SOUNDS IN SPACE: A FOLIO OF SPATIAL WORKS AND PRODUCTIONS USING GAME TECHNOLOGY

Daniel Pitman
University of Adelaide
Elder Conservatorium
Sonic Arts

ABSTRACT

This paper will overview the investigations of the author in creating spatial audio interfaces by exploiting recent developments in virtual reality and gaming technology. In keeping with the theme of this conference, spatial sound synthesis, game engines, and tracking technologies are common tools for the *mimesis* of space about us. Unlike simply recording binaural sound, tracking the listeners head and synthesising spatial audio in real-time allows for greater interaction with that space (including listener head movement), furthering immersion. However, new technologies come with new challenges as well as new potentials.

A folio of musical works is being composed and/or produced by the author; a result of investigations that apply game engines, VR, live streaming, and game physics to novel musical applications.

1. INTRODUCTION

Since the re-emergence of VR in 2014 interest in spatial audio has bloomed in the game industry along with the CPU/GPU power to implement it.

> *In general, I would say that unlike any other creative medium people are pushing the importance of sound from top to bottom. This is exciting for audio engineers. (Fodor 2016)*

This is certainly an exciting time for spatial sound artists and should be for musicians as well.

> *It is evident that sound in space remains a potent research area for contemporary composers and sound artists. (Lyon 2014, 851)*

Spatial sound interface is an expanding field of research for investigating interface to control sounds in spatial implementations, often combining developments in gesture and interface with acoustic propagation and diffusion. So varied are the problems seen in this field, and the solutions put forward in this field, it could be argued that the composer may be inclined to build their own interface as a specific solution or integral part of a work. In this sense the spatial sound interface incorporating location, scale, distance, and architecture, can be considered an instrument of its own with many possible designs depending on the specific needs of the performer/composer. The composer doesn't have to include all spatial features in every piece but might need to utilise any of them in any piece.

Several reasonable designs for spatial sound interface are being developed (a worthwhile review for another paper). But this project is investigating potential in 3D game development software for sound and music production and performance. In gaming, spatial acoustic propagation and gesture/interface have been paired together for decades. Modern game development software such as Unity or Unreal have been increasing support for various third party spatial sound plugins to the point where now basic spatial sound engines are considered standard options.

Further, game development software is no longer the domain of game development experts. One of the milestones achieved by both Unity and Unreal Engine is their accessibility. This paper proposes that "game development software" will evolve to be used as "development software" for tasks in many artforms like cinema, music, literature, and various industries like medicine, engineering, and education, as each industry finds new value in spatial or virtual experiences.

2. CREATIVE INVESTIGATIONS

Game technology has, as this paper will discuss, some excellent solutions to many of the problems that composers and sound artists currently face with spatial sound.

2.1. Propagation

Most obvious is practical acoustic propagation synthesis. It is currently (as of writing) too expensive to solve a full wave propagation solution in real-time, nor is it necessary for immersive musical and performance purposes. Sound engines such as the *Oculus Spatializer*, Google's *Resonance* and *Steam Audio Spatializer* (Based on *Phonon*) are designed for affordable real-time processing yet still feature accurately rendered early architectural reflections, occlusion/diffusion (objects being behind one-another), and other phenomena such as near-field effect and Doppler effect. For comparing some currently competing systems see Gould (2018).

Further advancements can be expected as new spatial audio solutions are emerging using GPU processing (Hamid and Kapralos 2009) and with raytracing now a feature of the latest range of Nvidia cards, the potential for complete wave propagation synthesis seems much closer.

The ability to design and dynamically change architecture (moving walls etc.) on the fly is difficult, as many architectural computations are 'pre-baked' (calculated in advance) to allow for faster computations of reflections and occlusion at run time.

Using game engines to design musical works encourages composers and the audience to consider architecture as a creative element in composition. The author's piece, *They Glow,* puts the listener in a deserted cityscape surrounded by giant, glowing, singing monoliths that circle around them through the city-scape.

