EXPLORING SPATIALIZATION: A METHOD FOR SUBJECTIVE ASSESSMENT OF SOUNDSCAPE PREFERENCE USING IMMERSIVE ENVIRONMENTS

Milena Jonas Bem, Samuel Chabot, Jonas Braasch

Graduate Program in Architectural Acoustics
Rensselaer Polytechnic Institute
Greene Building, 110 8th St, Troy, NY 12180, USA
jonasm@rpi.edu

ABSTRACT

This research explores the application of immersive environments, like the Panorama Screen System, for perception and psychoacoustic experiments. A system is developed to present architectural scenarios visually and sonically by using panoramic images combined with spatialized sound. These audiovisual results are displayed on a human-scale panoramic display, which is integrated with eight loudspeakers for spatial audio rendering. While immersed in this system, participants use a wireless interface as a controller to interact with the experiment and give subjective ratings for their preferences and perceptions of these scenarios. The findings show consistent patterns in preferences across different spatialized soundscapes, indicating the effectiveness of this method and the potential of immersive environments for gathering subjective assessments. This research demonstrates the practical use of spatialization techniques and how they can enhance immersion and facilitate the communication of sonic environments.

1. INTRODUCTION

Spatialization, the technique of positioning auditory sources within physical or virtual spaces to create a sense of location or movement for sounds, plays a crucial role in enhancing immersive audio-visual experiences. By systematically arranging and manipulating sound, spatialization allows sound sources to be perceived as if they were coming from specific directions and distances within the listening environment. Spatial audio has gained traction in evaluating acoustic environments, especially in soundscape studies, with the emergence of virtual reality (VR) [1, 2, 3] and augmented reality (AR) [4, 5, 6] technologies.

According to the ISO 12913-1 standard, a soundscape is defined as the "acoustic environment as perceived and understood by a person or people in context" [7]. Subjective studies are, therefore, essential for understanding the perception of sound environments and correlations with objective acoustic parameters in different settings. Assessment methods such as soundwalks, questionnaires, and guided interviews are commonly employed to investigate real or virtual (reproduced or synthesized) acoustic environments, both indoors and outdoors. These data collection tools and methods can be applied in situ or a laboratory [7].

In laboratory experiments, soundscape design and evaluation require appropriate test designs and rely on sophisticated spatial audio tools for sound capture and 3D rendering over headphones or speaker arrays [8]. Spatialized audio plays a fundamental role in this context and is essential to achieving natural realism to complement immersive visuals. Various techniques such as Vector Based Amplitude Panning (VBAP), binaural presentation with Head-Related Transfer Functions (HRTFs), and Wave-Field Synthesis (WFS) can be employed to create specialized sound [9]. These techniques decode or process recordings to recreate the immersive sound experience. In soundscape studies, ambisonics is the most commonly used recording technique, although general microphone arrays with suitable post-processing can also be employed [8].

Regarding sound reproduction, soundscape studies initially involved stereo configurations [10]. After 1950, multi-channel reproduction methods became popular and have been widely used in the reproduction of acoustic environments [11, 12]. More recently, wave field synthesis and ambisonics have been the other two physical reconstruction techniques that aim to create the same acoustical pressure field as the one present in the surroundings [13, 6]. The FOA-2D speaker array exhibited higher spatial acoustic fidelity and was perceived as more immersive and realistic than FOA-tracked binaural and FOA-static binaural methods in soundscape applications [14].

The environments in which spatial sounds are reproduced play crucial roles in the user experience. Immersive virtual environments (IVE) have become the forefront of experiential spaces and are a promising technology for exploring user experiences in response to various environmental stimuli (e.g., visual, thermal, and acoustic) [15]. This enables precise control of extraneous and independent variables, particularly when physical locations are inaccessible, restricted, or not yet built. These environments, increasingly applied in building science and occupant comfort research, provide ecologically valid tools in soundscape experiments [16], offering advantages over head-mounted displays in terms of group experiences, audio-visual congruency, and reduced fatigue and motion sickness [17]. Human-scale immersive environments, a type of IVE in which a large screen surrounds the user, offer significant potential in the soundscape assessment domain due to their potential to create perceptually accurate audio-visual scenes [8, 18], which is fundamental in the soundscape and perceptual studies.

