and was therefore not a viable option to compare against alternatives for the purposes of this project.

It was therefore decided fairly late on in the project to remove it from the testing as it would not have been fair to represent the method in that way (it would have affected opinion because of the unstable interface rather than because of the method itself).

However, it was felt that it was still important to have that sort of gestural control represented in the testing, so a replacement device was brought in to use as a representative of that approach. This took the form of a pre-existing piece of hardware, Essential Reality’s P5 Glove (Figure 6.1.1 & 6.1.2), which is a virtual reality device developed as a gaming controller, but which can also be used independently as a cursor controller. It contains 5 individual finger bend-sensors, 3 customizable buttons on the back of the hand and most importantly, 6 degrees of tracking (X, Y, Z, yaw, pitch and roll) as a result of infra-red emitters and a receiver.

This device was initially dismissed in the original experiments because of its reliance on the receptor, needing to be within 3 feet of it and in line of sight at all times. It was rejected during the search for controllers for proportional movement, but as a replacement for a controller outside the space it still had potential. To use this controller, the user must imagine an invisible box in front of them around 3ft cubed, a scaled down representation of the room essentially, where the user's hand would represent the position of the sound. This then stays within the limitations of the receptor and also allows the user to stay within view of the monitor while moving their hands. The other sensors on the glove also provided some useful controls. The buttons were used as selectors for which sound to control and the bend sensors were used as a simple switch to 'pick up' the relevant sound.

When the device is turned on, the user sees the position of the glove in the virtual space on the screen as a grey ring. Once the user selects which sound they wish to control using the buttons (the selected ball starts flashing), they can navigate to it and, by clenching their fist, pick up the sound to move it to a new location. Unclenching the fist will 'drop' the sound in its new location, while the ring can move away to pick up another sound. This way the user always knows where their hand is in relation to the sounds, allowing them to move to where the sound is situated to avoid unnecessary audible jumps in position when the sound is picked up.

The P5 Glove has a USB interface, but was not detected by the human interface (hi) object in Max. Fortunately, 'p5osc', a command line application, had already been created by Tim Kreger to turn the glove's messages into Open Sound Control data[17]. This allowed me to use the User Datagram Protocol as an input to Max and map the resultant outputs to the necessary parameters for use in Spatch.

[17] 'Simulus – P5 Glove Developments', *Simulus*, http://www.simulus.org/p5glove (Accessed 18th August 2009)

Figure 6.1.1: The P5 Glove

Figure 6.1.2: The P5 Glove in use

6.2. SpaceNav

For movement around the visual representation virtual environment it was desirable to avoid using the keyboard and mouse if possible, as that was the established barrier in the first place. The flexibility and reliability of the Space Navigator (Figure 6.2.1 & 6.2.2) being used for controlling the Visual Representation eventually earned it a place as an interface in its own right, as a separate controller for people to test. Using 'hi' as an interface into Max the 8 values that it produced could be then converted into values that could control Spatch. One challenge that presented itself was that the data that was displayed for movement was between positive and negative extremes, with 0 as the normal point. This meant that in order to create an XYZ position, the data would be constantly adding or subtracting to a value and would then always be affected by its previous value. However, once this problem was solved, it produced a useful side effect in that the further the user pushed the controller in a certain direction, the faster it would move that way. This allows for both fast movements by using harder gestures, or very precise miniscule movements by using light touches. The Space Navigator was one of the most reliable of all of the interfaces, mostly because it was a commercially built all-encompassing unit of which the user has no way to affect the reliability. Despite it being my personal preference of all of the interfaces, most users who tested it did not like it as much as the others because they felt it was too similar to the existing option of a keyboard and mouse. This is useful because it reinforces the original hypothesis of a keyboard and mouse not being the best option in the first place and also because it gives greater perspective on the other devices and shows the difference between them more in the same context (i.e. by using the same main patch, Spatch, in the same environment with a different interface).

Figure 6.2.1: The SpaceNav

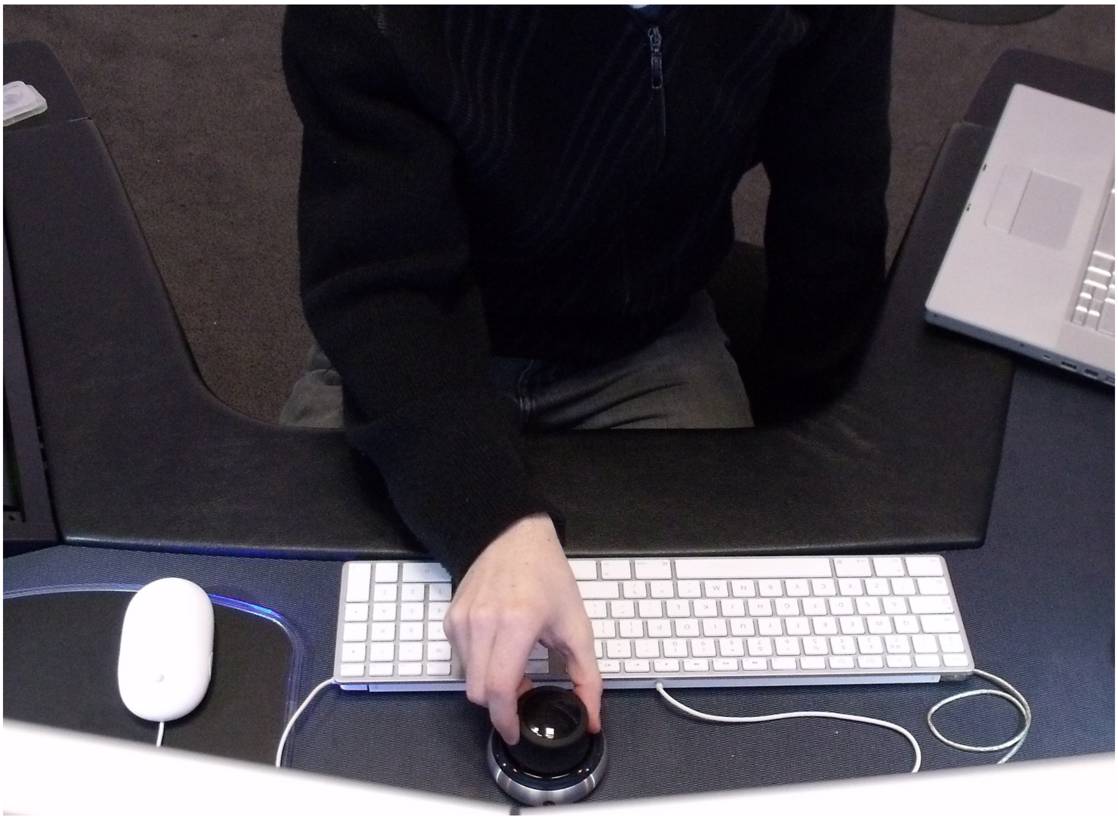


Figure 6.2.2: The SpaceNav in use

6.3. Wheel-e-bin

In addition to commercially available devices, it was found important in order to facilitate exploration of certain strategies to build prototypes of new interfaces. This gave rise to the 'Wheel-e-bin' (Figure 6.3.1 & 6.3.2). It was made from a collection of various materials, mainly because there did not seem to be anything existing that gave the desired result. It was also the culmination of much research into how to determine the heading of an object. The goal was to be able to position an object in a certain direction and on screen have a visual representation of exactly the direction (in 3D) it was pointing. This proved very difficult, even with a device designed specifically for this purpose. Infusion Systems' Orient 3D claims that 'With the data retrieved from an Orient 3D, heading, pitch and roll can be computed'[18]. Whist it does fulfill its brief, it did not serve to produce the required values needed accurately enough. So in trying to ascertain what exactly was desired, the problem was broken down into its component parts.

