A SuperCollider-Based Spatial Audio Performance System with AR Interfaces

Hyunkyung Shin
Georgia Institute of Technology
hshin336@gatech.edu

Henrik von Coler
Georgia Institute of Technology
hvc@gatech.edu

ABSTRACT

The presented real-time interactive spatial audio performance system is designed to connect an Augmented Reality (AR) interface with a musician or performer. This system enables dynamic sound manipulation and creates an immersive spatialization experience. It is built on SuperCollider and processes multi-source audio signals. OSC (Open Sound Control) and MIDI-based tracking are integrated to enhance interaction. Higher-Order Ambisonics (HOA) is employed for spatial audio rendering. The system allows the performer's movements to influence sound positioning, spectral transformation, and dynamic modulation through real-time motion tracking. A modular structure supports expansion with additional sensors, synthesizers, or external controllers. This adaptability makes the system suitable for virtual concerts, multimedia installations, and immersive experiences.

1. INTRODUCTION

Augmented Reality (AR) has expanded the possibilities of interactive performances by merging virtual and physical elements in real-time. Despite advancements in AR technology, controlling spatial audio in AR-based performances remains a significant challenge. This is primarily due to latency issues, complex spatialization algorithms, and the necessity for seamless interaction between performers and digital sound environments. In particular, real-time sound manipulation requires a system capable of capturing movement data, processing it efficiently, and applying it to a spatial audio engine with minimal latency. To address these challenges, this paper presents a SuperCollider-based spatial audio performance system that connects an AR interface with performers and enables real-time manipulation of sound through movement and virtual interaction. The system provides direct control over spatial audio rendering. OSC-based motion tracking captures positional and gestural data and maps it to real-time sound transformations. MIDI control mapping allows performers to interact with external synthesizers and hardware. Higher-Order Ambisonics (HOA) processes audio for three-dimensional spatial rendering and ensures the realistic positioning of sound sources in an immersive environment. Advanced sound design modules for AR performers apply spectral transformation, dynamic filtering, and spatial effects to enhance the performer's expressive capabilities. This study demonstrates how integrating motion tracking, spatial audio processing, and dynamic sound design enhances real-time audio interaction in AR environments. The system expands the expressive potential of performers, allowing them to manipulate spatial sound intuitively in virtual spaces. Beyond live performances and interactive installations, this approach paves the way for future applications in AR-based sound design and immersive musical experiences.

2. BACKGROUND AND RELATED WORKS

2.1 Computer-based Performance System

2.1.1 A computer music performance system (CMPS)

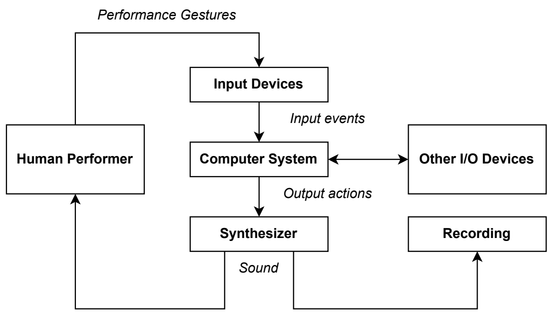

Figure 1. The components of a computer music performance system (Anderson and Kuivila, 1990).

A computer music performance system (CMPS) defines the fundamental structure of a computer-based performance system [1]. It comprises input devices, a control computer, synthesizers, and additional I/O devices, converting a performer's gestures into digital data for real-time processing and output. Input devices include general-purpose computer peripherals, MIDI keyboards and other traditional instrument-based controllers, and specially designed musical input devices. The control computer processes these inputs to generate sound control commands, such as note on/off signals and continuous timbre adjustments. Synthesizers receive control commands via the MIDI protocol, generate audio signals, and deliver them to performers and audiences. They can also support recording when required.

