

2CUBES: A Multi-User Augmented Reality Performance

Hyunkyung Shin

School of Music

Georgia Institute of Technology

hshin336@gatech.edu

Henrik von Coler

School of Music

Georgia Institute of Technology

hvc@gatech.edu

ABSTRACT

2CUBES is a multi-user augmented reality (AR) performance that explores new modes of musical interaction through spatialized sound and gesture. The performance features the ARCube, an AR interface that allows performers to manipulate virtual sound objects in real time using head-mounted displays (HMDs). Two musicians performing on double bass and modular synthesizer use hand gestures to control audio parameters and integrate the AR interface into their instrumental expression in improvisational collaboration. The performance creates an immersive audiovisual environment where sound can be layered, transformed, and spatially projected within a shared performance space. 2CUBES was presented in two public performances and reflects on the artistic and practical implications of AR-based interaction in live-electronic music performance.

1. INTRODUCTION

Augmented Reality (AR) offers new modes of interaction in musical performance by layering virtual elements onto physical space [1, 2]. In particular, AR enables performers to engage with sound in spatial and embodied ways through gesture-based interaction and visual feedback. While many AR systems in musical contexts emphasize solo performance or visual augmentation, fewer have examined how AR can support real-time

Copyright: ©2025 Hyunkyung Shin et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License 3.0 Unported](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

collaborative music-making with physical instruments. 2CUBES is a multi-user AR performance that explores how AR-based interaction can extend musical expression and improvisation between two performers. The performance uses the ARCube [3], an interface developed with Unity and SuperCollider, which allows each performer to manipulate virtual sound objects using hand gestures while playing their instrument. One performer plays a double bass, and the other controls a modular synthesizer, with both interfacing simultaneously with their own ARCube system.



Figure 1. ARCube with modular synth during 2CUBES performance.

Although each performer sees a different AR scene through head-mounted displays (HMDs), their actions influence a shared spatial sound environment in real time. The performance highlights how individualized AR interfaces can co-exist within a co-located setting and contribute

to an immersive sonic experience. This paper presents the design of the duo performance and examines how AR-based interaction shaped the performers' embodied engagement, based on insights from two public performances.

2. BACKGROUND AND RELATED WORK

Digital interfaces have been studied as tools to enhance collaboration among performers and expand musical expressiveness. Real-time manipulation of spatialized sound creates unified auditory experiences and emphasizes the role of spatial interaction in musical collaboration and creativity [4].

Multi-agent systems in virtual 3D environments enable rhythmic improvisation and the emergence of creative musical structures through self-assembly and synchronization processes. These systems have potential applications in AR environments [5]. AR has been identified as a key technology for reconfiguring traditional performance spaces. AR applications in stage performances provide performers with interactive tools and offer audiences immersive, visually enhanced environments [1].

Extended reality reveals the potential of AR to merge physical and virtual performance experiences. Technologies such as spatialized audio and HRTFs enable immersive and intuitive interactions between users and virtual objects [6]. Notably, AR interfaces provide performers with tools for spatial sound control, a critical parameter in modern music. Real-time control systems have significantly expanded the scope of musical performance practices. However, seamless integration of spatial control and sound generation requires carefully designed hardware and interfaces [7].

AR interfaces extend traditional music creation workflows by introducing interactive and exploratory elements. They create immersive music production environments and open new possibilities for artistic expression [8]. Multi-agent systems in mixed reality environments further enhance creative possibilities. These systems integrate performers, digital instruments, and virtual objects to create immersive and interactive musical settings [2].

Networked music performances using AR systems demonstrate the ability to maintain high-

quality audio experiences while supporting distributed collaboration [9]. AR interfaces seamlessly connect physical and virtual elements. They enable creative and expressive musical interactions through gestures and spatial interaction [10].

Previous studies demonstrate the transformative potential of AR in musical performances. AR technologies enhance user interaction, improve immersion, and overcome physical limitations through spatial interaction [11, 10]. Extended reality systems also integrate multi-user interaction and dynamic audio-visual mapping to provide immersive and collaborative experiences [12].