There were three requirements:

a) something to determine circular position on a horizontal plane (much like a compass),
b) something to determine height and
c) a third sensor to detect distance from a center point.

In the other interfaces these three variables were generally produced from the same device, but that did not mean that individual component devices would not work just as well, even if they did require a little more configuration to align them. The Orient 3D was not reliable enough for my purposes, however, its predecessor, the Orient 2D,

[18] 'Orient 3D – Manual', *Infusion Systems*, http://infusionsystems.com/support/Orient3D-manual.pdf (p.3) (Accessed 17th July 2009)

which had served me well in 'Diffuserguide', was the next choice. From past experience it was known that the sensor would only function correctly when lying totally flat, so the preferable approach, mounting on the body, was not an option. Instead it was mounted on a freely moving circular plate – a 'Lazy Susan'. This way when configured correctly, the direction in which the wheel was pointing could be mapped onto the screen. To provide the height values it was desirable to perform the physical movement of rising from the board. One way to do this whilst still maintaining manoeuvrability, was to create a rod pivoting at a point in the centre of the board and measure the variation in angle as the rod was raised. Placing a bend sensor on this hinge enabled the bend information to be received via the iCubeX digitizer. All that remained was the distance information. Now that there was a rod coming out of the centre of the board, a useful position was created to mount a 'push sensor' (essentially a stripped down fader), which would read distance information.
To complete the interface, I wanted some sort of dome to hide the electronics and give the user something to hold whilst moving it. The roll top of a waste paper bin fitted the brief and gave the interface its name. Using these 3 sensors from Infusion Systems and their iCubeX digitizer as an interface, there was everything necessary to feed all the data into Max necessary to position a sound in 3D.

The premise behind this device was that it could be used for fast movements like spins or spirals, something that in my opinion the other interfaces could not provide adequately. It was designed to be completely 'outside the environment', so the user could move the sounds around by focusing on something tangible in front of them, rather than manipulating a virtual environment. It was to bring the relationship between physical and sonic movement to as basic a level as possible, whilst still allowing for complex movements to be made.

The primary issue that presented itself with this interface was, that being a computer based system, it had to be plugged into the computer in some way, both for power and for the actual data transfer. This would certainly hamper the ability to move the device, especially for any sort of fast or spinning actions. This was overcome using a device built for the iCube (which itself was mounted inside the container) called Wi-Mi. This is a wireless transmitter for MIDI data, more specifically iCubeX data, using an early form of Bluetooth technology[19]. The iCubeX was powered by a set of AA batteries bundled together in series to produce the necessary 7.5 Volts for powering the digitizer. This gave the Wheel-e-bin complete wire independence and greatly enhanced its potential as an interface.

[19] 'I-CubeX Online Store – Wi-MI : Wireless MIDI interface' *Infusion Systems*, http://infusionsystems.com/catalog/product_info.php/products_id/57 (Accessed 1st January 2010)

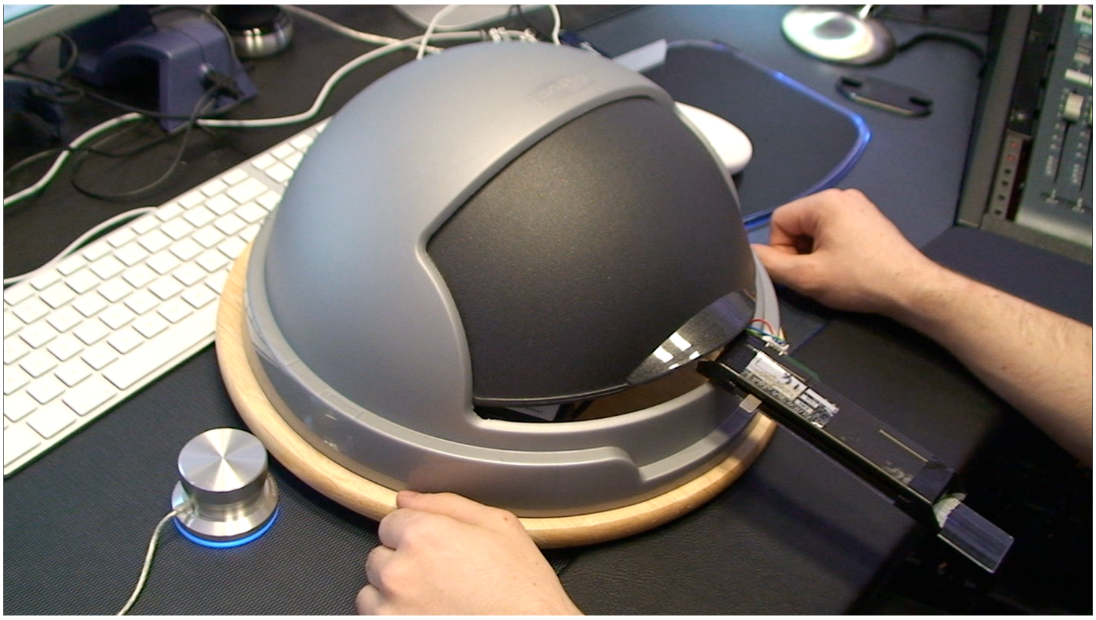


Figure 6.3.1: The Wheel-e-bin

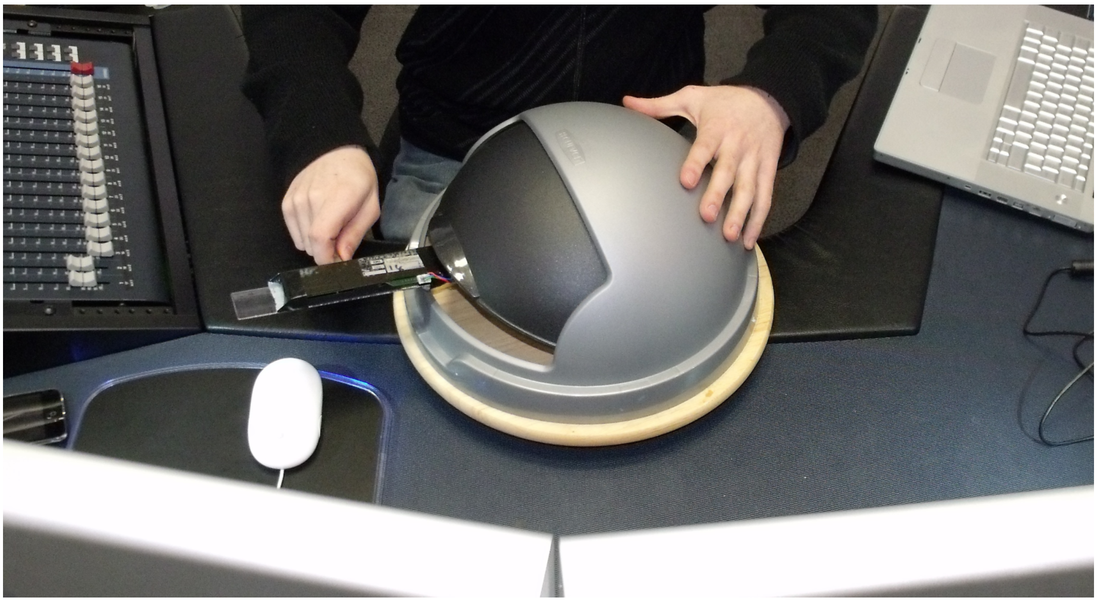


Figure 6.3.2: The Wheel-e-bin in use

6.4. Screenspat

Touch screen computers and multitouch surfaces are very popular in the current market, with products like Apple's iPhone and iPod Touch[20] and the growing range of multitouch smartphones bringing touchscreen interfaces to the general public. Though touchscreen computers do exist (e.g. ASUS Eee Top[21], HP Touchsmart[22], Microsoft Surface[23]), they have yet to become particularly mainstream and so have not been seen as a viable option for sound spatialization. Even those interfaces that are created more for use with audio applications (e.g. JazzMutant Lemur[24]), have a small screen and I personally find them restrictive in their graphics. I wanted to create something large enough to be easily viewed, did not have to be actually touched and that would provide a clear link between gesture and sonic movement.

