Audio Engineering Society

Conference Paper

Presented at the International Conference on Spatial Reproduction – Aesthetics and Science, 2018 August 7–9, Tokyo, Japan

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Spatialization Pipeline for Digital Audio Workstations with Three Spatialization Techniques

Tanner Upthegrove[1,2], Dr. Charles Nichols[1,2], Dr. Michael Roan[3,2]

[1] *Virginia Tech School of Performing Arts, 247 Henderson Hall, 195 Alumni Mall, Blacksburg, VA 24061*

[2] *Virginia Tech Institute for Creativity, Arts, and Technology, 190 Alumni Mall, Blacksburg, VA 24061*

[3] *Virginia Tech Department of Mechanical Engineering, 445 Goodwin Hall, 635 Prices Fork Road, Blacksburg, VA 24061*

Correspondence should be addressed to Tanner Upthegrove (upty@vt.edu)

ABSTRACT

As more artists find utility in spatial audio for producing live and recorded music, new tools are needed to meet the aesthetic goals of each artist. This paper proposes a method for controlling and mixing point-source spatial audio with the familiar interface of Cockos Reaper, a digital audio workstation (DAW), which controls a Cycling '74 Max Patch rendering Vector Base Amplitude Panning (VBAP), High Order Ambisonics (HOA), and 3-D convolution reverberation mixing busses. The Max patch allows mixes in Reaper to translate to any arbitrary loudspeaker array by modification of the quantity and location of loudspeakers.

1 Introduction

Venues hosting high-density loudspeaker arrays (HDLA) of 24 or more individually addressable loudspeakers [1] are increasing in number, as are applications for spatial audio in virtual reality (VR), arts, and research. Musical artists that have not used multichannel loudspeaker systems before are now increasingly interested in producing music for these systems.

This paper outlines a system developed at Virginia Tech to facilitate spatial audio music production that translates to arbitrary arrays of loudspeakers. Virginia Tech hosts four spatial audio spaces. The pipeline described in this paper was designed for musicians familiar with DAWs to easily mix in these venues, as well as have the capability to move the same mix between venues without having to make significant changes in routing to suit the new space.

This paper is in five sections. The first section details some of the benefits of spatial audio for music production. Second, a description of the DAW Reaper's inherent benefits for multichannel reproduction and plug-ins made with the JSFX programming suite. Third, the spatialization encoder and decoder Max patch is described, including future development opportunities. The fourth section details examples of the transport of audio and data information and decoding schemas. The last section discusses use cases, applications, limitations, and future development.

2 Benefits of Spatial Audio for Reproduction and Reinforcement

One major benefit of spatial audio for both sound reproduction and reinforcement is increased headroom over stereophonic reproduction. A virtue of spatialization, mixes with dozens or hundreds of tracks do not have to compete for headroom in a stereo mix. Moreover, 360 degree panning and elevation offer creative mixing options beyond a fixed perspective.

Most spatial audio reproduction systems are scalable and can contain virtual loudspeaker setups. Stereo, 5.1 and other formats are virtually reproduced within spatial audio frameworks, so previous work can adapt to new experiences in high-dimensional spatial systems. The spatial decoding itself is also scalable, which is addressed in the fourth section of this paper.

Another benefit of spatial audio for sound reinforcement is leveraging distributed audio systems with diffusion for a more controlled acoustic experience than a traditional front of house array.

New possibilities for signal processing are also indicated, including new time-based effects. Re-creating real acoustic spaces, as well as virtual and impossible spaces, with B-Format impulse responses, is one example detailed in this pipeline.

3 DAW Interface

Reaper was chosen as DAW because of the flexible routing, robust support community, and JSFX plug-in development. Project files transfer seamlessly between computers and operating systems.

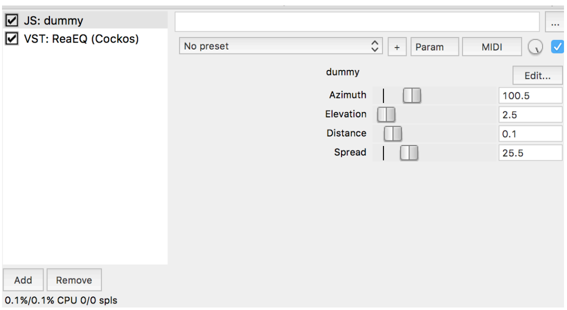

Figure 1. JS plug-in inserted on a track.