Based on this foundation, 2CUBES employs AR interfaces as a creative musical interface for performers. 2CUBES allows performers to manipulate spatialized audio in real time through HMDs and a multi-channel sound system. This study examines the potential of AR-based interfaces for multi-user interaction, live audio processing, and collaborative performance in AR.

3. SYSTEM DESIGN

3.1 ARCube Interface

The ARCube is an AR interface designed for three-dimensional spatial control [3]. It represents the physical space surrounding the user and allows for intuitive understanding and manipulation of spatial data. Users can interact with virtual sound objects using grab and pinch gestures to enable real-time spatial audio rendering. In 2CUBES, the visual design of the original ARCube was partially modified to enhance performers' intuitive recognition of the interface during live performance. As shown in Figure 1, the cube's surfaces are rendered semi-transparent to preserve the external appearance, so that internal movement can be observed. The previous cardinal direction labels on the cube's faces have been replaced with a light blue sphere positioned on the top and an arrow on the front side. The X and Y axis lines remain visible to support spatial orientation.

The original XYZ controller has been extended to support six degrees of freedom (6DoF). Virtual objects now take the form of tree-like structures with arrows representing pitch, roll, and yaw. These structures are 4 cm in length and retain the original red, green, blue, and yellow color coding.

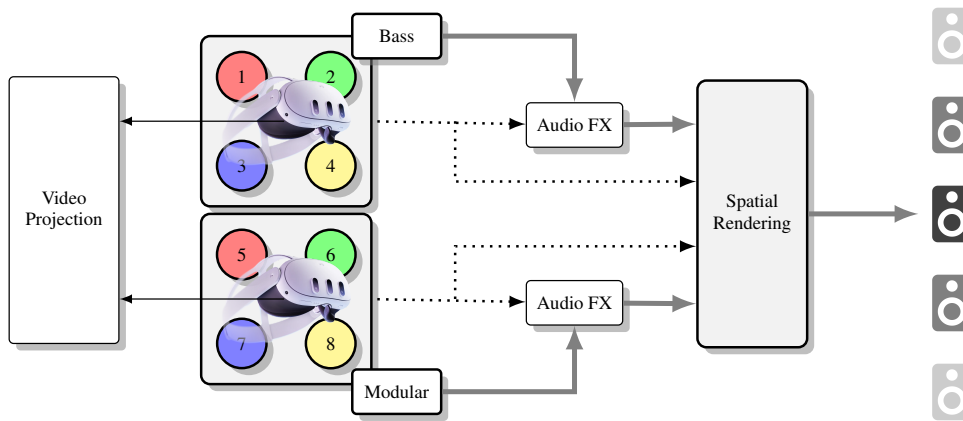


Figure 2. Signal flow for 2CUBES (dotted=OSC, gray=audio, black=video).

As with the original ARCube, performers can manipulate the cube or internal virtual objects using one or both hands. They can align the cube's orientation to their preferences, improving integration with physical objects such as musical instruments and enabling more precise, intentional control. The AR interface was implemented using the Unity engine (version 2022.3.22f1). Open Sound Control (OSC) messages were used to transmit orientation and position data to the sound system.

3.2 Sound Design and Interaction

The 2CUBES performance is based on a SuperCollider-based system that utilizes two AR-Cubes as musical agents [13]. Each performer wears an HMD and independently controls a sound layer tailored to their instrument using one ARCube (Figure 2). In this system, performers achieve real-time spatialization and modulation of sound by manipulating the arrangement, movement, and rotation of virtual objects relative to the cube. Each cube transmits 36 parameters derived from four virtual objects, including spatial coordinates (XYZ, AED) and rotational data (pitch, roll, yaw). As a result, a total of 92 parameters from both cubes are transmitted to SuperCollider via OSC messages, which facilitates dynamic and continuous sound control based on 6DoF data.

The sound system is organized into four effect chains, each corresponding to one of the four color-coded objects within an ARCube. This configuration supports a diverse range of musical expression depending on input characteristics.

