ARCube: Combining AR Interfaces with Physical Controllers for Hybrid Spatial Interaction

Hyunkyung Shin
School of Music, Georgia Institute of Technology
Atlanta, GA, USA
hshin336@gatech.edu

Henrik von Coler
School of Music, Georgia Institute of Technology
Atlanta, GA, USA
hvc@gatech.edu

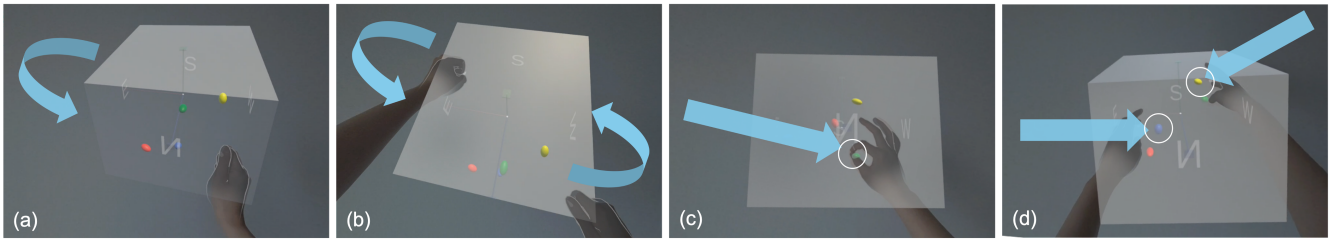

Figure 1: The ARCube interface in use: (a) grab gesture with one hand to place the cube; (b) grab gesture with two hands to rotate the cube; (c) pinch gesture with one hand to move a virtual object; and (d) pinch gesture with two hands.

ABSTRACT

The ARCube is an augmented reality (AR) interface for three-dimensional spatial control, that is designed to be used next to physical control devices. Users can freely move and place virtual objects within a cuboid that represents a simplified scale model of the surrounding space. Their relative position can be used as control data for arbitrary applications. To evaluate the user experience of the AR interface in combination with a conventional MIDI controller, a study was conducted inside an immersive audio environment. Participants explored the setup freely and performed a series of tasks related to the control of position and sound attributes of virtual sources. User feedback was collected through the think-aloud protocol with user surveys. Using an extended thematic analysis method, problems and opportunities of the AR interface and the hybrid approach could be identified. Faulty detection of gestures and issues with spatial sound perception were found to be the most critical problems. The majority of participants reported enhanced engagement and immersion.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; *Interface design prototyping*; User interface design.

KEYWORDS

Augmented Reality Interface, Spatial Audio Interaction, 3D Panning Interface, User Experience

ACM Reference Format:
Hyunkyung Shin and Henrik von Coler. 2024. ARCube: Combining AR Interfaces with Physical Controllers for Hybrid Spatial Interaction. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (SUI '24).* ACM, New York, NY, USA, 11 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Augmented Reality (AR) allows enhanced interaction with the surrounding environment and digital processes, making it easier to perform real-world tasks [7]. By adding virtual objects to the physical environment, AR supports task performance by providing information that users cannot directly perceive [11]. AR demonstrates the potential to create effective and immersive experiences, especially when combined with virtual audio content, by augmenting sensory perceptions [3, 47]. In AR, virtual objects are fixed, and users primarily approach them. Therefore, visual and physical references are crucial for navigation and object positioning in AR, and the freedom to scale visual scenes is limited when users simultaneously explore the virtual and physical worlds. These constraints and characteristics provide users with a limited immersive experience. To create a more interactive experience in AR, it is necessary to allow users to transcend these limitations [51], as providing an intuitive experience is critical for enhancing immersion [6, 26, 38].

Considerations for interaction within AR can be interpreted in the context of spatial audio, which requires appropriate input methods to provide 3D positional information. Due to the immutable characteristics of the physical world and the AR objects that depend on it, the impact of 3D sound on sensory perception in AR becomes significant.

SUI '24, October 07–08, 2024, Trier, Germany

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

In this study, we introduce ARCube, an AR interface designed to enhance user interaction and spatial audio autonomy by interpreting 3D positional information through gestures. ARCube represents physical space within an AR environment, allowing users to interact with virtual objects and control spatial sound data in real-time through gestures. The integration of a MIDI controller enhances these interactions by providing precise control over sound synthesis parameters, combining the strengths of physical and virtual interactions to overcome the limitations of traditional 2D interfaces and mid-air gesture controls [5].

The primary objective of this study is to evaluate how effectively users can control the movement of virtual objects within a 3D space and achieve desired sound spatialization in various audio scenarios. A user experience study involving seven participants employed both qualitative and quantitative methods, such as the Think Aloud Protocol (TAP) and the User Experience Questionnaire (UEQ). Specifically, TAP was utilized to analyze not only the problems but also the opportunities, thereby identifying improvements to enhance the user experience with the AR interface.

2 RELATED WORK

2.1 3D Interfaces for Spatial Control

Interfaces for 3D control and visualization are employed in various scenarios, utilizing tangible interfaces, haptic interfaces, mid-air gestures, and hybrid forms [11, 5]. Physical interfaces are effective in enhancing user engagement and interaction accuracy, allowing for more intuitive navigation of digital environments [41]. Haptic interfaces provide tactile feedback to users, enhancing the sense of touch in digital interactions. This tactile feedback significantly increases user immersion and creativity in 3D environments, making interactions more realistic and engaging [5, 23]. Mid-air gesture interfaces use sensors and cameras to enable users to control 3D environments without physical contact. This technology allows users to interact with digital content in a more natural and unrestricted manner, creating more dynamic and engaging interaction experiences [8, 5, 48]. Hybrid interfaces combine the strengths of physical control and mid-air gestures, offering versatile 3D interaction methods. By providing multiple input and feedback modes, these hybrid interfaces enhance user interaction and offer a more comprehensive and adaptable user experience [18, 5].

3D control interfaces are widely utilized in fields such as CAD, medical visualization, gaming, and educational simulations. For example, in CAD applications, tools like the *Gafinc* system allow users to manipulate complex geometric structures through gestures, significantly improving workflow efficiency compared to traditional 2D interfaces [39]. In the medical visualization field, 3D interfaces facilitate the detailed exploration of complex anatomical structures, supporting more accurate diagnosis and surgical planning [1, 32]. These interfaces enable medical professionals to interact with 3D models in ways that are not possible with 2D images, contributing to better patient outcomes. In the gaming industry, 3D interfaces provide more immersive and interactive environments, enhancing user experience [48]. The ability to navigate and manipulate 3D spaces adds depth and realism to gameplay, attracting a broader audience. Particularly in educational simulations that require spatial understanding, 3D interfaces help learners interact directly with educational content, allowing for a deeper comprehension of 3D concepts [29].

2.2 Interfaces for Spatial Sound Control

Spatial sound control has been extensively studied as a means to enhance human-machine interaction and enable efficient task performance through auditory feedback [4]. Traditionally, this research has utilized 2D interfaces for adjusting the position of sound sources [13, 12]. With advancements in technology, interfaces for spatialization have evolved, incorporating haptic feedback [31, 19] and employing actual sensors to create sound spatialization instruments [20].