Notable examples of immersive environments include the

Imaginarium at Deakin University, Reality Deck 2 at SUNY Stony Brook, and Allosphere at the University of California [17]. In addition to these examples, Rensselaer Polytechnic Institute (RPI) hosts two human-scale immersive environments, the Collaborative Research Augmented Immersive Virtual Environment Laboratory (CRAIVE-Lab) [9] and the Collaborative Immersive Room (CIR) [19]. These facilities, equipped with sophisticated video and audio systems, enable a wide range of research across different fields, offering a great potential to address the lack of acoustic stimuli in IVE-based multi-domain experiments [15]. Moreover, subjective studies, such as the assessment and reporting of soundscapes, often rely on questionnaires. By integrating questionnaires into human-scale immersive environments, researchers can administer assessments seamlessly and interactively [20], offering a more immersive and dynamic evaluation approach compared to conventional laboratory setups.

This research leverages the CIR to introduce a controller developed for subjective assessments in a human-scale immersive environment specifically designed to facilitate questionnaire administration while participants are immersed within the audio-visual experience. Methods for creating visual and audio content are presented to provide insights for future research. Finally, a practical application of this subjective assessment controller in a soundscape preference experiment is demonstrated.

2. COGNITIVE IMMERSIVE ROOM

The Cognitive Immersive Room (CIR) represents an immersive virtual environment developed within the cutting-edge infrastructure of the Experimental Media and Performing Arts Center (EMPAC) at Rensselaer Polytechnic Institute (RPI). Since its establishment in 2000, EMPAC has consistently maintained a leading-edge position in technological innovation. With an investment of 200 million, EMPAC is one of the world's largest multipurpose spaces [19]. The unique structure is home to an equally innovative program that combines technology and art [21], including a 1,200-seat concert hall, a 400-seat theatre, and two black-box studios spanning 2,500 and 3,500 square feet, respectively. All these rooms are distinguished by their flexible design and key technology, which allows several configurations and possibilities. All rooms are equipped with multimedia capabilities, and the integration of auditory and visual domains has been prioritized since their inception.

These shared attributes were essential in the deliberate selection of one of EMPAC's black-box studios (Studio 2) as the optimal location for housing the CIR. The enclosing room, Studio 2, spans dimensions of $18,20\,\text{m} \times 14,92\,\text{m}$ $(59' - 8'' \times 48' - 11'')$, resulting in a floor area of 2475 square feet (230 square meters), and a distance of $5.6\,\text{m}$ $(18' - 5'')$ from floor to walkable grid, with the screen positioned at its center. To optimize acoustics, computer-milled acoustic panels cover the four lateral walls specifically designed for this space (see Figure 1. These panels incorporate randomized holes of various sizes and depths, promoting the diffusion of a wide frequency spectrum. Additionally, absorptive banners can be deployed from the ceiling to overlay the diffusive walls.

Figure 1: Computer-milled acoustic panels designed for Studio 2.

Studio 2 rests on a separate foundation to prevent structure-born vibration transition. The background noise level was measured using Room Criteria of RC6, reflecting an exceptionally low noise floor below the threshold of human hearing. The ambient sounds naturally present in the room include the 'hum' from the projectors and the 'hum' generated by the computer fan. As measured by an Extech 407740 Sound Level Meter, background sound levels have been reported at 41.1 dBA with the projectors on, and the room's reverberation time measures about $0.625s$ at middle frequencies with deployed banners. Catwalks and control rooms situated above and outside the studios ensure unobstructed space (See Figure 2).