As mentioned earlier, whilst researching infra-red detection for motion tracking a method was discovered to make a cheap multitouch whiteboard using the Nintendo Wii remote controller, or 'wiimote' as an interface[25]. This was of interest for several reasons, not least because the cost of materials was relatively small compared to the incredible functionality and potential applications for the technology. The process involved using some custom software to hack the wiimote to track infra-red LEDs and use this data to control the computer's cursor. This meant that the user could, by creating a simple circuit including an infra-red LED, control their computer when projected on any surface.

---

[20] 'iPhone – Mobile phone, iPod and Internet device', *Apple (United Kingdom)*, http://www.apple.com/uk/iphone (Accessed 31st December 2009)

[21] 'Asus – Eee Top', *Asus,* http://event.asus.com/eeetop (Accessed 31st December 2009)

[22] 'HP TouchSmart PCs', *HP*, http://www.hp.com/united-states/campaigns/touchsmart (Accessed 31st December 2009)

[23] 'Microsoft – Surface', *Microsoft*, http://www.microsoft.com/surface (Accessed 31st December 2009

[24] 'JazzMutant – Lemur, multitouch modular controller for sequesncers, synthesizers, virtual instruments, vj and light', *Jazzmutant*, http://www.jazzmutant.com/lemur_overview.php (Accessed 31st December 2009)

[25] Chung Lee, J., 'Projects – Wii', *Jonny Lee.net*, http://johnnylee.net/projects/wii (Accessed 1st January 2010)

As most people are familiar with the computer monitor as a visual interface, it seemed a logical choice to use this principle, but on a larger scale. When there are numerous controls on a screen it is often a very small mouse movement on the display that creates a large movement in sound. Sometimes this is desirable, but other times it can weaken the relationship between the action and the result. It was felt that enlarging the screen and placing it directly in front of the user to manipulate, using their hands, would help strengthen this link. I was also inspired by the futuristic computer screen interface portrayed in the 2002 film *Minority Report*[26] (Figure 6.4.1).

Figure 6.4.1: The virtual screen used in the film.

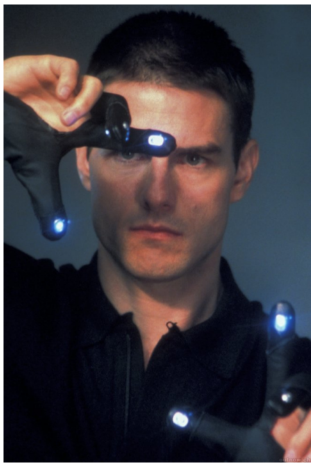


Figure 6.4.2: The *Minority Report* gloves

In the film, the premise was that the user would wear custom gloves (Figure 6.4.2) with three sensors of some sort and using a combination of gestures, control images on the glass in front of them, move them around, or start, stop and rewind video, all without touching anything.

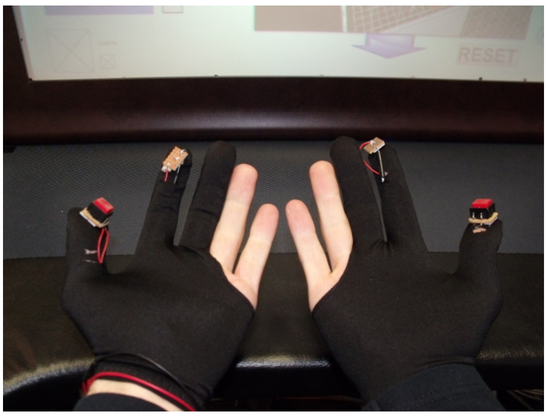
It is a very futuristic looking interface, but having seen Johnny Lee's wiimote tutorial, it is not really that far from being reality.

This interface clearly had great potential for use in sound spatialization. The first goal was to try and recreate a similar sort of movement as in the film, whilst also trying to

[26] Spielberg, S., dir., *Minority Report* (20th Century Fox, 2002)

remain practical. After trying several large screens, it was found that the only way to achieve the required breadth of movement was to use a projector and something larger than the screens at my disposal. As the infra-red detection worked best when the LED is pointing directly at the wiimote, the image was back projected onto a translucent screen that would allow the infra-red light through one way and the visible light through the other. In order to make this feasible within the available resources of the project, it was decided to develop a custom-built alternative. A large mirror was adapted by removing the mirror element from the frame and was replaced with some thick tracing paper. Once secured in place, it worked exactly as I had hoped and when an image was projected onto it there was very clear detail and no indication that it was not in fact designed for this very use (Figure 6.4.6 & 6.4.7).

The infra-red LED mounting stayed with the *Minority Report* theme, as can be seen from the design of the gloves (Figure 6.4.3 & 6.4.4). Starting with a pair of three-fingered billiard/snooker gloves, I cut holes and threaded though a custom LED and switch circuit (Figure 6.4.5) so that the LED was at the tip of the index finger and the switch was on the thumb (to be pressed by another finger, palm or side of index finger at the user’s discretion).



Figures 6.4.3 & **6.4.4**: The custom-made gloves that control Screenspat.

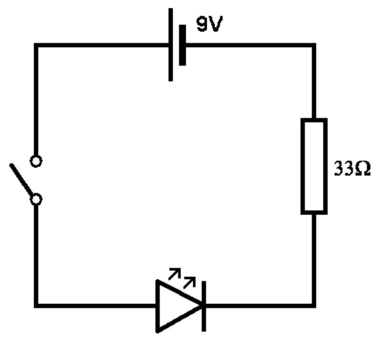


Figure 6.4.5: A diagram of the circuit used in the Screenspat gloves.

As the projected image was merely that of a computer screen, the interface had to be completely designed from scratch and due to its nature was quite different from the main Spatch and the other interfaces.

Using the interface as the main screen, it was necessary to have all required controls available on the screen at all times. This meant a slight redesign of the Spatch interface, whilst maintaining the controls that by this time the user had hopefully become familiar with. The actual controls for spatialization this time consisted of four 2D graphs (displayed on LCDs), with axes lines and a circle in the background to show the relative proximity of the speakers. This would show the horizontal position of the sound in space, much like on a surround sound panner found in a sequencer. Beside each of these is a slider that controls the height of the sound, the idea being that they could be used independently. With easy access to automation control, the user could record horizontal movement in one pass and then the vertical movement afterwards (which would also be displayed on the LCD in real time), although it is still possible to use both simultaneously. Also on the screen would be the visual representation of the room (this time the navigation being controlled by the virtual cursor / glove using clickable arrows and by dragging), as well as controls for playback start/stop, automation, overall level and metering.