Further developments in hardware, such as Head-Mounted Displays (HMDs), have propelled interaction with spatialization forward. These advancements allow for leveraging user movement in virtual reality (VR) applications to enable the spatialization and animation of individual sound sources [25]. Despite these various research efforts, there remains a gap in the development of interfaces for sound spatialization within augmented reality (AR). While some studies have explored augmented reality using smartphones [14], there is a notable scarcity of research focused on enhancing user interaction through tools like HMDs, which offer a more immersive VR experience.

In the realm of virtual reality, various approaches have been developed for controlling parameters in interactive audio environments [16]. One notable example is the VR music environment lyra^{VR}[1], which features a cube for controlling parameters in three dimensions. This highlights the potential of VR environments to enhance sound spatialization and improve user interaction. Additionally, plugins such as dearVR[2] applicable to virtual reality have emerged for 3D audio production. These plugins are utilized in fields such as mixing, mastering, and sound design.

2.3 AR Interfaces

Augmented Reality overlays virtual objects onto the physical environment to avoid collisions with the environment[27], enabling real-time interaction[10]. In this context, interaction becomes a key concept in AR[21]. To implement such interactions, various AR interfaces are utilized, and depending on the input method, different AR interfaces have been developed, including 3D user interfaces, tangible user interfaces, multimedia interfaces, natural user interfaces, and information browsers[42]. A specific example of these AR interfaces is their use in digital fabrication projects for visualization, where experimental and practice-based studies have emerged in recent years to assist unskilled workers with holographic on-site previews and instructional training in the field of architectural digital fabrication[40]. In Mobile Augmented Reality (MAR), users interact with MAR devices, such as smartphones and wearable devices, facilitating a seamless transition from the physical world to a mixed world with digital entities to enhance accessibility to digital content[10]. For human-robot interaction, AR interfaces have also been designed to convey robot motion intentions[43]. AR tools are employed to work with the absolute positions of augmented objects in the physical space[45, 36].

[1] https://lyravr.com/
[2] https://www.dear-reality.com/products/dearvr-pro-2

3 THE ARCUBE INTERFACE

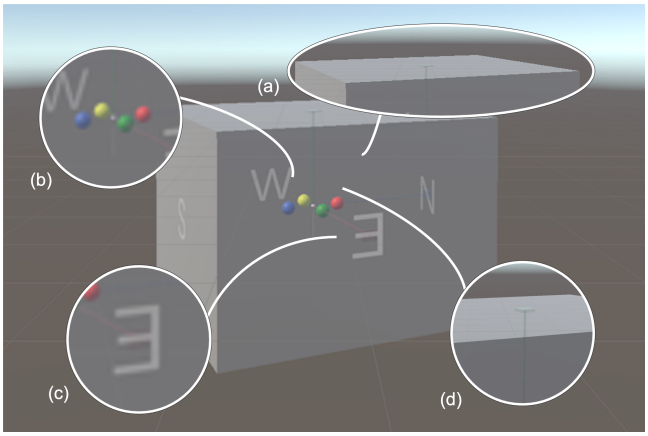

Figure 2: ARCube interface: (a) A cuboid shape proportional to the size of the room; (b) Four virtual sources positioned within the cuboid to represent spatialization; (c) The cardinal directions aligned with the spatial orientation of the physical space; (d) Axes and coordinate lines added to facilitate the initial assignment of X, Y, and Z coordinates.

The ARCube is an augmented reality interface that can be applied for three-dimensional spatial control in arbitrary use cases. ARCube represents a physical space, allowing users to intuitively understand and control spatial data. By moving virtual objects within the cube, users can achieve real-time spatial sound implementation in the physical space.

3.1 ARCube Design

Figure 2 shows the model of the augmented reality interface with highlited details. The ARCube is a cuboid that matches the physical dimensions of the spatial sound system. The proportions were calculated as follows:

$$W : L : H = 0.34 : 0.38 : 0.27$$

This formula was proportionally applied to the size of the cube, considering its use on a tabletop[35]. In particular, space for the MIDI controller used for sound synthesis and interaction through gestures was also taken into account during the experiment.

The cube's four faces are engraved with the abbreviations N, E, S, W, representing the cardinal directions (North, East, South, West), which helps align it with the physical orientation of the spatial sound system.

To allow for observation of internal movements while maintaining the external form, the cube's white surfaces were set to 60% transparency. Additionally, lines were added to facilitate the perception of the internal area, considering that surface recognition might vary depending on the user's movement. This visual aid helps participants understand the X, Y, and Z coordinates initially assigned to the cube, enhancing awareness of the physical space. Inside the cuboid, there are four virtual source-shaped ellipses, each with a diameter of 2 cm, colored red, green, blue, and yellow. These ellipses can be freely positioned both inside and outside the cube.

The augmented scene was generated using the Unity Engine (version 2022.3.22f1), a development platform widely used for creating games and simulations.

The cube allows for a variety of movements with hand gestures, but Y-axis rotation was constrained after releasing the grab gesture on the cube. It was designed to align the orientation of the cube with the spatial sound system and facilitate alignment with physical objects such as MIDI controllers. Additionally, *extosc*[3] was used to send OSC messages about the location of each virtual source.

3.2 Interaction

Interaction with the cube can be performed using either one hand or both hands. This is particularly useful for tasks requiring hand gestures such as rotation [17], and aims to facilitate simultaneous interaction with physical devices and AR interfaces by allowing free use of hands. In the same context, hand gestures with the cube are applicable to the cube itself or all internal source objects. The hand gestures are designed to be simple and non-overlapping, consisting of two types: grab and pinch 1.

Table 1: *Hand Shapes for Spatial Interaction.*

Gesture	Visualization	Uses
Grab		Move/Rotate Cube
Pinch		Move Sources

To move the entire cube, the grab gesture is applied. This gesture is chosen because it is suitable for moving the positions of fixed objects. It involves using the entire fingers [17], thus enabling connection with large objects. The pinch gesture is used for moving virtual sources inside the cube. This is because the pinch gesture is useful for selecting parts of a scene or moving specific objects [17]. To prevent overlap in the grab gesture, constraints were set for the pinch gesture using two finger combinations: either the index finger and thumb or the middle finger and thumb together.

4 INTERACTIVE SPATIAL SOUND SETUP

4.1 Overview

The ARCube is integrated in a spatial sound system to create an interactive immersive soundscape. Figure 3 shows the signal flow for the complete system, which is designed specifically for the concluding user study presented in Section 5. The spherical coordinates of the four augmented spheres relative to the cuboid's center are converted to Open Sound Control (OSC) [46] messages inside the Unity software, running on the Meta Quest 3. These messages are send to the spatialization software, running on a dedicated Linux

[3] extosc: https://github.com/Iam1337/extOSC.