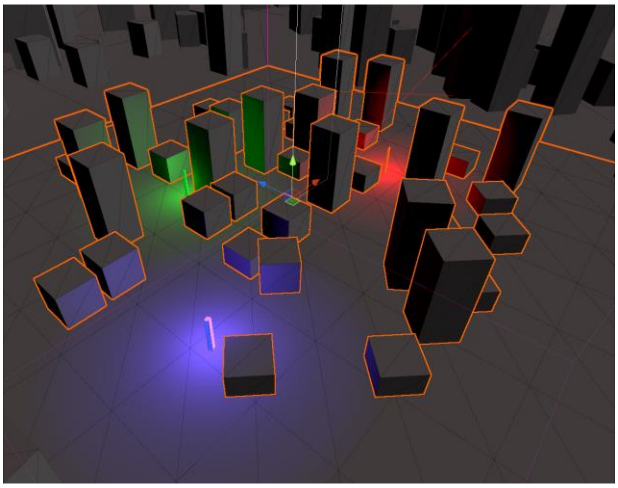

Figure 1. One possible city-scape arrangement in which the three (colored) monoliths move about.

The listener is free to watch in any direction they choose (either using typical FPS game controls or a VR headset), witnessing the eerie reflections of the light and sound about them and experiencing the architecture itself. The guttural singing of the monoliths (recorded using a bowed, fretless, electric guitar) acts as an 'exciter' that agitates the acoustic space.

2.2. Interface

Gaming is inherently tied to interface and gesture design. With the recent re-emergence of VR circa 2014, gaming technology arguably rests squarely at the forefront of spatial interface design. There is a history of relations between music and gaming going right back to the musically enhanced interface *Pong* (1972) and giving rise to rhythm challenge elements seen in *PaRappa the Rappa* (1997), dedicated hardware interfaces such as the guitar shaped controller in *Guitar Hero* (2005), game play fuelled generated music *Electroplankton* (2006), and even musical objects throughout the quasi-programmable world of *Minecraft* (2009). Game development and music share interface design as core concepts.

Moving objects in space is a common activity in this project. Almost every game and music controller offers a solution, however the ability to create feedback and return significant force to the user, sometimes quite unforgivingly, is quite unique to driving simulators and their controllers.

The author's folio piece, *Music for Rally Car and Track*, is a novel example of utilising that feedback-based interface. It sonifies the telemetry of a simulated rally car as it completes a bumpy track. It is inspired by and tributes the famous motorsport film *Climb Dance* (Mourney 1990).

Figure 2. Sonifying telemetry from the infamous Pikes Peak hill climb using the Dirt Rally simulator (image taken from *Dirt Rally* for PC published by Codemasters 2015).

This work can be played using most driving simulators (or a real car with telemetry sensors and output, such as most dedicated race cars). The center of gravity, gears, speed, weight distribution, and controls all contribute to variables in the generated piece. In this performance the car becomes an instrument, the track becomes the score, and the driver a performer 'following' the score. As the weight of the car shifts about, so too do the musical elements associated with it, developing an intuitive relationship between driving interface and the music that results. The system used for this piece may have potential as a feedback mechanism for drivers of increasingly quiet electric cars, or competitive driving.

2.3. Tracking Hardware

Some prototype hardware was devised as a part of this project, but since the HTC Vive and Oculus Rift went commercial circa 2014 many of the challenges of room scale tracking were reliably overcome.

Figure 3. Prototype tracking headphones combining infra-red, magnetometer, and gyroscopic tracking.

The HTC Vive uses relatively stable, inside-out, laser scanning and features a tracking device that is easily attached to peripherals, perhaps the most important feature of the HTC Vive (in the context of this project).