As shown in Figure 3, each color is linked to a custom SynthDef in SuperCollider, allowing performers to generate a wide range of sound effects in line with their creative intentions. Examples of parameter mappings include:

1. **The RED Object:** Pitch, gain, and spatial diffusion. Movements along the X-axis control pitch, the Y-axis controls gain, and the Z-axis adjusts spatial diffusion.
2. **The GREEN Object:** Feedback and delay effects. Delay time is controlled by movements along the Z-axis, and feedback intensity is controlled by the cube's rotation on the roll axis.
3. **The BLUE Object:** Bit-crush and freeze effects. The intensity of the bit-crush increases as the cube moves away from its center.
4. **The YELLOW Object:** Granular synthesis. Grain length is controlled by movements along the X-axis, and grain density is controlled by movements along the Z-axis.

For all objects, vertical movement along the Y-axis serves as a global control that adjusts the output gain of the effect. When the object is moved upward, the gain increases, and when moved downward, the gain decreases. This consistent mapping provides performers with intuitive dynamic control across all sound types.

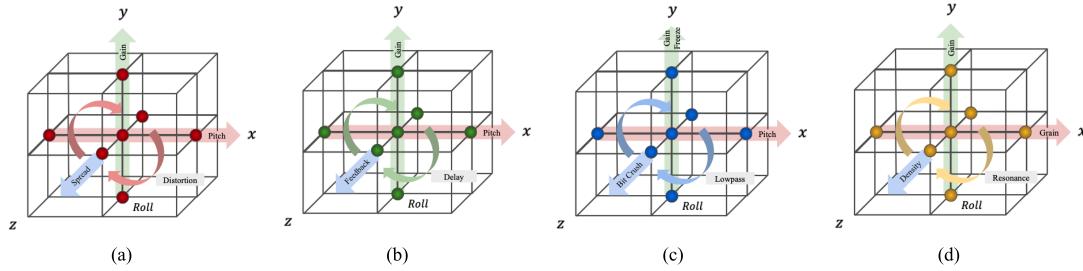


Figure 3. Sound mapping with 4 different virtual objects in AR interface: (a) Red, (b) Green, (c) Blue and (d) Yellow.

4. PERFORMANCE ANALYSIS

Various musical approaches were observed through the duo performance featuring the double bass and the modular synthesizer, both in the individual techniques of each performer and in their collaborative interaction.

To analyze the performance from the performers' perspective, two evaluation parameters were proposed (Table 1). These parameters are Hand Usage, which identifies whether one or both hands are required for interaction, and Input Persistence, which determines whether sound continues in the absence of ongoing interaction.. Using these parameters as a framework, traditional performance techniques were categorized to examine how each instrument interacts with the AR interface (Table 2).

The analysis also included aspects of multi-user interaction and visual projection. The multi-user interaction aspect observed how effectively the performers collaborated and maintained cohesion in a shared environment. The visual projection part described how real-time visual output was presented to the audience to convey the performers' gestures and their interaction with the AR interface

Table 1. Observation parameters for analyzing AR interface for creative musical interaction.

Observation Parameters	Operation
Hand Usage	One-hand vs. Two hands
Input Persistence	Manual vs. Automated

Table 2. Traditional Techniques of Double Bass and Modular Synth Examples classified with Performance Method and Hand Requirements.

	Double Bass	Modular Synth
Performance Method	Arco, Pizzicato	Sequencer, manual trigger, drone
Requirement	Two hands	One hand or Two hands
Techniques	Legato Marcato Tremolo Détaché Staccato Spiccato Con legno Position Change	Cutoff Modulation Frequency Modulation AM/Ringmod Bit Crush Wave Folding Sequencer Tempo Pattern Select

4.1 Performer-System Integration

In the context of traditional playing techniques [14], instruments that require both hands to play present limitations for performers attempting to execute techniques while simultaneously manipulating the cubes. A double bass player adjusted the virtual objects to modify and process the sound parameters before playing the resulting sounds to use traditional playing techniques (Table 2). This approach required the performer to temporarily stop playing the instrument. Nevertheless, it provided an intuitive means of exploring spatialization and location-based visual scoring, and the performer could predict the effects in order to achieve the intended expression.