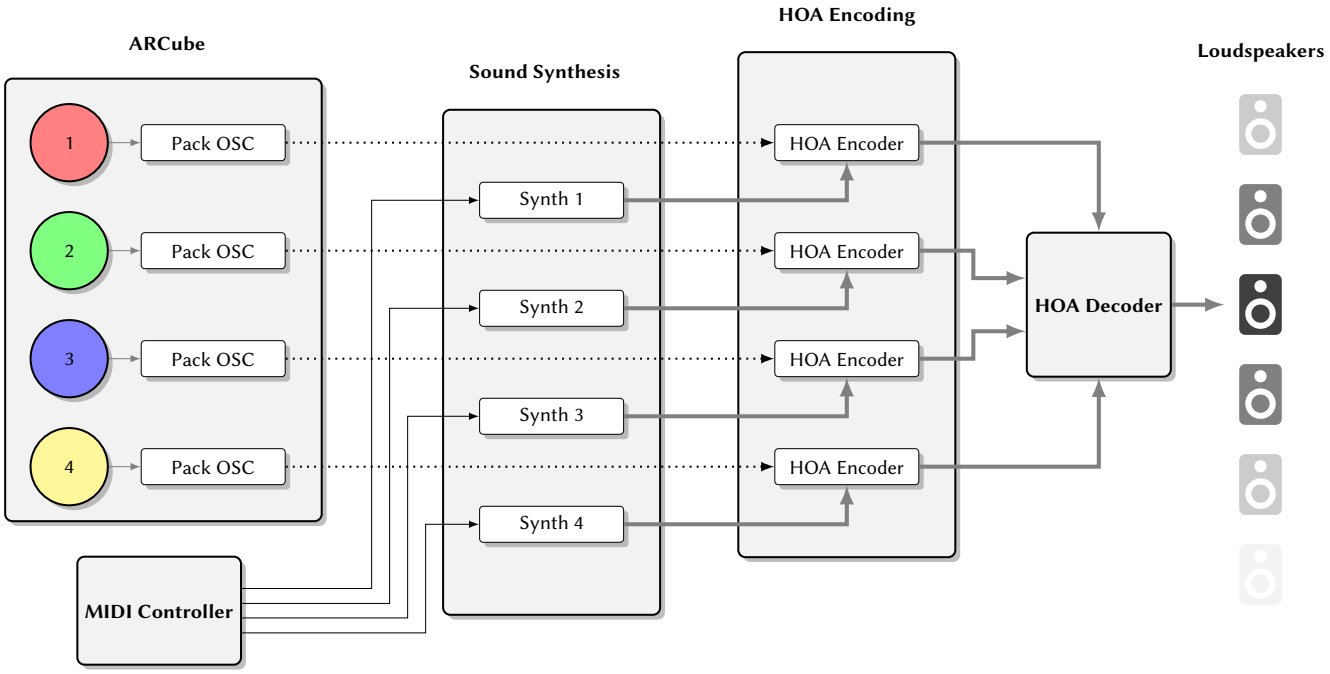

Figure 3: Signal flow for the user study (dotted=OSC, black=MIDI, thick gray=audio).

server (see Section 4.3), with a rate of 100 Hz via WiFi. An *Evolution UC-33* MIDI controller is connected to four sound synthesis processes running on the same server, with the resulting audio streams being sent to Higher Order Ambisonics (HOA) encoders. Combined with the OSC position data, four virtual sound sources are created, which are subsequently decoded and sent to the loudspeaker setup.

4.2 Sound Synthesis & Control

All four virtual sound sources are fed with an instance of the same general synthesis process as auditory stimuli, generated in Pure Data [34]. This sound synthesis process is designed to deliver localizable sounds with a wide enough tuning range to create four distinguishable sounds from the same basic process. In addition, these sound sources resemble abstract soundscape elements, rather than musical events or streams. This allows for a neutral audio-visual experience without a complex layer of music perception and interaction.

The basic synthesis process creates a repetitive sequence of randomly triggered events. Since noise bursts are known to be localizable and are thus frequently used in spatial auditory perception studies [37, 50], they are chosen as the basic element. Four adjustable parameters, as listed in Table 2, are used to tune the sound source via the four leftmost controller columns of the MIDI device, shown in Figure 4.

One individual column is used for each synthesis process, with the fader mapped to the signal's gain, ranging from 0 to 100 %. The dial above the fader controls the mean density of the random events, occurring at frequencies from 1 Hz to 100 Hz, with a variance of 10 %. Each triggered event is a created from a noise burst, processed with a resonating low-pass filter (using Pure Data's `bob~` object) and an exponentially decaying amplitude envelope. Decay time and filter quality are controlled with the mid row dial to change the character of the events from long noise-burst to short tonal impulses. Each triggered event has a random cutoff frequency, respectively pitch. The upper row of dials can set the random corridor from 300-600 Hz to 3000-6000 Hz.

Figure 4: Relevant elements of the MIDI controller: The four leftmost columns control the parameters of the four synthesis processes.

With the above introduced parameters, each sound source can be tuned from a sparse sequence of noise bursts to a dense texture of short tonal events with different frequency ranges.

Table 2: Sound Synthesis Parameters.

Parameter	*Min Value*	*Max Value*
Gain	0 %	100 %
Density	1 Hz	100 Hz
Frequency / Pitch	300-600 Hz	3000-6000 Hz
Character	Long noisy	Short tonal

4.3 Spatialization System

The four sound sources, generated with the synthesis process described in Section 4.2, are spatialized on a three-dimensional immersive sound system. The actual room dimensions are 4.7 m (W) × 5.2 m (L) × 3.7 m (H), which were measured to be the same area as the available range of the 28 speakers.

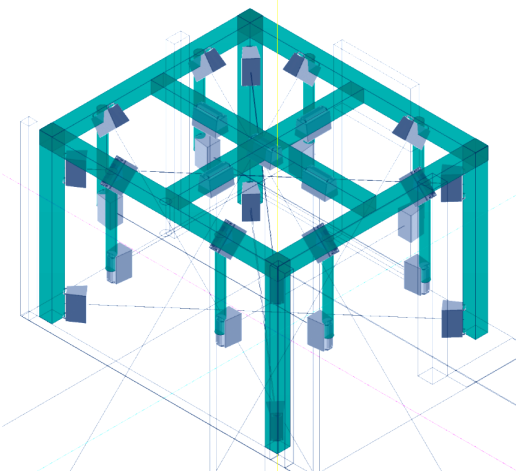

Figure 5: Immersive audio setup with 28 loudspeakers.

The rendering software runs on a Linux computer with a Klark Teknik DN9630 audio interface. A Jack[4] server is started with a sample rate of 48 000 Hz and a buffer size of 64 samples, resulting in an audio output latency of 1.33 ms.

A custom Ambisonics encoding system[5] is implemented, based on SC-HOA [22], a Higher-Order Ambisonics extension for SuperCollider [30]. This system utilizes fifth-order Ambisonics to generate four virtual sound sources, which are controlled through spherical coordinates received via OSC messages. The OSC processing functions implemented in SuperCollider constantly receive spatial positioning data from the head-mounted display.

The extracted data is then passed to the HOA encoder, which converts the input audio signals into HOA signals. A simple artificial reverb is added to increase the distance perception. The resulting Ambisonics signal is decoded with the standalone version of the AllRADecoder from the IEM Plugin Suite[6] for 28 Neumann KH120 loudspeakers which are distributed evenly on a truss system, as visualized in Figure 5.

[4]https://jackaudio.org
[5]https://github.gatech.edu/l42i/litespat
[6]https://git.iem.at/audioplugins/IEMPluginSuite

5 USER EXPERIENCE STUDY

5.1 Method and Scope

The presented study focuses on the user experience of the AR interface in combination with the physical controller inside the immersive audio setup described in Section 4.3. With our methods and setup we aim at answering the following two research questions:

R1 Can the augmented reality interface be used alongside a physical controller in a hybrid setup?
R2 Is the ARCube a suitable tool for spatial control tasks?