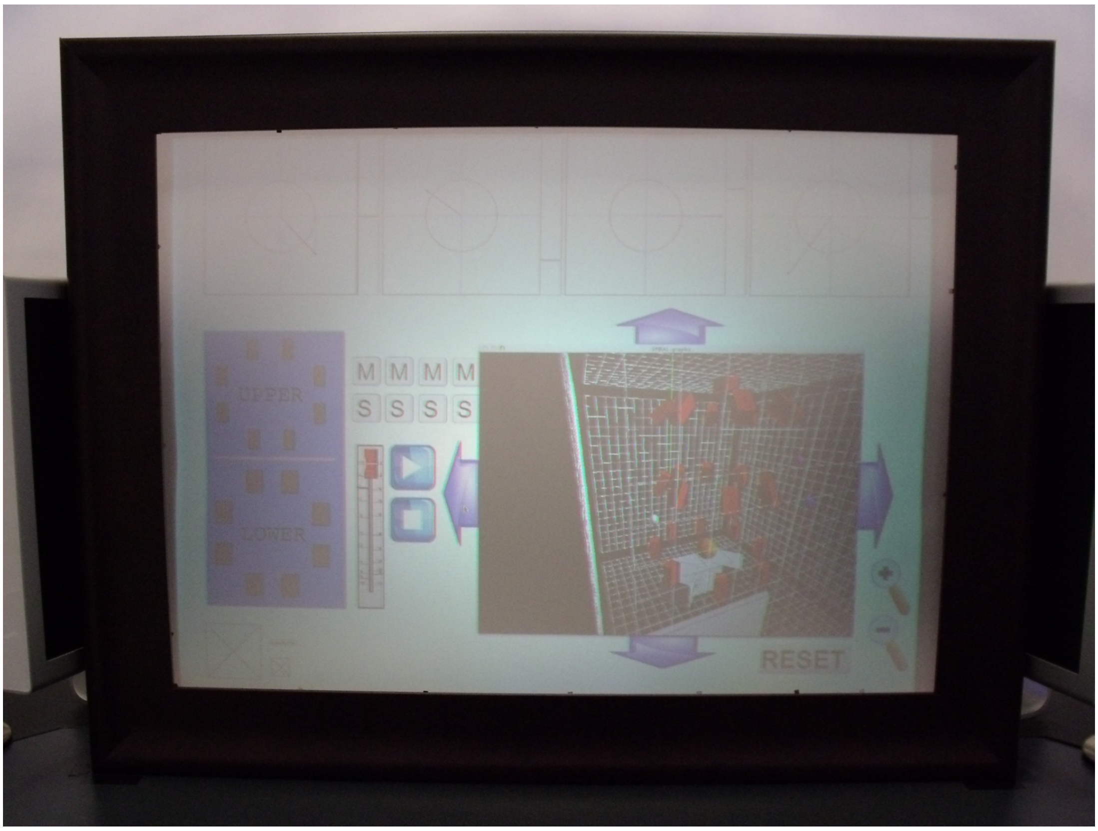


Figure 6.4.6: The custom projector screen, displaying the Screenspat interface.

Figure 6.4.7: A view from behind the screen showing the projector and Wii remote.

7. Testing:

Amongst those who were asked to come and test the interfaces were composers who had worked with 3D spatialization in the past as well as those that hadn't, a cross section of undergraduate students, postgraduate students and lecturers from the University.

Once the system was setup they were allowed approximately 15 minutes per interface to experiment. Though this is a relatively short space of time to become familiar with and assess a brand new interface, time constraints did not allow for any longer and it was hoped that this was still enough time to get an initial response and some useful feedback.

Each tester was offered the choice beforehand to bring in their own sound material to spatialize so that they would be familiar with the source material, but all of them opted to instead use a multitrack version of Stevie Wonder's [b. 1950] *Superstition* (1972) that had been used for test purposes during the research. Though this is not the intended sonic use for the system, I found that the clear definition between sounds was a great help in determining the effect of the spatialization and localization of the sounds, especially when sub-mixed into 4 channels: drums and bass, clavinet, horns and vocals. Also, it being a well known track meant that everyone was more or less familiar with it, making it more of a universally fair test than if they had all used different sources. Each user also tested the interfaces in a random order, in an attempt to vary comparisons between them.

Between testing interfaces they were asked to write a few words about the interfaces in categories of general comments, strengths and weaknesses and then an overall questionnaire after trying all of them. The following is a summary of the responses of the users. The complete questionnaires can be found in Appendix 10.4.

8. Results:

8.1. P5 Glove

Almost universally liked, the glove was generally found to be the most intuitive interface due to the direct kinaesonic[27] association with physical movement to sonic movement and general ease of use. The main concern was with the potential for RSI or arm ache with the need to maintain the position held horizontally in front. This may well be the case for periods of prolonged use, but the purpose of the 'grab' function was that when not controlling a sound the sensors are not connected and the arm can be lowered back to the mouse. One tester found the virtual nature of the device a hindrance as there was no physical feedback, only sonic and visual. The answer to this may be to add some sort of haptic feedback element to show where the boundaries of the room are, or where speakers are in relation to the sound. However, this addition would almost certainly add weight to the device, which is probably not desirable for something that is to be held at arms length for sustained periods.

One other issue that arose during the course of the tests was the compromise between speed of movement and sensitivity of the sensors. During development it was found that the infra-red sensor on the receiver is incredibly sensitive and is constantly adjusting the values being received from the glove. This means that there is very little delay between a gesture and its result on the screen, but when holding still there is the tendency for some jitter which when controlling spatialization can sometimes be audible. To combat this I introduced a smoothing control which averaged the incoming data. This made the static element much steadier, but had the adverse affect

[27] A term coined by interactive composer-performers Mark Bokowiec and Julie Wilson-Bokowiec, Kinaesonic is a term used to describe a relationship between bodily movement and sonic movement, in this case the direct mapping of gesture to the movement of a sound.
Wilson-Bokowiec, J. and Bokowiec, M., 'Kinaesonics: The intertwining relationship of body and sound', *Contemporary Music Review*, Vol. 25, Issue 1&2 (February 2006) (pp. 47-57).

of slowing down the response time for movements. Ultimately, rather than a compromise the decision was given to the user to control and give preference to one parameter at the expense of the other if they so desired.

8.2. Space Nav

The general consensus on the Space Navigator was that it was generally intuitive to use at first touch as it was clear how it worked, but would require a lot of getting used to before it could be used without conscious concentration. As with any interface, the more you use it, the better you will become. I personally favored this interface, probably because it was the first of the interfaces that worked fully, it was used most often and so familiarity breeds confidence. There was some concern about the reliance on the GUI for positioning of the sound precisely, and perhaps that depth perception was difficult with the wireframe graphical representation. One suggested solution was to have preset views for the wireframe changed by another controller and use the other Space Navigator to control an additional sound source. There was also concern regarding the speed at which the dots moved, even with the variable speed as a result of the strength of the gesture. This sensitivity is something that could easily be changed, or even given to the user to control on a sliding scale.

It seems that with practice and a few minor changes or additions to the programming, the Space Navigator could be a realistic option as a sound spatialization controller in the future.

