Audio Engineering Society

Convention e-Brief

Presented at the 139th Convention
2015 October 29–31 New York, NY, USA

Decorrelated Audio Imaging in Radial Virtual Reality Environments

Bryan Dalle Molle, James Pinkl, and Mark Blewett

Electronic Visualization Laboratory, University of Illinois at Chicago, Chicago, IL, 60607, USA
bryandallemolle@gmail.com
jimmypinkl@yahoo.com
markblewett@hotmail.com

ABSTRACT

University of Illinois at Chicago's CAVE2 is a large-scale, 320-degree radial visualization environment with a 360-degree 20.2 channel radial speaker system. The purpose of our research is to develop solutions for spatially accurate playback of audio within a virtual reality environment, reconciling differences between the circular speaker array, the location of a user in the physical space, and the location of virtual sound objects within CAVE2's OmegaLib virtual reality software, all in real time. Previous research presented at AES 137 detailed our work on object geometry, dynamically mapping a virtual object's width and distance to the speaker array with volume and delay compensation. Our recent work improves virtual width perception using dynamic decorrelation with transient fidelity, implemented via Supercollider on the CAVE2 sound server.

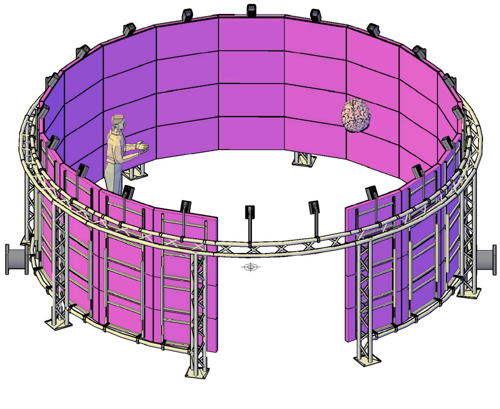

Figure 1 CAVE2 as located at the University of Illinois at Chicago's Electronic Visualization Laboratory

1. INTRODUCTION

CAVE2 has in place a large multichannel sound system consisting of 20 two-way, near-field reference monitors equally spaced around the circumference of the environment, as well as a pair of companion subwoofers (Figure 1). The sound system is fed from a SuperCollider audio server that uses the PanAz UGen [1] as a foundation for panning sound objects. The server, in turn, receives instructions from OmegaLib, CAVE2's custom middleware, sent via the proprietary Omicron SoundAPI as OSC messages. At the top control level, users create virtual environments using the OmegaLib Python API. At run time, the Omicron SDK receives information from the wand controller and tracking system, providing the SoundAPI with the user location data essential for dynamic audio imaging.

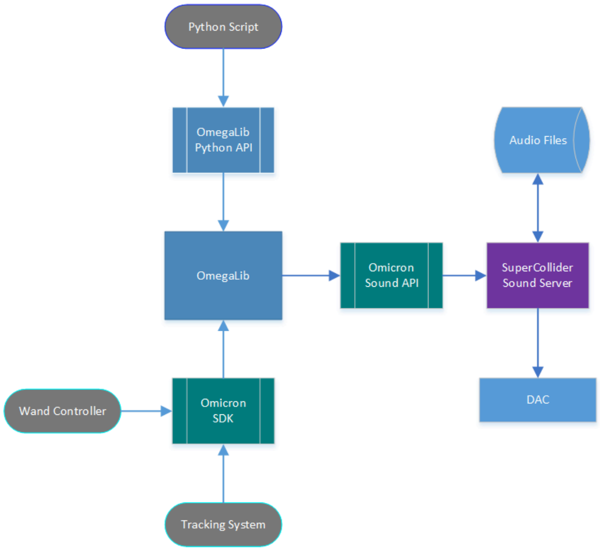

Figure 2 High-level representation of control in the CAVE2 sound system.

1.1 CAVE2 Spatial Techniques

Currently, three parameters are used to maintain spatial perception as a user and object move about the CAVE2:

1. The direction of the sound source relative to the user determines which speaker(s) should play back the sound (Figure 3).