2.1.2 Interactive Music System (IMS)

Interactive Music System (IMS) extends traditional musical interfaces by providing a 3D performance experience with a standard musical keyboard as the primary controller [2]. A multi-track looper enables the creation of multiple sound sources, referred to as Musical Agents, which function as loops of musical material. Real-time sound synthesis allows flexible timbre control through the physical interface. The performer's body and surrounding space are integral to the system, requiring sufficient physical room for interaction with sound sources. An ambisonic speaker array implements spatial audio, delivering an immersive sound experience without the need for wearable devices. The audio position of sound sources is visually represented in 3D space using MR technology. Motion tracking enables users to manipulate sound sources as physical objects and map their movements to target locations. Musical agents can operate autonomously, dynamically modifying musical lines based on user input and interactions with other agents while maintaining the performer's playing style. Users can control autonomy through the 3D visual interface, enabling or disabling autonomous behavior as needed.

2.2 Augmented Reality in Performance

Augmented Reality (AR) enables new forms of artistic expression and audience engagement in performance. In the field of music, AR-based interactive visualization and composition systems have been developed to facilitate real-time interaction with music by manipulating virtual elements through hand tracking or marker recognition [3, 4].

In theatrical performance, AR has been utilized to enhance storytelling through real-time digital overlays and interactive stage design. Recent studies have proposed technologies that allow performers to directly control AR-based visual elements and synchronize digital and physical components [5].

Additionally, Audio Augmented Reality (AAR) technology employs spatialized soundscapes to guide audience interaction, providing an immersive and participatory performance experience [6]. As a medium that connects physical spaces with virtual environments, AR suggests the potential to reshape the structure of performance. With further advancements in AR-based interaction technologies, its applications in performance are expected to expand [7, 8].

2.3 OSC-based Audio Connection

Open Sound Control (OSC) is a protocol widely utilized in music technology for real-time sound manipulation and spatial audio implementation. It is particularly effective in transmitting data between devices, enabling real-time control over sound parameters [9]. Due to these characteristics, OSC plays a crucial role in interactive audio and spatial sound design. OSC enables real-time interaction with sound through gesture-based control, enhancing musical expressiveness in live performances [10]. It is also a powerful tool for real-time sound synthesis, allowing for the manipulation of synthesis parameters across multiple devices in distributed music systems. In spatial audio applications, OSC is used to manage and control audio positioning [11] and facilitates precise sound placement [12]. Additionally, it is essential in 3D environments for effective sound source localization, ensuring accurate spatial perception of audio sources in immersive settings [13]. Furthermore, OSC enhances networked music performances by enabling multi-device synchronization and real-time collaborative sound manipulation[14].

3. SYSTEM DESIGN

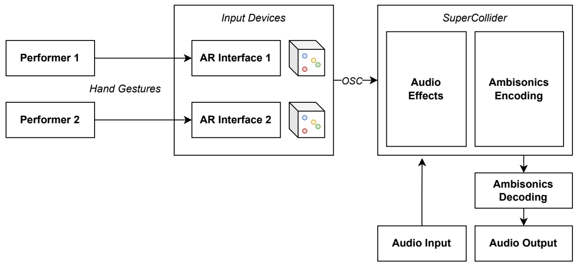

Figure 2. AR performance system with two AR Interfaces

The performance system integrates augmented reality (AR) interfaces for real-time, gesture-based control. It consists of input devices, audio processing modules, and an audio output stage, allowing performers to control spatialization using hand gestures. Motion data from the AR interfaces is transmitted via Open Sound Control (OSC) to SuperCollider, where it undergoes spatial audio processing and encoding. The processed signals are then decoded and output through loudspeakers, as shown in Figure 2.

3.1 AR Interface

Figure 3. 6DOF ARCube.

The performance system for AR performance integrates two AR interfaces to support a multi-performer environment. Performers interact with the AR interface by wearing a Head-Mounted Display (HMD). Performers can freely manipulate a cube and four objects color-coded in red, green, blue, and yellow using pinch and grab gestures (Figure 3). The cube transmits the 6-degree-of-freedom (6DOF) position data of each object via OSC.

Listing 1. OSC connection with AR Interface for xyz.

```
OSCdef('/quest/1/xyz',
    {
        arg msg, time, addr, recvPort;
        var x,y,z;
        var idx = msg[1]-1;
```

```
        ~quest_xyz_BUS[idx].setAt(0, msg[2]);
        ~quest_xyz_BUS[idx].setAt(1, msg[3]);
        ~quest_xyz_BUS[idx].setAt(2, msg[4]);
    }, '/quest/1/xyz');
```

3.2 OSC and MIDI Communication

OSC connections enable dynamic control of sound parameters by mapping external positional data from two AR interfaces and MIDI control signals to internal SuperCollider buses (Listing 1).