8.3. Wheel-e-bin:

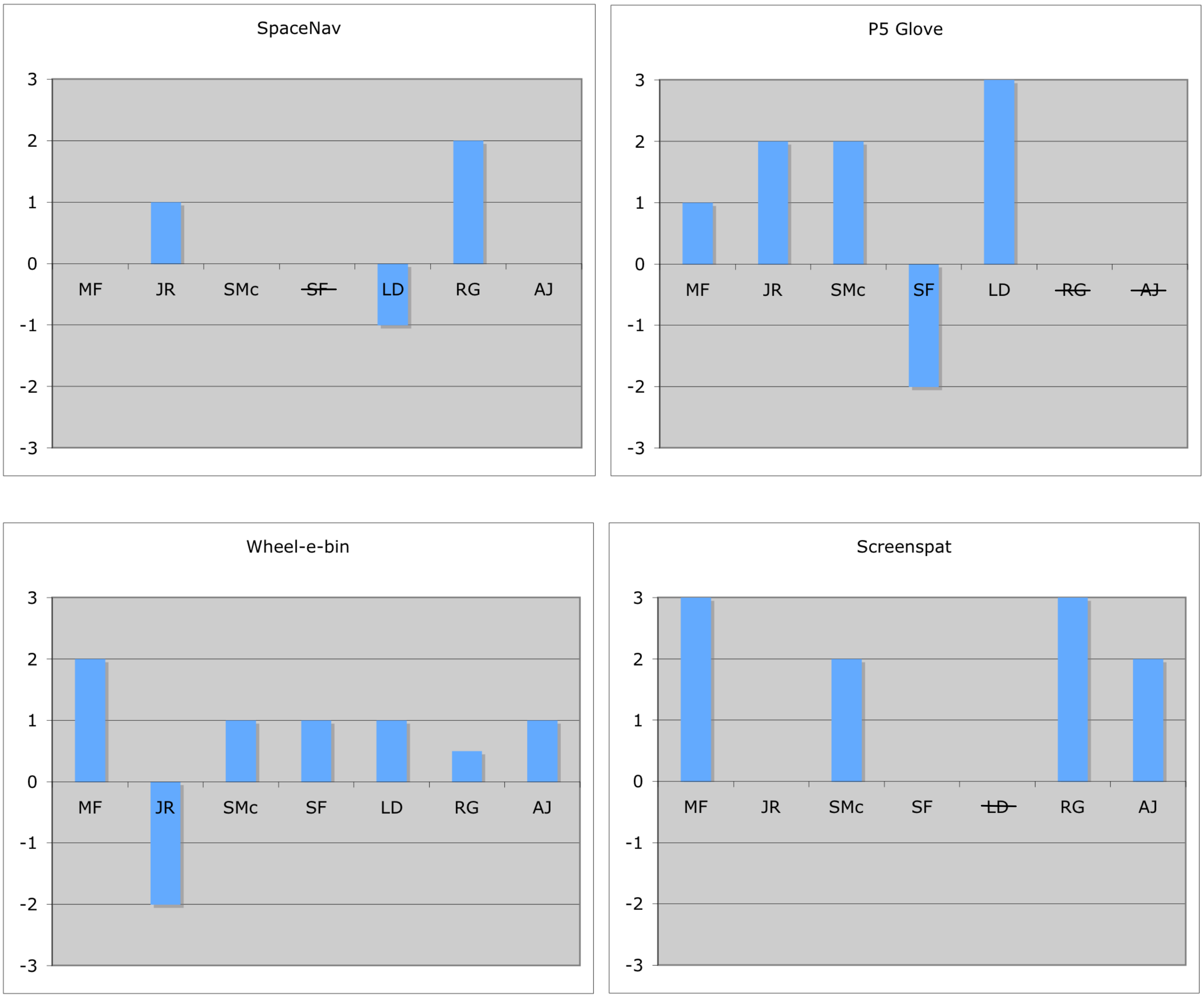
The Wheel-e-bin was surprisingly well received for such a rudimentary design. The testers liked its very intuitive nature, the physicality of the gestures and the direct relationship between gesture and sonic movement. Its popularity may have been partly due to the fact that no screen was necessary in its use because all the feedback was physical and auditory. This brings the user much more into the sonic environment, forcing them to listen more and react, rather than relying on information on a screen, which is often not precisely what's happening in sound. Some users felt that though it had strengths in certain areas (spins, circular movement, individual axis movement), these limited it in others (straight line movement, simultaneous sound control, selection switching). The main criticisms of it were that it took a long time to perform and record the movements, longer than the others and that it may not be as accurate either, meaning that any recorded automation would need to be adjusted, taking even more time. However, these are all things that could be improved with user familiarity and a more finely tuned prototype. Many users made suggestions for what they would like to see in a 'version 2.0', which judging by the reaction, could well be justified.

8.4. Screenspat

With one exception, this interface was the most popular which can probably be largely attributed to it being the most futuristic and different to what we know today, but without removing many of the elements we know from a standard computer interface. The users liked the intuitiveness of the interface, that once they knew what controlled what, the actual process of spatialization was straightforward. Some enjoyed the additional visualization of sound positions in 2D (the multisliders at the top) in addition to the 3D model, but conversely some felt that the visual displays detracted from using their ears to determine sound positions. One user felt that it might be difficult to scale up to use more than four tracks simultaneously, as more sliders would have to fit into the same space which would decrease the size on screen and so the accuracy of each controller. This could be addressed by either increasing the size of the screen, perhaps by having the user stand up in order to make use of the full width and height of the arms' range and/or by increasing the accuracy of the infra-red detection.

9. General Results Conclusions

In addition to comments, the testers were also asked to rate each interface in terms of its effectiveness as a spatialization system. They were asked to use a 7-point scale (with the mid point being equal to a standard keyboard and mouse system) so that there is essentially a scale of three in a positive and negative direction. Even though the primary interest is in gathering rich data from these tests, it was interesting to see if there were any patterns in people's experiences and so whether there was an overall preferred interface. Here are four graphs representing the interfaces and the user's grading (Figure 9.1-9.4).



Figures 9.1, 9.2, 9.3 & 9.4: Graphic illustration of the testers' preferences of interface. A strikethrough of the initials denotes that the user did not test that device.

In most cases, the problems that have been identified during these tests are reasonably small and are things that could be vastly improved or fixed in a second version of the interfaces. However, as with any interface it will always be that they will work better for some than for others because of individual preferences for ways of working. Having said that, there are some overall points and general consensuses that have been reached as a result of the tests. People will generally prefer a tactile environment to a virtual one, something physical to touch or respond to rather than only seeing or hearing the result on a screen. This was evident in the feedback from the Wheel-e-bin and the SpaceNav and I would suggest that if the other two interfaces had included some sort of touch element or haptic feedback, they would be even more popular. One issue that has arisen that is not so easy to solve is that of the physical position that gestural interfaces require of the hands and arms and that prolonged use could lead to tiredness, strain and even injury. The action of suspending our arms in front of us is not a natural one and is not yet realistically practical for the equivalent amount of time that many of us now spend at a computer with a keyboard and mouse. However, I would suggest that this is primarily because we are simply not accustomed to it. An orchestral conductor can have his/her arms suspended and moving in a similar way for sometimes up to two or three hours at a time because they have practiced and have become used to it. If gestural control does become more widespread in the near future, then it seems we will develop the ability to do likewise. The problem is that this needs to happen in order for it to become popular, but it needs to become popular in order for it to happen! It’s a catch-22 situation.

One limitation that was forseen before testing and deliberately avoided was that of the setting up of each interface. Each one requires quite a lot of preparation in order to work fully, in both the hardware and software domain. It was decided that as they were still at the prototype level the setting up would be done for them each time so that they could assess the devices equally as working interfaces, rather than being swayed by the setup complexity. Although this is something that will probably have a large impact on whether an interface is used in the real world, I think this is something that would be addressed at a later stage in development, rather than here at the research stage. Similarly with the software interfaces, the design has not been focused on an aesthetically pleasing GUI, so the inner workings of the patches are visible, but hopefully concealed enough to make it functional to the testers.

During the course of this short project, I have been able to investigate and experiment with various sensors and alternative methods for 3D sound spatialization and as a result have paved the way for some valid, potentially usable interfaces.

From the research and tests that were carried out it appears that there is a genuine demand and enthusiasm for gesturally controlled sound spatialization and as three-dimensional media becomes more common the demand for alternative methods of control will increase. This research has outlined the strengths and weaknesses of some of these alternatives and the limited testing has shown areas that could be improved upon in the future.

The research has reached a stage where testing has occurred and feedback has been received for potential improvements. The next stage will be to acknowledge these suggestions and use them to create more stable versions of some or all of these interfaces and hardware. It is hoped that whoever takes up this task will be able to use this software as a starting point to go on to bring gestural control into the professional spatialization world.