A JS plug-in was created to pass control data from Reaper to Max via Open Sound Control (OSC). A four-fader plug-in sends 3-D location data in the format of azimuth, elevation, distance, and spread. This plug-in is inserted on each channel and the channel number corresponds to a point source of the same number in the spatialization engine. It is currently possible to have 256 individual channels as sources, but limited to 64 in practice for most use cases for easy transport between the studios at Virginia Tech. Editing a Pattern Configuration file allows for select data to transfer via OSC.

```
desc: dummy
slider1:0.<0,359,0.1>Azimuth
slider2:0.<0.,90.,0.5>Elevation
slider3:0.1<0.,1.,0.01>Distance
slider4:0.1<0.,100.,0.1>Spread
```

Figure 2. Example JS Code.

Individual audio tracks can include up to 64 channels by default. Master buss effects can be added by bussing individual tracks to a buss. Some JS plug-ins scale to the number of channels on a track for processing of up to 64 channels, making parallel processing busses easy to set up and audition.

4 Max Spatialization

The graphical programming language Max is used as a spatialization engine. Individual audio channels are routed into Max. The version featured uses the ICST Ambisonics Toolkit [2], IRCAM Spat~ [3], and a B-Format convolution reverb designed by Upthegrove as separate mixing busses. Any spatialization suite that accepts Cartesian or polar coordinate data can be implemented.

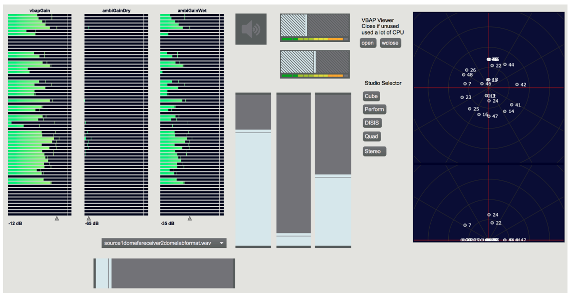

Figure 3. Max spatialization engine graphic interface.

The 3-D VBAP panner from IRCAM's *spat.pan~* object provides precise localization spatialization. This panner includes a *spread* method to increase the number of loudspeakers used, emanating radially from the source.

The ICST Ambisonics suite employs high order ambisonics for another spatialization option, which can be blended with the VBAP spatializer, for an effect similar to parallel compression, which, according to Bob Katz, is to “take a source, and mix the output of a compressor with it.” [4].

The B-Format convolution reverb uses the Huddersfield HISS Toolkit [5] to convolve encoded B-Format signals with B-Format impulse responses. At present, most impulse responses used come from the Open Acoustic Impulse Response Library [6]. Computer generated impulse responses have also been used [7] to gauge listener preferences. In future versions, real-time generated impulse responses generated for virtual environments will be added.

5 Audio and Control Data

The pipeline affords multiple avenues for live input as an alternative to Reaper. Currently, the DAW and spatialization engine are both hosted on the same computer. Each channel in Reaper is mapped to a hardware output. Using the loopback feature of a USB audio interface, each output is returned back to inputs on the same device to the spatialization engine. In practice, the audio hardware output of Reaper is the USB device. The hardware input to the Max spatialization engine is the same USB device. The inputs are mapped to the *adc~* objects in Max, which correspond to the track numbers in Reaper. In lieu of a hardware interface with a loopback feature, this method also works with inter-software transport applications at the expense of processing power. OSC data is shared locally on the computer with matching UDP ports in Reaper and Max.

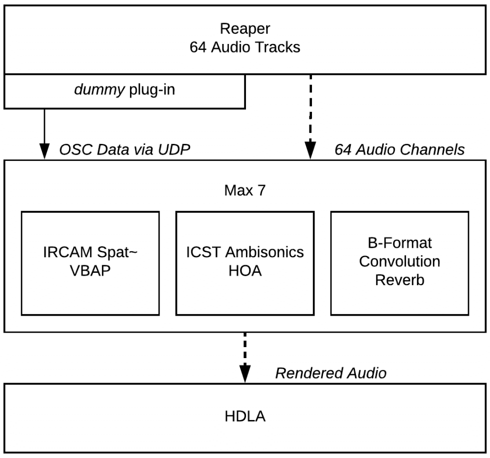

Figure 4. Signal flow from Reaper to Max to HDLA.