For example, Vive Trackers attached to headphones provide tracking data about the listener head position without the use of the headset (Figure 4). This is used for folio installation works like *Heavily Oxygenated Chromatic Undertow* that don't utilise the graphics in the headset. The listener must locate and find the small

sounds in the space that make up musical elements of the piece by ear and rearrange them using the Vive hand controllers as they see fit to reconstruct the music loops into new passages.
Tracked headphones are also used without graphics in the claustrophobic piece *Microbic,* where the listener is surrounded by a mist of tiny sounds (mouth noises, stick tape, crackling tissue etc.) in an unnatural simulated acoustic room environment that sounds as if it is shrinking.

Figure 4. The Vive Tracker with wireless headphones for tracking listener orientation and position.

Tracking devices can also be attached to binaural microphones, to combine real world spatial sound with digital spatial sound. They use real world head and ear geometry to preserve inter-aural level, inter-aural time and the geometric effects of the head and outer ear in the sound signal (collectively HRTF). This folio features a novel combination of an analogue binaural microphone tracked in VR so that it may double as a digital binaural synthesis listening point as well, allowing digital and real-world spatial sound to be mixed together around a central point of the listener/audience.
In this project Unity was used to develop a set of custom sound making and listening modules. The sound making modules are moveable ‘balls’ featuring various hand controller interfaces for different kinds of playback such as stochastic playback, rhythmic synching, looping, or being triggered on impact. The listening elements double as the virtual manifestation of the tracked headphones or tracked binaural microphone. These elements, along with a series of controllers and acoustics tools, are collectively referred to as *The Sound Sandbox.*

Figure 5. This binaural microphone was scratch built from ear moulds cast in a silicone that is designed to imitate skin. A Vive Tracker is attached to facilitate virtual tracking so that analogue and digital binaural sounds can be mixed during performance.

Their purpose is to be used in the creation of interactive spatial sound or music making environments, and various environments have been created by the author using these elements. One environment is designed for a scored performance piece utilising tracked binaural dummies. *Rain for Violin and VR* uses two separate dummies, one tracked digital dummy for the VR performer to perform with, and one analogue binaural microphone for a violin player who is choreographed about it throughout the piece. The piece exploits the combination of two approaches to explore a severe anxiety attack and schizophrenia like symptoms.
Similar environments, dubbed *Improvisation Environments*, have been developed as improvisation spaces, designed for online streaming performances that evolve according to real-time audience suggestions.

2.4. Distribution

Game engines also provide solutions for distribution. It is difficult to record and distribute spatial works currently, as there are still several format to choose from (each with unique demands for playback). Pieces in this folio often champion listener movement and dynamic reorientation (a feature which could arguably enhance recordings of traditional spatial of pieces by the likes of Brant, Ives, or Xenakis), lost in hard recording. Some of folio pieces also implement changes to audible architecture as a core feature, also a feature lost in any hard recording. Thus, a game ‘app’ makes for a potentially useful distributional format for such sound and music productions. It is intended that certain pieces and interactive environments from this folio will be distributable as ‘apps’.

2.5. Other Creative Possibilities

There are many proven technologies, considered elementary in modern gaming development, that offer potential solutions and un-taped creative potentials for musicians. Some of these include:

- **Networking-** Simultaneous long-distance performance, virtual installations, virtual busking, AI performance, or performer/audience interaction.
- **AI/NPC Implementation-** NPC (non-playable character) and AI design are staples among gaming. Sound or music projects involving AI might benefit from the infrastructure provided in game development software. This is a specific goal for future projects by the author, utilising affective algorithmic composition with neural networks (Pitman 2015).
- **Physics-** The various physics simulations available in game development software such as colliding, bouncing, particle physics, ropes and string, water flow and weather effects are powerful tools that have great potential for musical projects. Some investigation in the using GPUs and game engines for real-time physical modelling synthesis seems an imminent future research project.

3. COMPROMISES

Developing musical or sound experiences in game engines that are not designed for sound design, featuring only bare bones editing or synthesis tools, is a significant compromise. A game engine is really a tool for bringing together defined assets into an interactive whole.

In keeping with this workflow, a session for improvisation or performance needs to have sound assets heavily prepared beforehand in a DAW and in keeping with certain game and VR design principles that clash with typical stereo music production.