Bibliography

- Baluert, Jens, *Spatial Hearing : The Psychophysics of Human Sound Localization*, rev. ed. (London, 1997)

- Brown, Kenneth, Alcorn, Michael, and Rebelo, Pedro, *Sound Diffusion using Hand-held Light-emmitting Pen Controllers*, (University of Michigan, 2005)

- Clarke, Michael, 'Enmeshed: live in 3D fog~', In *Proceedings of the International Computer Music Conference 2006,* pp. 13-16. (New Orleans, Louisiana: International Computer Music Association, 2006)

- Clarke, Michael and Manning, Peter, *The influence of technology on the composition of Stockhausen's Octophonie, with particular reference to the issues of spatialization in a three-dimensional listening environment.* Organised Sound, 13 (3). pp. 177-187

- Roads, Curtis, 'John Chowning On Composition', in *Composers and the Computer* (Los Altos, CA, 1985) pp. 17-25

- Chris Dobrian, David Zicarelli, Gregory Taylor, Adam Schabtach, Joshua Kit Clayton, John, Richard Dudas, Luke DuBois 'MSP Reference Manual' (Cycling '74, 2001) p. 396

- Dodge, Charles and Jerse, Thomas, *Computer Music: Synthesis, Composition and Performance*, 2nd ed. (New York : London, 1997)

- Gardner, Wiliam, *3D Audio Using Loudspeakers* (Boston, 1998)

- Humphries, Christopher, *DiffuserGuide*, Interactive Sound Design 3 (BA (Hons) Music Technology, University of Huddersfield, 2008)

- Mac, Paul, 'Right Round, Baby – An Immersive Multi-Channel Experiment At Huddersfield Uni' *Audio Media*, ed. Marshall, Lanna (June 2009, Issue 223, p. 16)

- Roads, Curtis, *The Computer Music Tutorial*, (MIT Press, c. 1996)

- Roffler, S.K. and Butler, R.A., *Factors that influence the localization of sound in the vertical plane*, Journal of the Acoustical Society of America, 43(6): (pp. 1255-59) (1968)

- Rumsey, Francis, Spatial Audio (Oxford, 2001)

- Spielberg, S., dir., *Minority Report* (20th Century Fox, 2002)

- Stockhausen, K., *Kontakte*, (London, 1968)

- Wilson-Bokowiec, J. and Bokowiec, M., 'Kinaesonics: The intertwining relationship of body and sound', *Contemporary Music Review*, Vol. 25, Issue 1&2 (February 2006) (pp. 47-57).

Webography

- Waisvisz, Michael, '1984 - 1989 The Hands (first version)' http://www.crackle.org/The%20Hands%201984.htm (Accessed 8th January 2010)

- Bokowiec, Mark, 'Bodycoder Interactive Performance' http://www.bodycoder.com/bodycoderprogram.html (Accessed 10th January 2010)

- 'The Spatialisateur Project Overview', *IRCAM online*, http://support.ircam.fr/forum-ol-doc/spat/3.0/spat-3-intro/co/overview.html (Accessed 8th November 2009)

- 'Human Mocap', *Xsens*, http://www.xsens.com/en/company-pages/company/human-mocap (Accessed 31st December 2009)

- Chung Lee, Jonny, 'Johnny Chung Lee – Human Computer Interaction Research', http://johnnylee.net (Accessed 24th November 2009)

- Seldess, Zachary, 'OpenGL 3D Flight simulators' *Zachary Seldess.com*, http://www.zacharyseldess.com/z.glNav.html (Accessed 12th December 2009)

- 'Simulus – P5 Glove Developments', *Simulus*, http://www.simulus.org/p5glove (Accessed 18th August 2009)

- 'Orient 3D – Manual', *Infusion Systems*, http://infusionsystems.com/support/Orient3D-manual.pdf (p.3) (Accessed 17th July 2009)

- 'I-CubeX Online Store – Wi-MI : Wireless MIDI interface' *Infusion Systems*, http://infusionsystems.com/catalog/product_info.php/products_id/57 (Accessed 1st January 2010)

- 'iPhone – Mobile phone, iPod and Internet device', *Apple (United Kingdom)*, http://www.apple.com/uk/iphone (Accessed 31st December 2009)

- 'Asus – Eee Top', *Asus,* http://event.asus.com/eeetop (Accessed 31st December 2009)

- 'HP TouchSmart PCs', *HP*, http://www.hp.com/united-states/campaigns/touchsmart (Accessed 31st December 2009)

- 'Microsoft – Surface', *Microsoft*, http://www.microsoft.com/surface (Accessed 31st December 2009

- 'JazzMutant – Lemur, multitouch modular controller for sequesncers, synthesizers, virtual instruments, vj and light', *Jazzmutant*, http://www.jazzmutant.com/lemur_overview.php (Accessed 31st December 2009)

- Chung Lee, J., 'Projects – Wii', *Jonny Lee.net*, http://johnnylee.net/projects/wii (Accessed 1st January 2010)

- 'Stockhausen introducing "Kontakte" at The Royal College of Music in Stockholm 12th May 2001', *Stockhausen in Stockholm May 2001*, http://home.swipnet.se/sonoloco7/stockhausen/stockhausenmay2001.html (Accessed 28th January 2010)

Appendix 10.3: Hardware and Software used

Interface	**Hardware used**	**Software used**
General	- Mac Pro Quad-Core 3.2GHz, 8GB RAM with dual 23" Cinema displays - RME Audio interface (ADI-6432) - Genelec 8240A Monitors (x16) - 3D Connexion Space Navigator	- Cycling '74 Max/MSP v.5.0.4 (http://cycling74.com/products/maxmspjitter) - IRCAM Spatialisateur v.3.4 (http://forumnet.ircam.fr/692.html?L=1) - Steinberg Nuendo 4 (http://www.steinberg.net/en/products/audiopostproduction_product/nuendo4.html) - Zarchary Seldess' OpenGL 3D Flight simulator 'z.glNav2' (http://www.zacharyseldess.com/z.glNav.html)
P5 Glove	- Essential Reality P5 Glove	- Tim Kreger – 'P5osc' P5 Glove command line OSC routing software (http://www.simulus.org/p5glove)
SpaceNav	- 3D Connexion Space Navigator	N/A
Wheel-e-bin	- Infusion Systems' iCubeX Digitizer - Infusion Systems' Orient 2D - Infusion Systems' Bend Sensor - Infusion Systems' Push Sensor - Infusion Systems' 'Wi-Mi' wireless MIDI - MOTU midi express 128 - Griffin Powermate	- Infusion Systems' i-cube/o-cube Max plugins v2.10 (http://infusionsystems.com/catalog/product_info.php/products_id/83)
Screenspat	- MacBook Pro 17" Intel Core 2 Duo, 2.6GHz, 4GB RAM - Panasonic Projector (SPEC) - MOTU MkII Audio Interface - Yamaha 01V96 Mixing Desk - Nintendo Wii Remote Controller	- Cycling '74 Max/MSP v.4.5 (for plugin creation) (http://cycling74.com/products/maxmspjitter) - Wiimote Whiteboard (http://www.uweschmidt.org/wiimote-whiteboard)

Appendix 10.4: Questionnaires

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Richard Glover

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	**Comments**	**Strengths**	**Weaknesses**	**Grade 1-7**
SpaceNav	Very intuitive. Easy to use, visual display makes sense, could work well in studio or live situation	Fun! Tactile approach feels very intuitive. Ability for a constant sweep around the environment is well controlled.	Instead of the Space Navigators, perhaps a touchpad could work – this would help immediacy of control. The SNs mean that you have to pull sounds around the space, whereas with a touchpad, the sounds can be moved directly to any position. Could build in different tracking speeds.	6
P5 Glove	*[Not available at time of testing]*			
Wheel-e-bin	Good – perhaps not the best for overall control. Great for moving sounds with easy *around* the environment.	Geared towards circular movement of the sound, which works very effectively.	Difficult to move sound in a linear fashion from one point to another opposite. Also perhaps distance and height controls could be more intuitive (e.g. directly on the wooden board).	4.5
ScreenSpat	Tactile, easiest to use. Very versatile and quick to change parameters.	Very simple, immediate. Great control. Visual display very easy to follow. Also control of cursor means it is far more flexible. Very easy to immediately move sound from one point to the next – a great strength compared to the other two.	Build height into the same control as the xy scale.	7

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Richard Glover

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

They all did to varying degrees of success – certainly all easier than mouse and keyboard for spatializing.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

Certainly the ScreenSpat, perhaps the SpaceNav as well – although that feels a little clunky now compared to the Screenspat. Wheel-e-bin has limited use – i.e. it can only control the spatializing, and difficult to control visually.