To answer the two research questions, users interacted with the interface in the immersive audio setup through free exploration and by completing simple tasks. The setup and tasks challenged the users to use both the physical and the augmented interfaces simultaneously or in alternation.

User feedback collected with the think-aloud method [24, 2] and additional surveys, to get detailed insight into problems and opportunities of the proposed hybrid interaction system. During the active part of the experiment, participants were encouraged to verbalize their actions and thoughts for the the think-aloud method. The supervisor reminded participants to continue speaking if they fall silent for more than 10 seconds.

5.2 Participants

A total of 7 participants (Aged 23 to 33 years, Mean age = 25 years) were recruited for the study. According to the results of the Goldsmiths Musical Sophistication Index (Gold-MSI) [33] conducted before the experiment, the participants had a comprehensive understanding of music and proficiency in playing musical instruments, but limited experience with AR and spatial audio technologies. The mean scores for various dimensions of musical sophistication were as follows: Active Engagement (M = 5.3, SD = 1.2), Perceptual Abilities (M = 6.0, SD = 0.8), Musical Training (M = 4.5, SD = 1.3), Singing Abilities (M = 5.1, SD = 1.0), and Emotional Attachment to Music (M = 6.3, SD = 0.9). These scores culminated in an overall General Musical Sophistication score of 5.5 (SD = 1.0).

5.3 Experiment Setup

Figure 6 shows a participant while performing a task in the experiment setup. The participant's movement radius is constrained by the green square, measuring 1.8 m × 1.8 m, in the center of the spatial audio system. The MIDI controller is placed on a table, located at the front boundary of the square at a height of 1.2 m. Instructions are displayed on a Sony XBR-75X81CH monitor at a distance of 2.2 m and a height of 85 cm from the floor. In this study, we employed the Meta Quest 3 to establish an augmented reality (AR) environment, confined to the actual dimensions of a physical spatial sound system.

While interacting with the setup, a first-person view video was recorded from each participant's HMD. Additionally, a hanging microphone was used to capture the participants' think-aloud speech and sounds. The output values for the four sound sources from the MIDI controller and the audio played through 28 loudspeakers were recorded together in a 5th-order Ambisonics channels format.

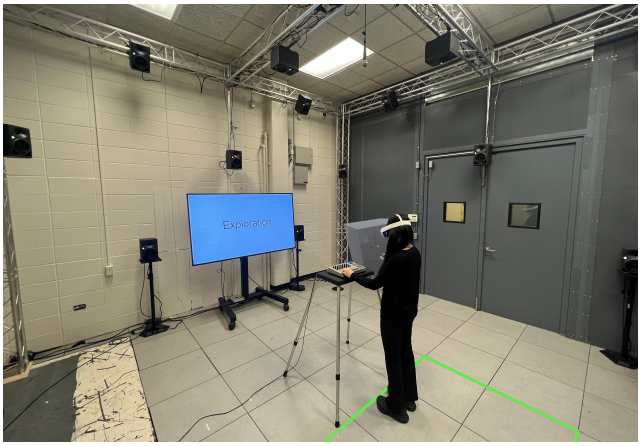

Figure 6: Participant interacting with the hybrid setup in the immersive audio system during the experiment.

5.4 Procedure

Each experimental session was conducted individually with a single participant and was divided into four parts:

(1) Scene Setup
(2) Exploration
(3) Static Tasks
(4) Dynamic Tasks

Prior to performing the first task, participants received detailed instructions about the experiment and completed a pre-survey, which included the Gold-MSI. After the experiment concluded, participants completed the User Experience Questionnaire(UEQ) and sonic interaction questionnaires (SID) [16] surveys.

5.4.1 Scene Setup. Participants are equipped with the head-mounted display, standing at the outer boundary of the spatial audio system. The augmented reality scene is started, showing the ARCube placed 2 m to their front, next to the instruction screen. They are then prompted to grasp the cube using a grab motion and move it towards a table. The participant is instructed to place the cube at an arbitrary position next to the MIDI controller using a release motion.

5.4.2 Exploration. The exploration phase provides participants with an opportunity to acclimate to the experimental environment and freely navigate the augmented and physical interfaces. Participants are instructed to manipulate all relevant knobs and sliders on the MIDI controller to activate sound synthesis, while simultaneously using the ARCube to explore the interaction between sound synthesis and the MIDI controller. This process does not impose restrictions on gestures, thereby enabling a diverse range of auditory experiences.

5.4.3 Static Tasks. Static tasks aim to adjust the sound position and synthesis parameters through interaction with the ARCube, thereby creating various audio scenes. Participants will sequentially follow the on-screen instructions to complete the tasks as following:

ST1 Activate all sound sources and set the synthesis parameters to minimum.
ST2 Each source should be moved to an individual letter on the walls of the cube.
ST3 The synthesis processes should be adjusted to result in four distinguishable sounds.
ST4 All sources should be moved to the center of the cube.
ST5 The synthesis processes should be changed to result in a different audio scene.
ST6 Each source should be moved to an individual upper corner of the cube.
ST5 The synthesis processes should be changed to result in a different audio scene.
ST7 Select arbitrary positions outside the cube for all sources.
ST8 Deactivate all sound sources.

5.4.4 Dynamic Tasks. In the Dynamic Tasks, participants interact with the four virtual sources using one or both hands. During the tasks, participants explore usability through repetitive circular motions, while the number of simultaneously active virtual sources is limited to facilitate the observation of gestures. The specific instructions are as follows:

DT1 Set all synthesis parameters to minimum.
DT2 Make only the first (red) source audible and perform a repeating circular movement inside the cube. Change the synthesis parameters during this activity to your liking. Repeat this section if desired.
DT3 Make only the second (green) source audible and perform a repeating circular movement around the cube. Change the synthesis parameters during this activity to your liking. Repeat this section if desired (one hand).
DT4 Activate only sources three (blue) and four (yellow) and perform repeating circular movements inside or outside the box with both sources at the same time (two hands). Adjust the synthesis parameters for this activity to your liking. Repeat this section if desired.
DT5 Perform a free improvisation using all parameters of the synthesis and spatial control as desired.
DT6 Deactivate all sound sources.

6 RESULTS

6.1 User Experience Questionnaire

Figure 7 illustrates the average UEQ scores compared to the benchmark scores. The results show that the system received an average score of approximately 1.69 (Good) for attractiveness. For perspicuity, the system had a mean score of about 2.00 (Good). The mean score for efficiency was close to 0.96 (Below Average). Dependability scored 1.64 (Good). Stimulation scored 1.89 (Excellent), emphasizing that participants found the system engaging and exciting to use. Lastly, the system's novelty was rated at 1.32 (Good). The system received good ratings in most categories, with efficiency being the only category rated lower at "Below Average."

6.2 Usability Problems from the TAP

To identify usability problems based on the recorded experiment process and transcribed verbalized responses for each individual

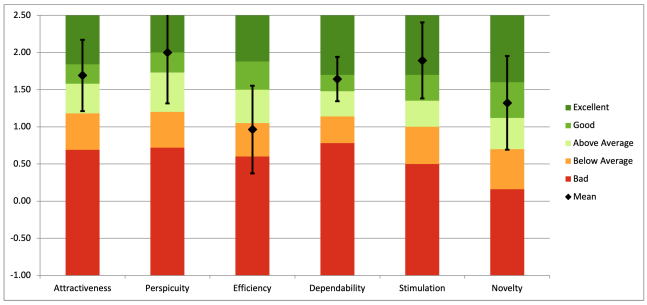

Figure 7: The average UEQ scores and the benchmark results.

participant, we employed the data analysis method for individual and final problems as proposed by Alhadreti and Mayhew [2]. This method was used because it allows for a detailed analysis of user experiences in TAP [49]. Participants reported a total of 40 usability issues through the TAP method, which were categorized into 13 unique problems described in Section **??**.