3.2.1 OSC Connection

The OSC (Open Sound Control) module is responsible for receiving and processing real-time motion-tracking data from an external device. For AR performance, OSC is used for connecting 2 AR interfaces with SuperCollider. The performance system listens for incoming OSC messages and maps them to relevant control buses, which enabled spatial tracking, positional modulation, and movement-based sound transformations.

Specifically, 6DOF data based on real-time interaction with objects are sent via OSC and mapped to each bus in SuperCollider (Table 1).

OSC Message	Data	Bus Assignment
`/quest/1/aed`	Azimuth, Elevation, Distance	`quest_aed_BUS[idx]`
`/quest/1/xyz`	X, Y, Z Position	`quest_xyz_BUS[idx]`
`/quest/1/pry`	Pitch, Roll, Yaw	`quest_pry_BUS[idx]`

Table 1. Core Sound Effects for four different virtual objects within ARCube.

3.3 Sound Design

The system integrates two different modules for each instrument for sound mapping, based on the 6 degrees of freedom of ARCube. two modules designed to be compatible with double bass and modular synthesizer. As shown in Table 2, the sound design integrates sound effects based on the location of four different sound sources.

Virtual Objects	Sound Effects
RED	Pitch and Spatial Diffusion.
GREEN	Feedback and Delay
BLUE	Bit-Crush and Freeze Effects
YELLOW	Grain Length and Density

Table 2. Core Sound Effects for four different virtual objects within ARCube.

To optimize sound processing according to the input characteristics of acoustic instruments, `/SynthDefs` are commonly used but also differ in certain aspects.

The following `/SynthDefs` are examples of common modules used for both instruments:

- **`frequency_modulation`**: Perform frequency modulation and wavefolding distortion.
- **`spectral_delay`**: Apply delay-based spectral diffusion.

However, `/SynthDefs` remain adaptable to different processing needs. For example, double bass and modular synthesizer share feedback and delay diffusion effects but use different parameters. In the `quad_pdf` synthesizer, both modules apply delay-based spectral diffusion, but the variation is controlled differently by the `delaybus` and `feedbackbus` parameters.

In `double_bass.scd`, the absence of explicit feedback constraints allows for more flexible modulation and broader diffusion effects. However, in `modular_synth.scd`, the `feedbackbus` suggests that delay feedback is tightly controlled, reducing the extent of time-dependent spectral diffusion.

3.4 Spatial Rendering

For the performance system, spatialization is implemented using SuperCollider with the Higher Order Ambisonics (HOA) class and multi-channel bus processing, which enable sound to be positioned or moved within a defined spatial field.

The following `/SynthDefs` serve as examples of spatialization implementations:

- **`hoa_mono_encoder`**: Encoding sound using HOA coefficients, allowing for detailed spatial localization.
- **`spatial_splitter`**: Handle spatial dispersion and multi-channel Ambisonic processing.

This spatialization technique proved effective in generating real-time synthesis based on gesture input within an AR interface. Spatialization modifies sound using position data obtained from real-time interactions between objects and gestures in the AR environment. The integration of OSC and SuperCollider facilitated the creation of new sound effects by utilizing gesture speed as a control parameter. Additionally, distinct sound effects were generated based on changes in input position with clearly distinguishable characteristics. As shown in Figure 4, when the green object moves diagonally from the center of the left face to the center of the top face, an increase in speed produces a modified sound effect resembling wind. For the yellow object, placing it at the left corner generates a small breath-like sound through a fast delay effect, while positioning it at the middle of the right side produces a breath effect similar to a strong wind sound.

4. FUTURE WORKS

Future research will focus on enhancing the technical performance and musical interaction of the AR performance system. System improvements will include integrating new gestures to refine Higher Order Ambisonics (HOA) spatialization. Optimizing sound effects will enable more precise and dynamic spatial diffusion by adapting to the input characteristics of various instruments. Developing flexible mappings between gesture data and sound parameters will allow performers to control sound more intuitively and expressively.