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

On SpaceNav and Wheel-e-bin I would have liked a quicker/easier ability to move the view around, it felt a little tricky at times. Screenspat has most room for innovation I think – could include volume controls on each channel, link in height control, etc. Difficult to see how the others could evolve easily.

All things considered, do you have a preference, and if so, why?

Screenspat – easily most intuitive and most fun. Has a great immediacy, but also allows for distinct subtlety. Two-handed approach makes control much easier and faster. Also would be the easiest for many different channels.

Any other comments?

Certainly the Screenspat could become an industry standard! This is a worthwhile project – specifically just aimed at Electroacoustic? I think if these kind of formats were introduced into the commercial world, many studios/engineers would be very interested: the Screenspat has such great versatility that it's difficult to see why industries wouldn't be interested in promoting it.

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Adam Jansch

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	**Comments**	**Strengths**	**Weaknesses**	**Grade 1-7**
SpaceNav	The most flexible method here, but as it doesn't deal with a direct, tactile surface it is the hardest to use initially.	Incredibly flexible, allowing access to spatialization parameters quickly. Useful for creating a path through the space.	Hard to get the hang of, would take practice to really refine the control of it. Could use a reset sound or jump to position button. Relies too much on the graphical interface.	4
P5 Glove	*[Not available at time of testing]*			
Wheel-e-bin	Really fun, allowing access to a different range of spatializing gestures, though I think limited otherwise.	Allows elegant control of specific aspects of sound position, like spinning. Gestures map intuitively.	I thought it is a bit bulky, and the distance/height control is a bit inelegant. Fairly limited outside its strengths.	5
ScreenSpat	I thought this one was cool, really easy to get the hang of. Being able to touch the screen would be cool, to give a tactile element.	Intuitive graphical interface, very approachable, fun.	Could be tiring after some use, maybe it could be positioned on a slant or flat? Also the graphical elements distract from actually listening.	6

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Adam Jansch

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

I think all the methods succeed in improving the relationship between spatialiser and the technology.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

For stereo work I would certainly see the ScreenSpat as a useful interface. For very specific multichannel gestures the Wheel-e-Bin would also be useful.

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

The ability to unlatch certain parameters would be useful, so that automation for each parameter could be recorded more precisely. The 3D representation was a bit ungainly and took focus away from listening to where the sound was. Being able to set the distance range would also be helpful.

All things considered, do you have a preference, and if so, why?

With practice the SpaceNav would probably be the most flexible system, though if mounted horizontally I would choose ScreenSpat, due to its intuitive approach.

Any other comments?

I think that the three solutions presented have some overlap but really specialise in particular areas, so it's hard to definitively compare them. SpaceNav is good for any kind of non-linear path, Wheel-e-Bin specialises in circular position and motion, and ScreenSpat fits it nicely as a direct positioner. The ultimate interface may take elements from all three.

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Sam Freeman

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	Comments	Strengths	Weaknesses	Grade 1-7
SpaceNav	*[Not available at time of testing]*			
P5 Glove	Found the virtual nature of this type of interface is a hindrance, the motion of ones hand is relative only to the onscreen environment such that navigation requires focussed attention to the 3D GUI	It did seem to work pretty much as designed to	This shares weakness of ScreenSpat while also lacking the strengths of Wheel-e-bin.	2
Wheel-e-bin	Once I had a source selected it was no longer necessary to look at the monitors as I could use the controller to move the sound focusing purely on auditory feedback. Perhaps more buttons could be mounted on top of the bin lid to facilitate functional interaction such as source selection, play/stop…	3D placement controller is very good; a transparent interface with physical motion translated (perceptibly) directly into the audio result.	I had to resort to mouse interaction to grab a source to be able to move it – this is mostly due to my own skills, but it would have helped if, for example, they grey hoop (when not carrying a blob) were to turn the colour of a blob that it is in range of 'picking up' – that way user would not have to guess when to click the select button.	5 – the wheel-e-bin is great, but I cannot grade higher as for me, the mouse was still needed more than it might have been
ScreenSpat	no problems with the physical interface which worked very well – I would have been more comfortable if I were stood up to use it. The software interface is ok, but better labelling would help.	Having IR on each hand very good – even tho the system is not multi-touch the user is able to jump the cursor from one point to another very quickly. Have the laptop right there to revert to if needed a very good idea	The gloves pose a barrier to the interaction – maybe a pen like object to hold would be better.	4 = same as mouse and keyboard because while the mouse action is better (see note at strengths) the keyboard element has been removed.

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Sam Freeman

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

Yes, wheel-e-bin inparticular, and the 'touch screen' hack of the spat one are definitely moving toward this goal

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

Yes, those cited above

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

All, or at least some, of the onscreen push buttons could have been mapped to qwerty

All things considered, do you have a preference, and if so, why?

Wheel-e-bin is a fantastic controller and deserves further development

Any other comments?

Please do let me know how these things develop in the future

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Scott McLaughlin

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	Comments	Strengths	Weaknesses	Grade 1-7
SpaceNav	(4th used) Nice, but some downsides.	As with glove, very intuitive and clear where the sounds are going.	- Slow to move, sound moves at just one speed: smoothness is excellent but more speed variety is important for gestural music. - I would prefer 2nd mouse was another voice and that the view was fixed, don't think the view needs to move. If we had four of these mice that would be even better: that would be best compromise between this and screenspat.	4
P5 Glove	(3rd used) Great, simpler selection process than wheel-e, easier to learn. Interface more intuitive also.	As with screenspat, using my hands is most intuitive, allows fine control of speed and position. 'grab' gesture is also very clear, are there more like this that can be unambiguously encoded to allow more functionality?	Compromise between smoothness of motion and speed of motion could be problematic, but is more of a performance/virtuosity/experience issue.	6
Wheel-e-bin	(2nd used) good interface, slightly more awkward and slower than screenspat due to drop-move-move method	Wheel-e-bin itself is great, more intuitive control for 3D than screenspat.	Click-move-click action could be simpler and all one single interface: use footswitch, thumbclick, etc?	5
ScreenSpat	(first used) Very attractive idea, because of the clear link between gesture and response, very intuitive	- Flexibility, intuitive and easy to navigate when the array of tracks is small.	- Will be tricky to scale upwards I think, as the number of tracks increases, but ultimately no harder than standard mixing desk approach. - Sensors are a little inexact but this is a prototype issue and easily fixed.	6 (ignoring issues of equipment not being as exact as could be)

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Scott McLaughlin

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

Yes, definitely. These methods add a greater level of intuitive control of the spatialisation as well as (theoretically) more specific control.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

Yes,

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

No

All things considered, do you have a preference, and if so, why?

The glove was most intuitive, if you could make the glove select between tracks as well then that would be best.

Any other comments?

All the interfaces have their good and bad points, but overall there is an excellent and much needed approach in here. This is very important research, I think you can find the best points of all approaches and combine them.