6.2.1 Unique Usability Problems.

UP1 *Errors in detecting hand gestures with the Head-Mounted Display (HMD)*

UP2 *Errors in the HMD's viewpoint detection lead to discrepancies in the cuboid's position.*

UP3 *Participants experience confusion in selecting gestures.*

UP4 *Confusion between Grab/Release and pinch gestures.*

UP5 *The rotation constraint along the Y-axis of the cube limits creative actions.*

UP6 *Unclear distinction between front (S) and back (N) sounds.*

UP7 *Unclear perception of positions above and below.*

UP8 *Weakened spatial recognition regarding abstract position selection.*

UP9 *Weakened spatial recognition due to the similarity of sounds.*

UP10 *Confusion caused by abrupt speed changes in gesture movements.*

UP11 *Misalignment between the cube and the physical environment in relation to cardinal directions.*

UP12 *Confusion in simultaneous spatial awareness of four virtual sources.*

UP13 *Weakened spatialization outside the cube.*

6.2.2 Categorization of the Usability Problems. Usability problems were categorized into four types (PCs) based on a thematic analysis of the identified issues [2, 49].

PC1 *Misoperation Due to Inexperience with HMD Device:* Users showed inexperience in using the HMD due to a lack of experience with AR scenes, which could lead to difficulties in task performance.

PC2 *Confusion in Object Manipulation Through Gestures:* Gestures can be confused with certain movements or lack of the visual detections from HMD.

PC3 *Confusion in Sound Localization and Differentiation Due to Spatialization:* Users listened to spatialized sounds based on the position of objects, but the degree of perception varies depending on the location.

PC4 *Limitation on Creative Tasks Due to Restricted Movement:* Users felt limited in their creative or freedom of the movement while interacting with AR interface.

6.2.3 Distribution of the Usability Problems: The experimental procedure was divided into two stages: Initial Stage (S1) and Task Stage (S2). S1 included the Scene Setup and Exploration phases, while S2 encompassed Static Tasks and Dynamic Tasks. The severity of individual usability problems was categorized into four levels based on their impact on participants' performance [2, 49]. Additionally, the number of participants experiencing the problem was used as a criterion for determining the severity.

C Critical, the problem is explored by all (7/7 users) or the problem prevented the completion of a task;

Ma Major, Major, the problem is explored by more than 75 % (5/7 users) or the problem caused significant delay more than 30 seconds or frustration;

Mi Minor, the problem is explored by more than 50 % (3/7 users) or the problem had minor effect of delay on performance or slight frustration;

E Enhancement, the problem is reported over 10 % (1/7 users) or participants made suggestions or indicated a preference, but the issue did not cause impact on performance.

Table 3: *Distribution of Usability Problems.*

	S1					*S2*				
	SUM	*C*	*Ma*	*Mi*	*E*	*SUM*	*C*	*Ma*	*Mi*	*E*
PC1	4	1	1	2	1	0	0	0	0	0
PC2	5	2	1	1	1	7	2	1	2	2
PC3	9	1	2	2	4	11	2	1	1	7
PC4	2	1	0	1	0	2	0	0	1	1

The distribution according to this classification is shown in the Table 3, and below, the unique problems measured more than once from individual items are reported.

The most frequently reported usability problem was Type PC3 (Confusion in Sound Localization and Differentiation Due to Spatialization). A total of 20 issues related to confusion in spatialization were reported across S1 and S2. PC1 (Misoperation Due to Inexperience with HMD Device) only occurred in S1 and was not mentioned in S2.

6.3 Usability Opportunities from the TAP

We inverted the traditional approach of identifying problems [2] to also detect usability opportunities. Our focus is on the benefits and possibilities of augmented and hybrid interfaces, and we propose a new approach for the evaluation of qualitative data. Participants reported a total of 57 usability opportunities through the TAP method,

which were categorized into 14 unique opportunities detailed in Section 6.3.1.

6.3.1 Unique Usability Opportunities.

UO1 *Free placement of the Cube Next to the Table.*
UO2 *Hybrid Use of the Cube and Physical Devices.*
UO3 *Gesture-Based Interaction with the Cube.*
UO4 *Enhancing Auditory Functions through Positional Changes.*
UO5 *Improving Spatial Auditory Function through Movement.*
UO6 *Natural Circular Movement around Sources.*
UO7 *Movement for One Hand and One Source within the Cube.*
UO8 *Movement for Two Hands and Two Sources within the Cube.*
UO9 *Movement for One Hand and One Source outside the Cube.*
UO10 *Movement for Two Hands and Two Sources outside the Cube.*
UO11 *Adjusting Movement Speed.*
UO12 *Narrative Attribution Based on Sound Characteristics.*
UO13 *Adjusting Source Positions Based on Sound Characteristics and Narrative.*
UP14 *Movements with the Body of the Cube.*

6.3.2 Categories of Usability Opportunities. We analyzed usability opportunities from the transcription of the think-aloud protocol (TAP) and video recordings within four different categories (OCs) using the text coding method [28, 2]. We also considered the crucial design principles for the AR interface [15].

OC1 *Spatial Awareness of the Physical Environment:* AR user interfaces can be naturally and intuitively placed within the physical environment. Users consider the position and orientation of digital content in relation to the physical environment based on spatial awareness [9].

OC2 *User Engagement with the Spatial Interaction:* AR user interfaces enhance user engagement and immersion through spatial interaction. Users can manipulate digital content naturally and engagingly through gestures, and experience spatialized sound based on the position of virtual objects.

OC3 *Realism and Simplicity in Use:* AR user interfaces enable users to easily interact with AR content through simple and intuitive virtual object designs and gestures.

OC4 *Contextual Awareness and Adaptability:* AR user interfaces enable users to perform creative spatialization tasks. Users dynamically adjust virtual objects based on their location, surroundings, and previous interaction experiences, enhancing personalized experiences with spatialization.

6.3.3 Distribution of the Usability Opportunities. To measure the extent of usability, we proposed a method to explore the level of opportunities based on their impact on usability. We observed how many users commonly identified these opportunities and how much interest users showed in their behavior. Interest was measured through the duration of the behavior and whether users expressed positive reactions. Similar to how the severity of usability problems was classified into four levels [2], the degree of opportunities identified was also categorized into four levels.

C Critical, the opportunity is explored by all (7/7 users) and enhances task completion;
Ma Major, the opportunity is explored by more than 75 % (5/7 users) or engages users for more than 20 seconds and elicits excitement;
Mi Minor, the opportunity is explored by more than 50 % (3/7 users) or engages users for more than 5 seconds and slight excitement;
E Enhancement, participants made suggestions or indicated a preference of the opportunity or it is reported over 10 % (1/7 users).