The pipeline is also designed to be as source agnostic as possible. Another tested method uses an audio host computer running Reaper on the same local network as the decoder computer running Max. Control data transmitted via UDP, and audio over Ethernet with Audinate Dante Virtual Soundcard makes 64 channel spatialization possible with one Ethernet cable.

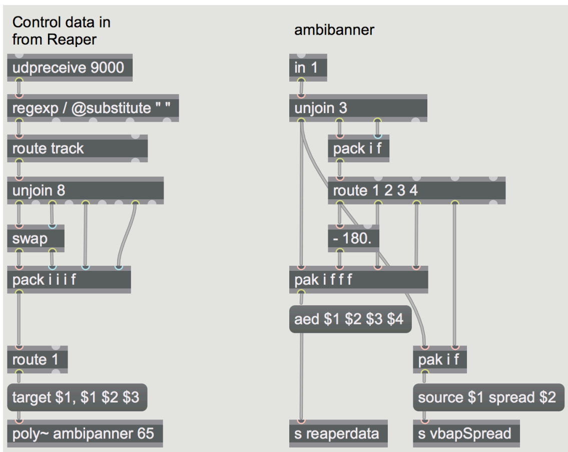

Figure 5. Max OSC data parsing.

Using known loudspeaker coordinates in a venue, the pipeline decodes to any arbitrary loudspeaker array. As VBAP and ambisonics are both used, and each requires a minimum number of loudspeakers for optimal performance, there are limitations. The recommended minimum number of individually addressable loudspeakers is 24 to meet the definition of a HDLA, but the pipeline will function with fewer loudspeakers.

In addition to Reaper, other spatialization data sources have been used, such as motion capture optical tracking, LEAP Motion gesture tracking, MIDI interfaces, and scripting.

6 Use Cases, Applications, and Limitations

This pipeline was first developed to allow rapid translation of spatial audio mixes between four diverse spatial audio venues at Virginia Tech. These venues are The Cube with a 134.6 channel array, the ASPIRe lab with a 58.2 channel array, the Perform Studio with a 24.4 channel array, and the Digital Interactive Sound and Intermedia Studio (DISIS) with 16.2 channels.

Upthegrove has utilized this pipeline for his compositions *rabies* and *The Jury*. Upthegrove has also brought in external collaborators to use the pipeline, and often uses components of the pipeline to meet the needs of visiting researchers and artists.

Figure 6. Virginia Tech Percussion Ensemble with performers on the theatrical catwalks. Photo courtesy Susan Sanders, Virginia Tech.

Resonance on the Walls II featured Annie Stevens and the Virginia Tech Percussion Ensemble for an hour-long concert of live percussion and fixed media in the Cube in February 2017. Percussionists were stationed at various locations in the Cube. Their instruments were amplified with microphones that fed into the spatialization pipeline. Microphone signals were encoded with their actual spatial positions and convolved to simulate real acoustic spaces, which changed for each musical piece performed. Drew Wordin's piece *Volume of a Cube,* was composed specifically for this concert, and reprised May 2017.

The first exhibition outside of Virginia Tech was Moogfest 2017 where Roan and Upthegrove designed a mobile 25.4 channel HDLA. The spatialization pipeline was used for live performances, listening sessions, and workshops.

Family-oriented workshops featured Moog synthesizers with a variety of controllers so audience members could try performing live spatialization Listening sessions had up to 100 people at a time listening to spatial audio examples.

Figure 7. Roan and Upthegrove at Moogfest 2017. Photo courtesy Charles Nichols.