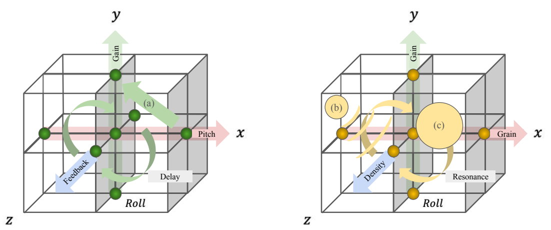

Figure 4. Distinct sound effects achieved through spatialization: (a) speed-based effects, (b) A location-based small breath effect, and (c) A location-based big breath efffect.

For creative applications, the system will be expanded to support collaborative manipulation of spatialized sound objects, enabling both performers and audiences to interact within the AR environment. Implementing an adaptive machine learning-based gesture recognition model will provide an interaction framework tailored to individual movement styles. These advancements will establish AR-based music performance as a more versatile and expressive medium for live concerts and interactive installations.

5. CONCLUSION

For augmented reality (AR) performances, the system integrates OSC-based motion tracking, SuperCollider-based synthesis, and Higher Order Ambisonics (HOA) for dynamic spatial sound rendering. The proposed framework enables performers to manipulate sound through AR-based gestures, creating an immersive and expressive performance environment. The modular structure allows for adaptability across different instruments, which ensures flexibility in sound design. Experimental validation demonstrated that gesture speed and object positioning significantly influenced spatialized sound effects and enhanced real-time interaction. The findings highlight the potential of AR-driven spatial audio in live performance contexts and offer new possibilities for gesture-based sonic interaction and multi-performer collaboration.

6. REFERENCES

[1] D. P. Anderson and R. Kuivila, "A system for computer music performance," *ACM Transactions on Computer Systems (TOCS)*, vol. 8, no. 1, pp. 56–82, 1990.

[2] P. Lucas and S. Fasciani, "A Human-Agents Music Performance System in an Extended Reality Environment," in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. NIME, 2023.

[3] G. S. R. Reddy and D. Rompapas, "Liquid Hands: Evoking Emotional States via Augmented Reality Music Visualizations," in *ACM International Conference on Interactive Media Experiences (IMX '21)*. ACM, 2021, pp. 305–310.

[4] K. Poddar and B. Sharma, "Enhancing Music Composition with Marker-Based Augmented Reality Technology," in *2024 4th International Conference on Sustainable Expert Systems (ICSES)*. IEEE, 2024, pp. 1319–1323.

[5] D. Lisowski, K. Ponto, S. Fan, C. Probst, and B. Sprecher, "Augmented reality into live theatrical performance," in *Springer Handbook of Augmented Reality*. Springer, 2023, pp. 433–450.

[6] A. N. Nagele, V. Bauer, P. G. T. Healey, J. D. Reiss, H. Cooke, T. Cowlishaw, C. Baume, and C. Pike, "Interactive Audio Augmented Reality in Participatory Performance," *Frontiers in Virtual Reality*, vol. 1, 2021.

[7] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[8] W. Li, "Musical Instrument Performance in Augmented Virtuality," in *6th International Conference on Digital Signal Processing (ICDSP)*. ACM, 2022, pp. 91–98.

[9] M. Wright, "Open Sound Control: an enabling technology for musical networking," *Organised Sound*, vol. 10, no. 3, pp. 193–200, 2005.

[10] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," *Computer music journal*, vol. 26, no. 3, pp. 11–22, 2002.

[11] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt, "Rendering spatial sound for interoperable experiences in the audio metaverse," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–15.

[12] M. Zbyszyński, H. Déjardin, D. Marston, and G. Le Nost, "ADM-OSC: an industry initiative for communicating object-based audio data," in *2023 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2023, pp. 1–6.

[13] Y. Masuyama, Y. Bando, K. Yatabe, Y. Sasaki, M. Onishi, and Y. Oikawa, "Self-supervised neural audio-visual sound source localization via probabilistic spatial modeling," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 4848–4854.

[14] A. Schmeder, A. Freed, and D. Wessel, "Best Practices for Open Sound Control," in *Proceedings of the Linux Audio Conference (LAC)*. Utrecht, Netherlands: LAC, 2010.