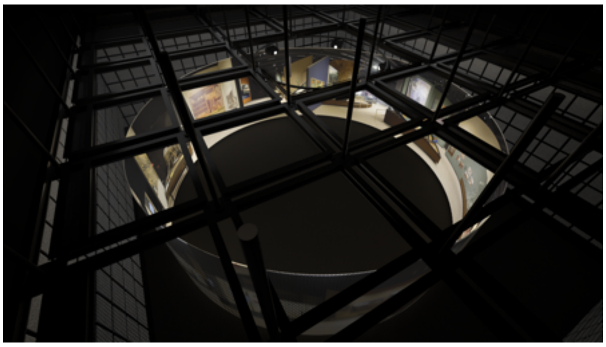


Figure 2: View from the wire grid ceiling above the screen that provides a walkable surface and a framework for suspending equipment, such as the five short-throw projectors, motion trackers, or any other necessary apparatus for specific applications.

2.1. Video Rendering

The video rendering capabilities of CIR are based on the PanoramaScreen, a collaborative development between Jeffrey Shaw's team at UNSW and Bernd Lintermann's team at ZKM [19]. Serving as a test environment for immersive audiovisual experiments at a human scale, the video system comprises a front-projection setup consisting of five projectors and a 360° projector screen. The screen is a perfect circle in shape, measuring 12 m in diameter, with a door opening of 85 cm wide, and a metal frame behind the screen serves as a frame and keeps the screen taut and smooth in the corners.

The front-projection screen stands 4.3 m in height. The screen surface area is $158\,\mathrm{m}^2$, and the screen form volume is $511\,\mathrm{m}^2$. Measuring from the screen's center, the vertical field of view reaches a maximum of approximately 36°. While typically capturing the most pertinent elements within the scene, this field of view may occasionally truncate floor, ceiling, or vertically extended content. The expansive floor area of the screen, totaling $113\,\mathrm{m}^2$ $(1216\,\mathrm{ft}^2)$, can accommodate up to 49 participants simultaneously (in accordance with fire code regulations), surrounding them in the immersive experience. There is a single entrance that disrupts the screen continuity, and due to its cylindrical form, there is no intrinsically exerted particular "front" to the space.

The five projectors are short-throw, allowing users to stand about one meter away from the screen without creating shadows on the display. The five projectors employed in the setup are carefully aligned to overlap their images, resulting in a seamless, continuous display across the entire screen. Utilizing the Pixelwix software, the projectors and screen undergo meticulous calibration to ensure precision. Additionally, Pixelwix facilitates the seamless stitching of the projectors' output, correcting for the room's geometry and seamlessly blending the individual projections into a cohesive, unified display. Each projector operates at a resolution of 1920-by-1080 pixels, collectively forming a unified and uninterrupted desktop spanning 9600-by-1080 pixels.

A single custom workstation equipped with two NVIDIA Quadro P5000 cards renders seamless content across all five projectors. The audio and visual control machines are kept separate and distinct. They are located outside the environment to maintain an uninterrupted immersive experience, but the display machine can be controlled on the screen within the space. Additionally, Kinect sensors are positioned below the base of the screen to detect gestures and monitor head orientation. Time-of-flight sensors are also installed in the ceiling to track users' spatial positions.

2.2. Audio rendering

Regarding audio, spatial audio is a primary driver for the CIR system design and is crucial for this multimodal immersion facility. The CIR's video and audio rendering capability together deliver congruent audiovisual content with balanced fidelity. This capability allows the sound field to be coextensive with the image field, which means the sound emanates from its corresponding locations within the image space.

Eight JBL 308p MkII Powered Monitor loudspeakers are strategically arranged behind the screen at ear level and equidistant from the center to optimize spatial audio rendering (See Figure 3. This setup offers flexibility, accommodating various configurations, including the potential expansion to a 512-channel wave field synthesis array, making this space capable of multiple spatial audio techniques.

The screen is a Gerriets GAMMALUX, Matte-White Finish Micro-Perforated Projection Screen. This micro-perforated PVC material is acoustically transparent and plays a crucial role in sound transmission, allowing the eight loudspeakers to be positioned behind it for sonification purposes. This unique feature ensures seamless integration of audio and visual elements, with the loudspeakers remaining concealed behind the screen. Thus, the immersive experience is enhanced while offering flexibility in loudspeaker placement. The loudspeakers are individually addressable via a Mac mini with a MOTU 16A. The machine that controls the audio system is located outside the screen area.