Alternatively, new extended techniques were developed to address the limitations of hand usage. For example, a one-handed technique was intro-

duced that allowed the double bass player to perform actions such as tapping, pizzicato, or string pressing with one hand while manipulating the cubes with the other. This approach enabled performers to use audio effects in real time by using gestures to control virtual objects. As a creative use case, sound processing techniques were employed to generate continuous input. This approach allowed the contrabass performer to reuse previously processed sounds as new input for further transformation. For example, the blue object enabled the performer to freeze sounds and treat them as static layers, onto which new material could be added to construct complex harmonic textures. In addition, physical interaction made it possible to apply effects such as bit-crushing in real time, allowing the performer to shape the resulting sound dynamically.

Furthermore, The double bass performer discovered new ways to expand the expressive potential of the virtual objects by introducing external audio input, such as bodily sounds usable during performance. For instance, breath input was integrated into the yellow object, and voice input was added in subsequent performances.

The modular synthesizer has the ability to sustain sound without continuous physical input. This characteristic provides an environment in which the performer can interact with the interface freely, without physical constraints, and manipulate or transform sound precisely using both hands. By controlling two virtual objects simultaneously, the modular synthesizer performer was able to shape the atmosphere of the sound and maintain control over the overall sonic output. Synchronizing gesture movement with audio effects contributed to more efficient sound manipulation. For example, lifting the red object diagonally from the center to the upper rear side of the cube and then quickly lowering it to the bottom produced changes in both pitch and delay. This motion created a sense of tension, which was then immediately released through the gesture.

4.2 Multi-User Interaction

The performers on double bass and modular synthesizer were able to create an integrated soundscape by sharing the same sound design for each virtual object, despite the timbral differences between acoustic and electronic sounds (Figure 3). This demonstrates the potential of using interface-

based sound design to bridge the timbral gap between traditional acoustic and electronic instruments. On the other hand, even when using objects of the same color and applying the same processing methods, the distinct timbral qualities and performance techniques of each instrument resulted in variations in sound length and texture. As a result, in duo performances, strategies such as allocating different frequency ranges or clearly distinguishing melodic and rhythmic roles were necessary to ensure a balanced and cohesive interaction.

4.3 Visual Projection

During the performance, the audience observed the performer's movements within the AR interface through a real-time streaming display of the performer's first-person view (Figure 4). By directly seeing the performer's gestures in the air within the AR environment, the audience gained a clearer understanding of the interaction between the performer and the AR interface. Audience feedback indicated that linking the visual and auditory experiences enhanced their comprehension of the performance. Additionally, the projection served a dual purpose by providing real-time updates about mechanical issues, such as HMD operational errors during the performance, and delivering a more immersive experience to help the audience perceive how the performer's movements influenced the sound changes.

5. LIMITATION AND FUTURE WORK

Several limitations remain in the 2CUBES performance. Instruments that require both hands face challenges in simultaneous interaction with the system, often relying on pre-configured parameters or persistent effects to minimize interruptions. The steep learning curve for mastering gesture mapping, coupled with the complexity of the interactions, may hinder accessibility for some performers. Moreover, the reliance on hardware like HMDs and multi-speaker systems limits portability and installation flexibility, and their technical constraints reduce the system's adaptability. Gesture recognition with HMDs often becomes less accurate in low-light environments. Additionally, loudspeakers can fail to deliver effective spatial audio depending on the venue's acoustics and layout. The audience may also find



Figure 4. Performers with projection in the background.

it difficult to fully grasp the subtle interplay between gestures and sound generation through visual projection alone, which relies on the first-person view provided by the HMDs.