Table 4: *Distribution of Usability Opportunities.*

	S1					*S2*				
	SUM	*C*	*Ma*	*Mi*	*E*	*SUM*	*C*	*Ma*	*Mi*	*E*
OC1	3	2	0	1	0	3	2	0	1	0
OC2	12	5	4	3	0	13	6	4	2	1
OC3	4	3	0	1	0	8	4	1	2	1
OC4	7	2	0	3	2	7	2	2	1	2

The most frequently reported usability opportunity was Type OC2 (User Engagement with the Spatial Interaction). A total of 25 issues related to confusion in spatialization were reported from S1 and S2 (see Table 4).

6.4 Open Feedback

Participants were asked to provide one open-ended feedback about the experiment after its completion. The question was as follows:

Q1 What are your comments on the use of the the AR interface in this setup?

Participants felt more connected to the physical world while moving to grab virtual objects (5/7 users). One participant stated, *"[...] it made me feel like I am actually in this physical environment and I especially liked having to move around to grab virtual objects. This made me feel more connected to the physical world."* The use of the MIDI controller further enhanced this experience, comparing it to the difference between reading a physical book and an e-book. One participant mentioned, *"[...] the MIDI controller is simplistic but offers a lot of creative potential due to the synthesis parameters. Turning the knobs felt like controlling something real, which was very satisfying."* Another participant noted that manipulating sound sources in 3D with hand gestures and camera pass-through maintained the feeling of being present in the room.

Some participants mentioned that the system did not always accurately detect gestures (3/7 users). One participant commented, *"[...] the gesture detection was sometimes inaccurate, and it overly detected unintended gestures."* The comfort and brightness of the headset were also concerns, with one participant stating, *"[...] the headset was slightly uncomfortable and too bright, which was worrying."*

Suggestions for improving the user interface (5/7 users) included allowing more simultaneous actions, enhancing gesture detection reliability, and integrating virtual knobs and faders for a fully virtual

interface experience. One participant pointed out, *"[...] the limitation is that only two things can be changed at a time. I don't understand the purpose of rotating the cube."* Another participant suggested, *"[...] it would be better if there were more visual feedback for virtual sources, such as numbers and movements."*

ARCube was described as user-friendly, with minimal errors and fast response times (7/7 users). Participants found the experience enjoyable and engaging. One participant described, *"[...] the AR interface was very user-friendly. The process was easy, and there were almost no errors."* The duration of the experiment was generally deemed appropriate (4/7 users), and some participants expressed interest in additional features such as polyphony and the ability to customize sounds (2/7 users). One participant said, *"[...] I liked the range of sounds available, but I would like to explore a wider variety of sounds that I can patch myself."*

Overall, participants generally expressed positive feedback about the AR interface experiment, emphasizing the immersive experience and user-friendly nature of the system.

7 DISCUSSION

7.1 RQ1: Hybrid use of the AR Interface with Hardware Controller

Initially, in the classification of problem categories, PC1 (Misoperation Due to Inexperience with HMD Device) is most closely associated with Hybrid Use. Regarding PC1, participants reported four usability issues only in S1 (Severity: C:1, Ma: 1, Mi: 2, E: 1). This indicates that as participants became more familiar with using the HMD, they were able to use the hybrid setup without issues. The problems reported in S1 can also be explained by the technical limitations of the HMD hardware (UP1, UP2). This inexperience with HMD use reflects the limited experience participants had with AR environments and may factor into the "below average" efficiency rating in the UEQ. Since all participants encountered HMD-related problems in the early stages of the experiment (PC1, S1, C:1), the efficiency of the experimental process (M = 0.96, Below average) was likely rated lower compared to attractiveness (M = 1.69, Good) or stimulation (M = 1.89, Excellent).

Additionally, participants experienced confusion when the default orientation of the cube did not match during hybrid use (UP11). This underscores the importance of aligning the orientation guidelines of the AR interface with the physical environment for proper spatial awareness.

In terms of usability opportunities related to OC1 (Spatial Awareness of the Physical Environment), participants identified the same number and degree of opportunities in both S1 and S2 (Opportunities: C:2, Ma: 0, Mi: 1, E: 0). This indicates that participants were able to sufficiently discover the usability of the AR interface for hybrid use and successfully perform hybrid tasks regardless of the task level. Participants could freely place the cube next to physical objects according to their personal convenience of interacting with the AR interface via gestures (UO1), and they could use various movements with one or both hands to interact with virtual sources (UO3). Specifically, participants were able to use both the MIDI controller and the ARCube simultaneously to perform both sound alteration and spatialization (UO2).

Furthermore, the interaction between the AR interface and the MIDI controller enhanced the sense of realism by bridging the gap between the virtual and physical worlds (Open Feedback, 5/7 users). One participant remarked, "[...] turning the knobs felt very satisfying, as if I were controlling something real." Another participant directly mentioned, "[...] moving to grab the virtual objects made me feel more connected to the physical world."

User feedback highlighted areas for improvement in gesture detection accuracy and the comfort of the AR interface. For instance, some participants noted that the system did not always accurately detect gestures (Open Feedback, 3/7 users). One participant stated, "[...] gesture detection was sometimes inaccurate, and the system often over-recognized unwanted gestures." This suggests that while spatial control and hybrid interaction were beneficial, the reliability of gesture detection remains a crucial area for improvement.

7.2 RQ2: Suitability for Spatial Control

In the study, various usability issues related to spatial interaction were identified and classified into specific problem, indicating that interactions involving gestures could lead to problems (PC2) and that users might experience confusion in the perception of spatialized sound (PC3). Notably, PC3 (Confusion in Sound Localization and Differentiation Due to Spatialization) was the most frequently reported usability issue, with a total of 20 instances (S1: 9, S2: 11). Additionally, participants reported feeling limited in creativity during spatial interactions (PC4).

Specifically, participants experienced confusion while using grab and pinch gestures (UP3, UP4), which was also related to sudden changes in movement speed (UP10). The inaccuracy of these gestures could be associated with recognition errors by the HMD (UP1). Participants reported constraints in creative interaction due to limitations in the Y-axis rotation of the cube (UP5). However, this constraint also had a positive aspect, as it alleviated spatial confusion when the orientation of the cube did not match the physical environment (UP11).

In recognizing virtual sources, participants particularly reported confusion when perceiving each sound from a stationary position after placing the virtual source (UP6, UP7). The degree of confusion increased when the sounds had similar frequencies (UP9), when the location was abstract or distant from the cube (UP8, UP13), and when the number of sounds to be recognized increased (UP12).

The cube also presented opportunities for various spatial interactions, which correlated with OC2 being reported as having the most potential. Users were able to interact with the AR interface using only two gestures, grab/pinch, and this simple design (OC3) elicited positive reactions from participants (Open Feedback, 7/7 users). The grab gesture and virtual objects could interact within a 1cm distance after initial detection by the HMD (UO3).

Participants performed a variety of movements, such as circular, figure-eight, vertical/horizontal, and crossing patterns, both inside and outside the cube when interacting with virtual sources (UO7, UO8, UO9, UO10). Circular movements (UO6) and speed adjustments (UO11) were particularly natural methods for participants to explore space without specific instructions. When experiencing spatial confusion with static virtual sources, participants improved spatial perception by moving the virtual sources (UO4, UO5), helping

to resolve the front-back confusion observed in UP6, as identified in Wightman's study [44].