It is worth noting that Unity, with some expertise or extensive use of third-party tools, can be coaxed into accepting standard sound protocols such as MIDI, OSC, or live sound input, however in the author's experience there is some room for developing more reliable tools for this purpose.

3.1. Recording Assets

Stereo approaches are not relevant at all in game design (outside of non-diegetic music). Any individual sound source, like a flying insect or a beer being opened, is recorded as a mono channel. A larger (or closer) object like a piano or organ, might need to be implemented as a multisource object, where those sources are intended to make up various 'sound parts' of the larger object. If the recording is captured in an acoustically live space, this will clash with the synthesised acoustic space provided by the game engine.

3.1.1. Loudness for Size or Scale?

Sound assets need to consider scale. In good spatial audio, any object that is close has different inter-aural time difference for the listener than the same object far away, even if the loudness is effectively the same despite the distance. Varying of the loudness of an asset in the DAW can create strange illusions of scale and size (such as very loud mosquito that still seems one hundred metres away, resulting in the illusion of an enormous mosquito). Loudness and scale are usually best handled in the game development software and can be adjusted there, however in the DAW itself every sound asset is typically normalized to prevent confusion later.

3.2. Tracking Hardware

There are a few limitations to the HTC Vive system. Firstly, the tracking device only works when within the laser field of a lighthouse emitter. This limits the working area to two to five metres squared. When demonstrating at events where other Vive systems are in use, the lighthouses of several systems will interfere with each other. Some conferences and live venues with a lot of natural light can be difficult to perform at, however most indoor venues and studios would make perfect conditions for laser tracking.

3.3. Why Binaural?

Binaural spatial synthesis, which simulates the locational information detected by our hearing system, is not a new technology. Real-time digital binaural synthesis hardware was developed in the late Eighties by NASA (Wenzel, Wightman, and Foster 1988), and examples even commercially released for PC gaming in the late Nineties, used by game titles like *Unreal*, and *Quake*. Digital binaural synthesis has become increasingly relevant in recent years due the re-emergence of VR, but typically without dedicated hardware.

This paper, and the project it stems from, feature a strong focus on binaural spatial synthesis (where headphones are used to deliver separate channels of audio rather than speaker arrays). There are several reasons behind this in addition to the acceptance by game developers that headphones are 'standard issue' for VR.

There are some excellent examples of artists who have successfully incorporated binaural technology into their work. Live theatre performances, where each member of the audience gets allotted headphones, are becoming increasingly common; *Fiction* (Rosenberg 2015), or *Long Tan* (Chris Drummond 2017). The BBC also has a long repertoire of binaural and semi-interactive radio shows (Mant 2002; Sachs 1978; Bank 2014).

3.3.1. Headphone Culture

Lyon points out that many venues do not have consistent or predictable infrastructure for spatial sound performance via speakers. It could be argued that this problem extends to binaural performance which requires headphones, but there are several factors worth considering, not least of which is the headphone's ubiquity in our culture.

Headphones are taking an ever increasing role in media consumption in general, not only conquering the mobile and portable music world as suggested by Bull (2007, 2000) but also at homes where, since 2013, headphones connected to home PCs appear to account for more than half the headphones in use throughout the US (NPD 2014). Headphone-laden audiences include gamers, fans of podcasts, e-sports, and live game streaming; all excellent templates for live performances.

3.3.2. Potential Audiences

Live game streaming involves a 'host' playing computer games of some kind (usually with some skill and/or humor) while an audience watches in real-time, often contributing with comments, donations, or requests for certain actions. There are several well-established platforms already in place for live streaming digitally, such as *Twitch.tv* (launched in 2011 and recently acquired by Amazon for just under a billion US dollars), *Mixer* by Microsoft, and *YouTube-Gaming* by Google. These streaming channels feature thousands of channels, many of whom specialise in certain games (or perform music) for a range of tastes, and often include various forms of monetisation via donation or subscription systems.