To address these challenges, future developments will focus on improving gesture adaptability and broadening input modalities. Machine learning could be integrated to dynamically adapt gestures to performance contexts, allowing for greater control and expressivity. Novel input methods, such as motion tracking and voice commands, could further expand creative possibilities. Furthermore, exploring interdisciplinary applications in fields like dance, theater, and interactive installations would highlight the system's versatility, enabling new forms of artistic expression. These advancements aim to establish 2CUBES as a flexible and accessible platform, redefining the boundaries of interactive performance art.

6. CONCLUSION

2CUBES performance demonstrates how AR can expand expressivity and immersion in live musical performance by integrating physical gestures with virtual object manipulation. It offers a new direction for extending the expressive potential of acoustic instruments into digital environments. The duo performance with double bass and modular synthesizer showcased real-time spatial audio control, performer collaboration, and audience engagement. However, challenges remain, including the complexity of gesture mapping, hardware dependency, and limitations for two-handed

instruments. Future work will focus on reducing these constraints and exploring interdisciplinary applications in dance, theater, and installation art to further broaden the system's creative scope.

7. REFERENCES

- [1] N. Petrović, “Augmented and Virtual Reality Web Applications for Music Stage Performance,” in *Proceedings of the 2020 International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST)*, 2020, pp. 33–36.
- [2] P. 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)*, 2023, pp. 10–20.
- [3] H. Shin and H. von Coler, “ARCube: Hybrid Spatial Interaction for Immersive Audio,” in *Proceedings of the 2024 ACM Symposium on Spatial User Interaction*, 2024, pp. 1–3.
- [4] M. Amazonas, V. Vasconcelos, A. Brandão, G. Kienem, T. Castro, B. Gadelha, and H. Fuks, “Collaborative Music Composition Based on Sonic Interaction Design,” in *Proceedings of the 6th International Conference on Distributed, Ambient and Pervasive Interactions (DAPI)*, 2018, pp. 335–346.
- [5] P. Lucas, A. Szorkovszky, S. Fasciani, and K. Glette, “Self-Assembly and Synchronization: Crafting Music with Multi-Agent Embodied Oscillators,” in *Proceedings of the 2024 IEEE International Conference on Autonomous Computing and Self-Organizing Systems (ACSOS)*, 2024, pp. 145–150.
- [6] J. Paterson and O. Kadel, “Audio for extended realities: A case study informed exposition,” *Convergence: The International Journal of Research into New Media Technologies*, 2023.
- [7] A. Pysiewicz and S. Weinzierl, “Instruments for Spatial Sound Control in Real Time Music Performances. A Review,” in *Musical Instruments in the 21st Century: Identities, Configurations, Practices*. Springer Singapore, 2017, pp. 273–296.
- [8] V. Bauer and T. Bouchara, “First Steps Towards Augmented Reality Interactive Electronic Music Production,” in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021, pp. 90–93.
- [9] R. Hupke, S. Preihs, and J. Peissig, “Immersive Networked Music Performance: Impact of Extended Reality on the Quality of Experience,” in *2024 IEEE 5th International Symposium on the Internet of Sounds (IS2)*, 2024, pp. 1–9.
- [10] Y. Wang, M. Xi, M. Adcock, and C. P. Martin, “Mobility, Space and Sound Activate Expressive Musical Experience in Augmented Reality,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2023, pp. 128–133.
- [11] C. P. Martin, Z. Liu, Y. Wang, W. He, and H. Gardner, “Sonic Sculpture: Activating Engagement with Head-Mounted Augmented Reality,” in *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*, 2020, pp. 39–42.
- [12] G. Santini, “A Case Study in XR Live Performance,” in *Proceedings of the International Conference on Extended Reality (XR Salento 2024)*, 2024, pp. 286–300.
- [13] H. Shin and H. von Coler, “A SuperCollider-Based Spatial Audio Performance System with AR Interfaces,” in *Proceedings of the SuperCollider Symposium 2025 (SC2025)*, 2025.
- [14] J. Bradetich, *Double Bass: The Ultimate Challenge*. Music For All To Hear, 2009.