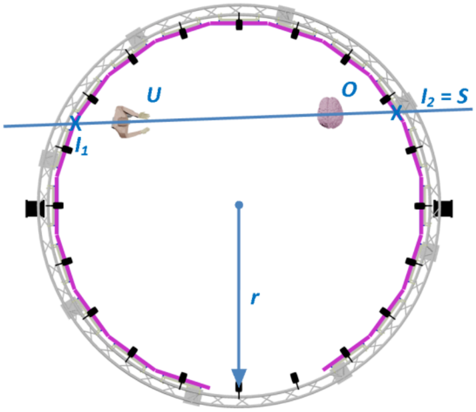

Figure 3 The known geometry of the environment and line UO are used to determine the proper point along the CAVE2's wall to which the sound is centered.

2. The source volume adjusts the amplitude of playback based on user and object location.

$$1 \geq V_{dynamic} = V_{defined}\left(\frac{1}{\left|\overrightarrow{UO}\right|}\right)$$

3. The width of the audio object is used to dynamically alter the number of speakers playing back an object's audio.

$$1 \leq W_{dynamic} = \frac{W_{defined}}{\left|\overrightarrow{UO}\right|} \leq 20, \left|\overrightarrow{UO}\right| \geq 0.25m$$

The source width varies linearly in inverse proportion to the user's distance from the object. The algorithm limits the dynamic width to values between 1 and 20. The minimum object to user distance of 0.25 meters was experimentally determined to provide reasonable results [2].

1.2 Problems with Width

When a sound object increases in width and spreads to adjacent speakers, spatial accuracy begins to degrade. When identical signals are played from two or more speakers, the signals are perfectly correlated, and as a result a phantom image appears in the middle of the space between the two speakers. As a result, when a sound object increases in width and spreads to it's neighboring speakers, the perceived source remains near the center of the speaker array rather than spanning the width of the speaker array. Additionally, interference between correlated signals emanating from neighboring speakers causes timbral coloration, as combing results when identical signals with different amounts of delay combine at the listener [3].

1.3 Decorrelation

Controlling the amount of correlation, a measure of similarity between two signals, is an effective method to prevent the negative effects of wide sounds. The correlation parameter is a normalized value between -1 and 1, where 0 indicates no correlation. A signal of large width that has been decorrelated will be perceived as spanning the speaker array. The principal method used to decorrelate a signal is to introduce small randomized phase shifts or time delays to the source before playback.

1.4 The Applause Problem

The addition of randomized time delays prevents negative effects for steady sounds only. Because transients behave like an impulse, these sounds are susceptible to temporal smearing. For example, playback of applause, is a particularly difficult signal to treat because it contains both the transients of clear claps as well as the more steady state sound of distant claps [4]. The method we aimed to implement was one that extracted transients and treated the remaining, steady-state sounds separately.

2. METHODS

2.1. Dynamic Decorrelation

We initially investigated implementing our decorrelator using the critical band approach proposed by [5]. In this method, the source signal is decomposed into a 24 partitions, referred to as critical bands, by filtering the source through an equivalent rectangular band filter bank. Randomized delays upper-bound by the band's wavelength were applied to each partition before being summed back together as output. However, with 24 critical bands, this approach became too computationally expensive for SuperCollider to handle a single sound object in real time.

As a result, we decided to use an existing implementation in the 3rd party PV_Decorrelate UGen for its simplicity and real-time efficiency. PV_Decorrelate is a reimplementation of SuperCollider's built in diffuser, the PV_Diffuse UGen, with an additional scaling factor. PV_Decorrelate operates in the frequency domain, taking an input FFT chain and applying scaled randomized phase shifts to each frequency bin [6]. In our application, we scale each bin by a decorrelation coefficient between 0 and 1, derived from the object width.

2.2. Transient Fidelity

To retain transient fidelity when decorrelation is introduced required us to develop a specialized signal processing routine, called a Unit Generator or UGen in SuperCollider, for transient extraction. Our transient extraction UGen is derived from an algorithm to improve decorrelation in 5.1 surround sound systems proposed in [7]. At a high level, the algorithm takes as an input single time-domain source, separates transient material from the steady-state signal upon detection, and then fades the transient frequencies back into the steady-state signal with a release. The steady-state signal is then decorrelated, while the transient material is randomly panned to one of the five output speakers. For our purposes, it was undetermined as to the appropriate treatment of the transient signal, so the transient signal is simply sent unaffected to the appropriate output speakers based on object location and width.