This emerging performance format and culture serves as an excellent template for spatial music performance using game technology. The audience is typically headphone-laiden, and the infrastructure to broadcast is reliable and in common use. The author's *Improvisation Environment* is built specifically around performing in such a way: gathering audience suggestions, potentially leaving the performance environment at times to develop new elements for as a part of that performance, and even generating and incorporating new assets with the audience's while streaming.

4. POTENTIAL FUTURE RESEARCH DIRECTIONS

GPS was investigated in this project but currently lacks the accuracy (rarely better than 5-10 metres) and stability required for spatial audio synthesis. GPS is less reliable indoors.

The potential of room-scale self-tracking VR equipment is immediate, however the amount of effective floor space available is still quite limiting for performance. This may be resolved when high definition GPS becomes available, or as laser-tracked VR implementation progresses (next generation headsets that support an array of lighthouses are released as of writing).

Various commercial, self-tracking, headphones products have become available or are in development, such as the *Waves NX* device and the *Ossic-X* headphones. To what extent and how these devices become accepted into public use may open new possibilities.

It is likely that future entertainment will involve something akin to today's multiplayer gaming's 'speactating', where the audience log in via a game client to witness performances and may even interact. This serves as a potential enhancement to limitations with streaming.

Controlling sound variables with VR controllers is difficult, and there is much room for further development in user interface.

5. SUMMARY

This paper describes a folio of works that is a novel contribution towards live binaural performance and spatial composition, where composers and artists can wield location, architecture, and space itself with as much confidence and affect as they do harmony or rhythm now. This paper discusses many of the technicalities involved and creative potential and limitations around those technicalities. The folio itself is a novel contribution toward a musical fluency using game development technologies. It is hoped that the project's folio will contribute to understanding of the specific approaches required for manipulating this spatial and interactive sonic medium effectively as a musician or sound artist, and the potential for realising new works.

6. REFERENCES

Bank, Rob da. 2014. "3D Headphone Special." Binaural. *3D Headphone Special*.

Bull, Michael. 2000. *Sounding Out the City: Personal Stereos and the Management of Everyday Life*. Bloomsbury Academic.

———. 2007. *Sound Moves: IPod Culture and Urban Experience / Michael Bull.* International Library of Sociology. Abingdon, Oxon ; New York: Routledge.

Chris Drummond. 2017. "Long Tan." Brink Productions.

Fodor, Travis. 2016. VR Audio: Trends and Challenges of Pioneering a New Field Interview by Adriane Kuzminski. Web. http://designingsound.org/2016/08/vr-audio-trends-and-challenges-of-pioneering-a-new-field/.

Gould, Richard. 2018. "Let's Test: 3D Audio Spatialization Plugins." *Designing Sound* (blog). 2018. http://designingsound.org/2018/03/29/lets-test-3d-audio-spatialization-plugins/.

Hamid, Foad, and Bill Kapralos. 2009. "A Review of Spatial Sound for Virtual Environments and Games with Graphics Processing Units." *The Open Virtual Reality Journal* 1 (1).

Lyon, Eric. 2014. "The Future of Spatial Computer Music." In *ICMC*.

Mant, Izzy. 2002. "The Dark House." BBC Radio 4.

Mourney, Jean Louis. 1990. *Climb Dance*. Peugeot.

NPD. 2014. "Premium US Headphone Market Surpasses $1B." NPD. 2014. https://www.npd.com/wps/portal/npd/us/news/press-releases/premium-us-headphone-market-surpasses-$1-billion-in-2013-according-to-npd/.

Pitman, Daniel. 2015. "Developing Methods for Predicting Affect in Algorithmic Composition," 178.

Rosenberg, David. 2015. "Fiction." Binaural. *Edinburgh Festival*.

Sachs, Andrew. 1978. "The Revenge." *BBC Media Centre*. BBC Radio 3.

Wenzel, Elizabeth M., Frederic L. Wightman, and Scott H. Foster. 1988. "A Virtual Display System for Conveying Three-Dimensional Acoustic Information." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 32 (2): 86–90.