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Mike Fisher

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	**Comments**	**Strengths**	**Weaknesses**	**Grade 1-7**
SpaceNav	Very quick to use but can be confusing due to 3D. Would be great for 2D i.e. only 1 ring of speakers.	Quick – easy to move all 4 tracks to a general position. Buttons to scroll through tracks makes things very fluid. Moving around the 3D model is intuitive.	Strafing would be useful for moving around the 3D model, as would up/down (shame you've run out of axes!) Depth perception is very difficult. SpaceNavs are very sensitive – moving side to side moves a little up and down too.	4
P5 Glove	Very cool!	"Grab" is brilliant! Better link between hands and sound than spacenav but not as good as wheelebin.	Quite shaky and depth perception is difficult though not as bad as spacenav.	5
Wheel-e-bin	Definitely facilitates spat. User has a visual idea of where the sound is in the room by looking at ruler angle and fader position.	Direct link between hands and position of sound Circular motion reflects orientation of speakers 3 axes are very intuitive	Doesn't provide accuracy – no problem for creative mixing of music but would be very difficult to use for audio/visual work i.e. films. Written automation would need to be tweaked.	6
ScreenSpat	The best of the 4. Having a 2D and 3D represenation is v. useful – no depth perception issues. I prefer using the 3D model as a visual guide rather than part of the method of moving sounds around.	V. quick and easy. Nice big screen to use. Levels for each speaker are very cool and very useful. Radii showing perpendicular distance are helpful when the speakers are shifted 12.5^{o} from the clock i.e. 12,3,6,9. Also shows which speakers are being used and the ratio of volumes.	Finger is accurate to about 1 square centimetre; most of the times its plenty but can be frustrating near the edges of the screen.	7

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Mike Fisher

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

All of the methods are more gestural than using a keyboard and mouse and all the but the SpaceNav are more intuitive. I think the bridge between conception and execution is achieved by there being a physical link being the gestures and the resulting position of the sound i.e. moving your hands around rather than using your hand to move a computer component.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

No, but only because I have limited interest in working outside of Stereo. If perfected however I would hypothetically use the wheel-e-bin; having a visual representation created by the ruler bearing, angle and fader distance provides a combination between the physical link of using your hands described above without the inaccuracies of the "hands-on" methods.

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

All the patches are very easy to use and very user friendly. I could deal with them being more complicated in fact e.g. having four faders on the ScreenSpat method would have been useful. I like the professionalism of user simple interfaces covering up very complicated programming underneath.

All things considered, do you have a preference, and if so, why?

My preference is the ScreenSpat method. I found the greatest sense of immediacy between what I was doing and how the sounds were moving, and I like idea of a glove – pointing your finger at the screen and having the button on your thumb is nifty! However this is closely followed by the wheel-e-bin because of the visual representation of the 3 axes discussed above.

Any other comments?

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Liz Dobson

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	**Comments**	**Strengths**	**Weaknesses**	**Grade 1-7**
SpaceNav 1	I was aware of putting too much mental focus on the visual display, rather than just listening for the movement. This might be because of not being so familiar with the tool.	Nice to move a dial and for the sound to move up, down as much as to the sides and surround.	The dial stops and its relation to the speed and continual movement of a dot is not great. Would be better if you could turn the dial by an amount and have a dot follow with closer relation to this rather than the same speed. If the speed does vary this is hard to control so feels like the dot is arriving late.	3
P5 Glove 2	Apart from holding my arm in a difficult position – not sure how it'd cope after a couple of hours, this is extremely easy to use.	Close feeling of being in contact with the sound movement (good for recording automated movement gestures). Highly intuitive and precise. Good relation of movement speed with the glove. Seems very precise and natural way of manipulating sound in space to me.	Only that the physical position could cause repetitive strain!	9
Wheel-e-bin 3	Impressive piece of kit. Found myself leaning into the screen a lot and with my head so close to the screen, wondering if I was hearing the movement as I should.	Again, close relationship between movement and position. Immediate, like the other glove. Would work well I think for movement on either the horizontal or vertical dimension.	With one mouse to play with it isn't possible to move sounds diagonally in the space. This is a shame because using both hands together this would've been interesting. Easier to switch between sounds (more like a conventional mixer).	5
ScreenSpat 4	*[Not available at time of testing]*			

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Liz Dobson

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

Some methods act as more of a barrier because they seem to add another level of cognitive engagement; the relationship between hand gesture, sound movement and visual movement seems complex at times. The two dials were a good example of the kit getting in the way for me. The two gloves were an improvement but they prevent movement on diagonal axis. By far the most intuitive to use was the P5 glove. It seems to enable all directions at a pace and relationship which matches my intention.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

I would certainly work with the P5 glove. The two gloves were also interesting however not as flexible for movement as I say. The two dials would take a lot of time to get used to. Perhaps, whilst the P5 is intuitive, maybe the dials can do more with practice.

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

I would have liked to control the volume of each track (two gloves) and I liked the interfaces that looked more similar to a mixing desk. I can't think of needing any other controls however. It was good to be able to mute and solo and this could be a feature for the other interface perhaps?

All things considered, do you have a preference, and if so, why?

Definitely. The P5 glove. I really felt like I was able to grab and position sounds. I could focus more on the audio positioning (sonically rather than visually) also.

Any other comments?

Impressive work! This is useful research because the current interfaces are of course lacking for sonic movement in spaces like this. Quite Minority Report!

Thank you for taking the time to help me with my research.

Alternative Methods for 3D Spatialization – Questionnaire 1

Name:

Jonathan Rich

Please comment on your experience using each of the controllers. What were the individual strengths and weaknesses? Were they practical? Did they facilitate the spatialization or hamper it? Please also grade each system 1-7 on its effectiveness as a spatialization system where 7 is high and 1 is low (4 being a standard keyboard and mouse combination).

Sensor	Comments	Strengths	Weaknesses	Grade 1-7
SpaceNav	A nice and convenient method to use for spatializing. I can imagine using it when mixing multichannel setups.	Precise control over sound positions. Good for mixing in a studio rather than in a live performance.	Slow to move sounds. Only move one sound at a time. More controls needed. Difficult to grasp at first.	5
P5 Glove	Feels the most natural to use. Wouldn't need a screen to use. Could be developed to include more controls using hand movements.	Quick movements. Easy to cycle through sounds. Easy to visualise. Quick to grasp the concept. Would be the best for live spatialization.	Only move one sound at a time. Arm aches after a while. Restricted movement to one area. Should allow to be used anywhere in the room.	6
Wheel-e-bin	Quite difficult at first to grasp movements. Good concept though. Would be good smaller like the size of a mouse.	Self-contained. No need for wearing any accessories.	Longer process to move individual sounds compared to other methods. Restrictive in terms of movement. Harder to visualise. Only move one sound at a time. Would be the worst for live mixing.	2
ScreenSpat	Easy to use and understand. Quick to grasp the concept	Gave a lot of control for each sound in terms of positioning. Easy to visualise on the screen	Not enough controls. Use a touchscreen would give more freedom. Only move one sound at a time	4

Alternative Methods for 3D Spatialization – Questionnaire 2

Name:

Jonathan Rich

The purpose of this project is to produce an interface that is more intuitive / gestural than the keyboard and mouse in order to bridge the gap between conception and execution of sound movement. Do you think any of the methods fulfilled these criteria?

All of them apart from the bin interface provided a more intuitive experience for 3d spatialization than a keyboard and mouse. The glove has the most potential to be an excellent tool when diffusing or mixing in a studio.

These devices are still in the prototype phase. If perfected/developed further and once you became familiar with them, could you see yourself using any of these interfaces in your work?

I could see myself using the glove and the SpaceNav to help mix. The Screenspat would be better if it was simply touchscreen and the Wheel-e-bin would be better if it was smaller and produced at a high quality standard.

How did you feel about the software environment? Were all the controls you wanted present / accessible? Were there things you would like to have done but couldn't?

More controls were needed, including spreading individual sounds wider or narrower, controlng volume without having to use the mouse, starting and stopping, mute, solo etc without using the mouse. Also it would be good to develop a method for pairing sounds or being able to move two or more sounds at the same time.

All things considered, do you have a preference, and if so, why?

I prefer the P5 glove because it provides the most natural, fun and convenient interface. With a bit more development it has the potential to be useful for manipulation, and mixing sounds in 3d with ease, more so than a mouse and keyboard.

Any other comments?

Good work!

Thank you for taking the time to help me with my research.