Through experiencing spatial interactions, participants engaged in narrative activities, such as naming or assigning roles based on sound characteristics (UO12). These activities sometimes contributed to creating an overarching story for the entire audio scene. To implement spatial characteristics resulting from sound position changes, participants differentiated between background and primary sounds and arranged them spatially (UO13).

Additionally, for more creative and abstract interactions, participants attempted to arbitrarily adjust the default orientation of the cube or impart movement to the cube itself (UP14). When the grab gesture was maintained on the cube, interactions with all axes were possible regardless of the Y-axis constraints, potentially resolving the issue described in UP5.

Overall, participants identified a similar number of positive opportunities in both S1 and S2 (S1: 26, S2: 31). As described in Figure 7, participants rated the system as above average on all aspects except for efficiency in the UEQ. These results suggest that participants generally had a positive experience interacting with the AR interface, including spatial interaction elements.

7.3 Limitations

The unresolved limitation was detected with the HMD when implementing an AR environment. The Meta Quest 3 used in the experiment showed variations in gesture recognition levels depending on hand rotation within the AR environment, leading to errors in interactions with virtual objects. Additionally, the grab and pinch gestures are the only two types of gestures designed for interaction, which can add confusion during detection.

Spatial audio can cause confusion in sound localization when the sound is delivered from a static position, and it becomes difficult to accurately identify when far from the center of the cube or when selecting abstract positions.

The experiment had limitations in sample size and diversity of participant backgrounds, as each experiment was conducted with a single participant. Additionally, the MIDI controller was the only physical device used in the experiment.

The Y-axis rotation constraint was implemented to maintain physical consistency, but it was found to limit users' creative interactions.

8 CONCLUSION

In this study, the usability of the ARCube interface was investigated through immersive augmented reality with a spatial audio system. The research extended traditional usability problem methods to explore usability opportunities through user surveys. While participants provided positive feedback in most categories, the interface was rated below average in terms of efficiency in the UEQ. Through the think aloud protocol (TAP), 40 usability issues were identified and categorized into 13 unique problems, with frequent reports of confusion regarding sound localization. Conversely, 57 usability opportunities were identified and categorized into 14 unique opportunities, indicating that the AR interface enhances connectivity with the physical world and provides a sense of immersion.

Despite the positive usability of the ARCube interface, several limitations were identified, including gesture recognition accuracy and constraints on creative interactions due to Y-axis rotation limitations. The study only presented two gestures to facilitate interaction, but expanding and comparing the types of possible gestures could help identify more efficient ways of interaction. Additionally, using the interface simultaneously by two or more individuals or each using the interface separately within the same AR environment could enable more creative interactions.

There is also potential for the ARCube interface to be used in conjunction with other hardware devices. This could facilitate collaboration among users through spatial control across various hardware devices. Future versions of ARCube could explore integrating additional rotational degrees of freedom without compromising physical orientation consistency. Improvements might include adding visual interaction methods with the cube or objects to enable more flexible interactions. This could involve incorporating buttons that allow participants to set the cube's constraint limits themselves or adding gesture-based interaction capabilities.

Future research could explore the integration of customized acoustic synthesis parameters and more complex spatial audio scenarios to overcome cognitive limitations associated with fixed auditory cues. Additionally, conducting user studies with larger and more diverse participant pools would provide further insights into the usability and effectiveness of the ARCube interface across various user demographics and contexts. These enhancements are expected to further improve the usability and effectiveness of the ARCube interface.

REFERENCES

[1] Mauren Abreu de Souza, Daoana Carolaine Alka Cordeiro, Jonathan de Oliveira, Mateus Ferro Antunes de Oliveira, and Beatriz Leandro Bonafini. 2023. 3d multimodality medical imaging: combining anatomical and infrared thermal images for 3d reconstruction. *Sensors*, 23, 3, 1610.

[2] Obead Alhadreti and Pam Mayhew. 2018. Rethinking thinking aloud: A comparison of three think-aloud protocols. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, 1–12.

[3] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair MacIntyre. 2001. Recent advances in augmented reality. *IEEE computer graphics and applications*, 21, 6, 34–47.

[4] Durand Begault, Elizabeth M Wenzel, Martine Godfroy, Joel D Miller, and Mark R Anderson. 2010. Applying spatial audio to human interfaces: 25 years of nasa experience. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society.

[5] Lonni Besançon, Anders Ynnerman, Daniel F Keefe, Lingyun Yu, and Tobias Isenberg. 2021. The state of the art of spatial interfaces for 3d visualization. In *Computer Graphics Forum* number 1. Vol. 40. Wiley Online Library, 293–326.

[6] Eleonora Brivio, Silvia Serino, Erica Negro Cousa, Andrea Zini, Giuseppe Riva, and Gianluca De Leo. 2021. Virtual reality and 360 panorama technology: a media comparison to study changes in sense of presence, anxiety, and positive emotions. *Virtual Reality*, 25, 303–311.

[7] Frederick P Brooks Jr. 1996. The computer scientist as toolsmith ii. *Communications of the ACM*, 39, 3, 61–68.

[8] Dom Brown, Chris Nash, and Tom Mitchell. 2018. Understanding User-defined Mapping Design in Mid-air Musical Performance. In *Proceedings of the 5th International Conference on Movement and Computing*. Genoa, Italy, 27.

[9] Douglas S Brungart, Julie Cohen, Mary Cord, Danielle Zion, and Sridhar Kalluri. 2014. Assessment of auditory spatial awareness in complex listening environments. *The Journal of the Acoustical Society of America*, 136, 4, 1808–1820.

[10] Jacky Cao, Kit-Yung Lam, Lik-Hang Lee, Xiaoli Liu, Pan Hui, and Xiang Su. 2023. Mobile augmented reality: user interfaces, frameworks, and intelligence. *ACM Computing Surveys*, 55, 9, 1–36.

[11] Julie Carmigniani, Borko Furht, Marco Anisetti, Paolo Ceravolo, Ernesto Damiani, and Misa Ivkovic. 2011. Augmented reality technologies, systems and applications. *Multimedia tools and applications*, 51, 341–377.

[12] Thibaut Carpentier. 2018. A new implementation of spat in max. In *15th Sound and Music Computing Conference (SMC2018)*, 184–191.

[13] Thibaut Carpentier. 2015. Tosca: an osc communication plugin for object-oriented spatialization authoring. In *41st International Computer Music Conference (ICMC)*, 368–371.

[14] Adrià M Cassorla, Gavin Kearney, Andy Hunt, Hashim Riaz, Mirek Stiles, and Damian T Murphy. 2020. Augmented reality for daw-based spatial audio creation using smartphones. In *Audio Engineering Society Convention 148*. Audio Engineering Society.

[15] Sophia Chen. 2024. Augmented reality user interfaces: analyzing design principles and evaluation methods for augmented reality (ar) user interfaces to enhance user interaction and experience. *Human-Computer Interaction Perspectives*, 4, 1, 15–27.

[16] Michele Geronazzo and Stefania Serafin, (Eds.) 2023. *Spatial Design Considerations for Interactive Audio in Virtual Reality. Sonic Interactions in Virtual Environments*. Springer International Publishing, Cham, 181–217. ISBN: 978-3-031-04021-4. DOI: 10.1007/978-3-031-04021-4_6.

[17] Julien Epps, Serge Lichman, and Mike Wu. 2006. A study of hand shape use in tabletop gesture interaction. In *CHI'06 extended abstracts on human factors in computing systems*, 748–753.