Auragami, a special fixed media concert at Virginia Tech's CubeFest, featured the music of Secret Chiefs 3 remixed from stereo to the Cube's 134.6 HDLA in August 2017. Composer and bandleader

Trey Spruance worked with Upthegrove over a three-day residency to unfold stereo mixes. "Where we are on literally virgin earth is in breaking out 100+ channel mixes of orchestrated musical instruments, with all of their raw, untainted frequency characteristics into 64 discreet channels — to be heard anew in much deeper spatial relationships. Without need for traditional group (bus) compression used for shoehorning complex multiple instrument elements down to size in a 2-channel mix, a new universe of relationships opens up," notes Spruance [8].

Upthegrove collaborated with Stephen Vitiello for a commissioned sound art installation for the opening of Virginia Commonwealth University's Institute for Contemporary Art in Richmond, Virginia. The installation, titled *whether there was a bell or whether I knocked*, is inspired by *The Garden of Forking Paths* by Jorge Luis Borges. The pipeline was used to edit and sculpt 3-D sound during a week-long engineering session at ICAT's Perform Studio, allowing easy editing and modification. Spatial control data was partially generated by image analysis created by Virginia Tech School of Visual Arts Assistant Professor Zach Duer with Processing. The image selected was a Borges self-portrait, created in later stages of life when overtaken by blindness.

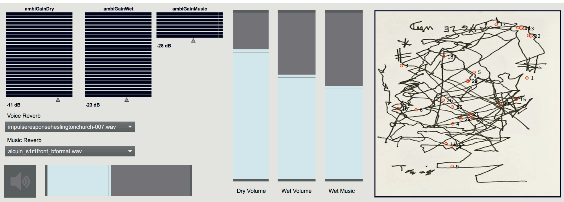

Figure 8. Max spatialization engine with image analysis of a self-portrait by Borges, gleaning spatial trajectories.

One last example is a theatrical production for Thomas Murray's documentary drama *The Right of Way*. *The Right of Way* is a new documentary play that explores the true story of Hector Avalos, who was killed by a drunk driver as he rode his bicycle home from work. Actors and spatial audio trace the true story of the fallen bicyclist through interviews with transportation engineers, urban planners, lawyers, and bicycling activists. Upthegrove designed sounds simulating what it may be like to be on a bicycle while cars zoom by, followed by a crash, at the beginning the play to immerse audiences.

Planned future use cases include mixing for film, virtual reality, and non-linear immersive environments.

A major limitation of the system is the computational power needed. For large systems where dozens or hundreds of channels are rendered in real time for HDLAs, significant load is put on the CPU. The convolution reverb currently has been tested with third order ambisonics. As higher order impulse responses are available, there will be increased demand for computational power. One way to improve performance for the future is to bundle the spatialization engine in a plug-in format that runs in the DAW, saving computational power with all audio rendered locally instead of transporting between applications. Moreover, the plug-in must be the first plug-in for each track in Reaper.

Room acoustics, loudspeaker type, and calibration are not included in the scope of this paper, yet are necessary considerations.

References

[1] E. Lyon, T. Caulkins, D. Blount, I. I. Bukvic, C. Nichols, M. Roan, T. Upthegrove, "Genesis of the Cube: The Design and Deployment of an HDLA-Based Performance and Research Facility." *Computer Music Journal,* vol. 40, issue 4, pp. 62– 78

[2] J. Schacher, P. Kocher, "Ambisonics spatialization tools for max/msp." *Omni* 500.1 (2006).

[3] URL: http://forumnet.ircam.fr/product/spat-en/

[4] R. A. Katz, *Mastering Audio: the art and the science, Second Edition*, Focal Press, p. 133 (2002).

[5] A. Harker, P. A. Tremblay, "The HISSTools impulse response toolbox: Convolution for the masses." *Proceedings of the International Computer Music Conference*. The International Computer Music Association (2012).

[6] URL: http://www.openairlib.net/

[7] M. Ermann, A. Hulva, T. Upthegrove, R. J. Rehfuss, W. Haim, A. Kanapesky, T. Mcmillon, C. Park, A. Reardon, J. Rynes, S. Ye, C. Nichols. “Subjective listening tests: Perception and preference of simulated sound fields.” *The Journal of the Acoustical Society of America*, vol. 141, issue 5 (2017).

[8] URL: http://www.webofmimicry.com/cubefest/