Transients are detected by analyzing the signal with the Short Time Fourier Transform (STFT) with small window sizes and large overlap between frames. In our implementation, STFT frames contained 128 samples (2.7 ms) with 50% frame overlap. A signal is said to contain a transient if a smoothed average of magnitude for a set of adjacent frequency bins (Figure 4) crosses a user-defined threshold (Figure 5). The source code for the plugin can be viewed at https://github.com/bdallemolle/PV_TransExtract.

$$
\begin{aligned}
\Omega[s,k] &= \min_{j=0}^{J-1}(\omega[s-j,k]) \\
\text{where } \omega[s,k] &= \frac{1}{I}\sum_{i=1}^{I}(|X[s-i,k]|)
\end{aligned}
$$

Figure 4. Algorithm to compute smooth averages of frequency bin *k* at time slice *s* from source signal *X*. Parameters *I* and *J* are user defined determine the number of previous bins to average [8].

$$|X[s,k]| > \alpha\Omega[s,k]$$

Figure 5. The final comparison determining a transient, with user parameter α and the smoothed average Ω from Figure 4 [8].

3. FURTHER WORK

Though our work has been implemented and tested at various stages in the CAVE2, a more formal assessment of our methods should be conducted to determine effectiveness from a subjective user's perspective. Test scenarios have been created and evaluated, but should be expanded to test a variety of different sound sources to determine if our solutions are general enough to handle arbitrary audio playback in the CAVE2. We plan on conducting formal user studies to present informative results of effectiveness.

More to the last point, initial testing of our transient extraction algorithm has demonstrated the need to optimize parameters. The parameters suggested in [8] provide a good starting point. However, while these values provide reasonably good performance on sources like applause, timbral changes in sources like a multi-track a rock song are more audible. An ideal solution would include dynamic analysis of the "transientness" of the source, adjusting parameters to best suite the source material.

And finally, the reflective properties of the CAVE2 present a challenge to any spatialization techniques in the space. Analyzing the acoustic properties of the space in an effort to produce filters to reduce the effects of the interference from reflections is planned.

4. ACKNOWLEDGEMENTS

This work was supported by the Electronic Visualization Laboratory at the University of Chicago at Illinois. Graphics drafted by Jonas Talandis. Special thanks to Maxine Brown, Dr. Andrew Johnson, Arthur Nishimoto, and Alessandro Febretti. Also, thank you to Zachary Seldess and Michelle Daniels of Sonic Arts Research and Development, UCSD, for additional support.

5. REFERENCES

[1] Reference to SuperCollider help file URL: https://github.com/supercollider/supercollider/blob/master/HelpSource/Classes/PanAz.schelp

[2] M. Blewett, J. Pinkl, and B. Dalle-Molle: "Dynamic Audio Imaging in Radial Virtual Reality Environments" *Audio Engineering Society 137th Convention,* e-Brief 162, 2014

[3] G. Kendall, "The Decorrelation of Audio Signals and Its Impact on Spatial Imagery" *Computer Music Journal* vol. 19, no. 4, 1995, pp. 71-87

[4] R. Penniman, "A General-Purpose Decorrelation Algorithm with Transient Fidelity" *Audio Engineering Society 137th Convention,* Convention Paper 9170, 2014

[5] M. Boueri and C. Kyirakakis, "Audio Signal Decorrelation Based on a Critical Band Approach" *Audio Engineering Society 117th Convention,* Convention Paper 6291, 2004

[6] Reference to Beast Mulch SuperCollider Libraries: http://www.birmingham.ac.uk/facilities/ea-studios/research/mulch.aspx

[7] R. Penniman, "A High-Quality General-Purpose Decorrelator with Transient Fidelity" Master's Thesis, *University of Miami,* 2014

[8] N. Juillerat, S. M. Arisona, and S. Schubiger-Banz, "Enhancing the Quality of Audio Transformations Using the Multi-Scale Short-Time Fourier Transform in *Proceedings of the 10th IASTED International Conference, Signal and Image Processing*, 2008, pp. 379-387