[18] Anna-Katharina Frison, Philipp Wintersberger, Andreas Riener, and Clemens Schartmüller. 2017. Driving hotzenplotz: a hybrid interface for vehicle control aiming to maximize pleasure in highway driving. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*, 236–244.

[19] Steven Gelineck and Dan Overholt. 2015. Haptic and visual feedback in 3d audio mixing interfaces. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound*, 1–6.

[20] Florian Goeschke. 2022. The ioscahedron: developing a hybrid spatialization instrument. In *Proceedings of the 17th International Audio Mostly Conference*, 151–154.

[21] Eg Su Goh, Mohd Shahrizal Sunar, and Ajune Wanis Ismail. 2019. 3d object manipulation techniques in handheld mobile augmented reality interface: a review. *IEEE Access*, 7, 40581–40601.

[22] Florian Grond and Pierre Lecomte. 2017. Higher Order Ambisonics for SuperCollider. In *Proceedings of the Linux Audio Conference 2017*. Saint-Etienne, France.

[23] Philipp Pascal Hoffmann, Hesham Elsayed, Max Mühlhäuser, Rina R Wehbe, and Mayra Donaji Barrera Machuca. 2023. Thermalpen: adding thermal haptic feedback to 3d sketching. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–4.

[24] Monique WM Jaspers, Thiemo Steen, Cor Van Den Bos, and Maud Geenen. 2004. The think aloud method: a guide to user interface design. *International journal of medical informatics*, 73, 11-12, 781–795.

[25] Steven Jiang, Lawrence Lim, and Misha Sra. 2023. Spatializing music in virtual reality. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*, 1–3.

[26] Jung In Kim, Sining Li, Xingbin Chen, Calvin Keung, Minjae Suh, and Tae Wan Kim. 2021. Evaluation framework for bim-based vr applications in design phase. *Journal of Computational Design and Engineering*, 8, 3, 910–922.

[27] Max Krichenbauer, Goshiro Yamamoto, Takafumi Taketom, Christian Sandor, and Hirokazu Kato. 2017. Augmented reality versus virtual reality for 3d object manipulation. *IEEE transactions on visualization and computer graphics*, 24, 2, 1038–1048.

[28] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research methods in human-computer interaction*. Morgan Kaufmann.

[29] Patrick Maier and Gudrun Klinker. 2013. Evaluation of an augmented-reality-based 3d user interface to enhance the 3d-understanding of molecular chemistry. In *International Conference on Computer Supported Education*. Vol. 2. SciTePress, 294–302.

[30] James McCartney. 1996. SuperCollider: a new real time synthesis language. In *Proc. International Computer Music Conference (ICMC'96)*, 257–258.

[31] Frank Melchior, Chris Pike, Matthew Brooks, and Stuart Grace. 2013. On the use of a haptic feedback device for sound source control in spatial audio systems. In *Audio Engineering Society Convention 134*. Audio Engineering Society.

[32] Georgios M Minopoulos, Vasileios A Memos, Konstantinos D Stergiou, Christos L Stergiou, and Konstantinos E Psannis. 2023. A medical image visualization technique assisted with ai-based haptic feedback for robotic surgery and healthcare. *Applied Sciences*, 13, 6, 3592.

[33] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9, 2, e89642.

[34] Miller S. Puckette. 1997. Pure Data. In *International Computer Music Conference 1997. Proceedings: Thessaloniki, Hellas. 25-30 september 1997*. The International Computer Music Association. Thessaloniki, Greece.

[35] Zhimin Ren, Ravish Mehra, Jason Coposky, and Ming Lin. 2012. Designing virtual instruments with touch-enabled interface. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 433–436.

[36] Giovanni Santini. 2019. Composing space in the space: an Augmented and Virtual Reality sound spatialization system. *Sound and Music Computing 2019*, 229–233.

[37] Eitan Schechtman, Talia Shrem, and Leon Y Deouell. 2012. Spatial localization of auditory stimuli in human auditory cortex is based on both head-independent and head-centered coordinate systems. *Journal of Neuroscience*, 32, 39, 13501–13509.

[38] Jaeyoung Shin and Jin-Kook Lee. 2019. Indoor walkability index: bim-enabled approach to quantifying building circulation. *Automation in Construction*, 106, 102845.

[39] Junbong Song, Sungmin Cho, Seung-Yeob Baek, Kunwoo Lee, and Hyunwoo Bang. 2014. Gafinc: gaze and finger control interface for 3d model manipulation in cad application. *Computer-Aided Design*, 46, 239–245.

[40] Yang Song, Richard Koeck, and Shan Luo. 2021. Review and analysis of augmented reality (ar) literature for digital fabrication in architecture. *Automation in construction*, 128, 103762.

[41] Wenxin Sun, Mengjie Huang, Chenxin Wu, Rui Yang, Yong Yue, and Miaomiao Jiang. 2024. Tangible and mid-air interactions in hand-held augmented reality for upper limb rehabilitation: an evaluation of user experience and motor performance. *International Journal of Human–Computer Interaction*, 1–18.

[42] Toqeer Ali Syed, Muhammad Shoaib Siddiqui, Hurria Binte Abdullah, Salman Jan, Abdallah Namoun, Ali Alzahrani, Adnan Nadeem, and Ahmad B Alkhodre. 2022. In-depth review of augmented reality: tracking technologies, development tools, ar displays, collaborative ar, and security concerns. *Sensors*, 23, 1, 146.

[43] Michael Walker, Hooman Hedayati, Jennifer Lee, and Daniel Szafir. 2018. Communicating robot motion intent with augmented reality. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 316–324.

[44] Frederic L Wightman and Doris J Kistler. 1999. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America*, 105, 5, 2841–2853.

[45] Aaron Willette, Nachiketa Gargi, Eugene Kim, Julia Xu, Tanya Lai, and Anıl Çamcı. 2020. Cross-platform and cross-reality design of immersive sonic environments. In *Proc. Int. Conf. New Interfaces Musical Expression*, 127–130.

[46] Matthew Wright and Adrian Freed. 1997. Opensound control: a new protocol for communicating with sound synthesizers. In *International Computer Music Conference 1997. Proceedings: Thessaloniki, Hellas. 25-30 september 1997*. The International Computer Music Association, 101–104.

[47] Jing Yang, Amit Barde, and Mark Billinghurst. 2022. Audio augmented reality: a systematic review of technologies, applications, and future research directions. *journal of the audio engineering society*, 70, 10, 788–809.

[48] Adriel Yeo, Benjamin WJ Kwok, Angelene Joshna, Kan Chen, and Jeannie SA Lee. 2024. Entering the next dimension: a review of 3d user interfaces for virtual reality. *Electronics*, 13, 3, 600.

[49] Xuesong Zhang and Adalberto L Simeone. 2022. Using the think aloud protocol in an immersive virtual reality evaluation of a virtual twin. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction*, 1–8.

[50] Yunfang Zheng, Jacob Swanson, Janet Koehnke, and Jianwei Guan. 2022. Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: a review. *American journal of audiology*, 31, 3, 819–834.

[51] ZhiYing Zhou, Adrian David Cheok, Yan Qiu, and Xubo Yang. 2007. The role of 3-d sound in human reaction and performance in augmented reality environments. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37, 2, 262–272.

Received 20 June 2024; revised xx July 2024; accepted xx July 2024