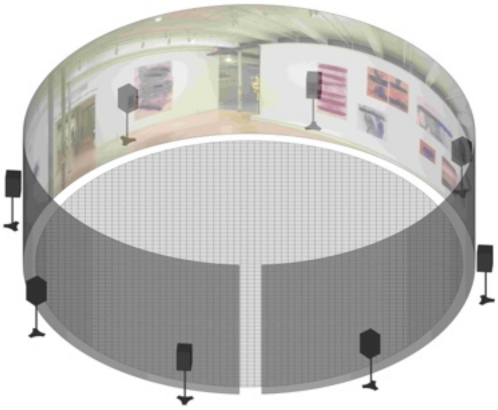


Figure 3: Speaker configuration around the panoramic visual display.

3. CONTENT CREATION FOR IMMERSIVE EXPERIMENTS

3.1. Visual Content Creation

Creating visual content for the panoramic screen involves various methods, each offering unique advantages and challenges. This section primarily focuses on commonly utilized methods for generating visual content, not just for soundscape assessment experiments but also for broader applications. Regardless of the chosen method, it is important to ensure that visualizations are tailored to the entire width of the screen for optimal presentation. This often involves configuring the workspace within software programs to match the specific pixel dimensions of the panoramic display. This section presents techniques for creating immersive visual content, including photography, 360° recording, street view imagery, online datasets, and computer-generated environments.

3.1.1. 360° Photography and HDR Imaging

Photography remains a fundamental method for capturing visual cues in immersive environments. Static panoramic photos, particularly those employing High Dynamic Range (HDR) imaging, offer detailed and realistic representations of physical spaces. The process involves visiting the location with specialized equipment, such as a Panasonic Lumix DMC-GH4 4K camera paired with a Rokinon 12-mm T2.2 lens or similar, mounted on a rotating tripod, preferably has angle markings in the horizontal plane, to capture a full 360° view. To minimize disruptions to the scenes, remote control via a mobile phone app is important during the capture process. The camera should be approximately at eye level. Post-processing using software like Adobe Lightroom is then necessary to merge and edit the images seamlessly. Figure 4 shows an example of an image at the end of this process. While effective, this method can be time-consuming and requires physical visits to capture locations, posing challenges due to geographical limitations.

Figure 4: This image is an example of 360° HDR static photo. To create this image, the 360° field was divided into 30° angles, with 7 photos taken at each interval to capture a range of exposures for HDR processing to ensure optimal image quality across different lighting conditions. After capturing and processing the 12 HDR photos, Adobe Lightroom was utilized to merge and edit them seamlessly.

3.1.2. 360° Recording Devices

360° recording devices, such as the Freedom360 GoPro rig, enable the capture of immersive videos covering the entire field of view. These devices, consisting of multiple cameras mounted in a specific configuration, record simultaneous footage from all directions. After recording, specialized stitching software synchronizes and blends the footage to create seamless panoramic videos that can be encoded for playback on the panoramic screen. This method immerses viewers in a dynamic visual experience, complementing other static imagery methods.

3.1.3. Street View Imagery

Online datasets, particularly Google Street View imagery, offer a convenient alternative for sourcing visual cues. The Pin-Up platform [22] allows users to search for desired locations and retrieve accurately positioned images on the panoramic screen (See Figure 5. As this crowdsourced dataset is typically viewed on flat displays like computer monitors, the Pin-Up integrates the Street View Static API to ensure that the content is suitable for the immersive display [18]. This method offers a convenient alternative to address some of the disadvantages of the first method of capturing static panoramic photos on location. However, it is only possible to project internal and external environments that have already been documented.

3.1.4. Computer-Generated Environments

Computer-generated environments provide extensive creative freedom for visual content creation. Within architectural visualization workflows, a variety of software tools are available to create immersive 360° images and animations. Utilizing 3D modeling software such as SketchUp, Rhino, or Blender, users can construct intricate virtual spaces with precision and versatility. Rendering engines like Unity, Unreal, and V-Ray then transform these models into visually stunning 360° images or animations. Each of these software tools has its unique strengths and applications, catering to various requirements. While some specialize in static image rendering, others excel in animation production, and a few integrate both 3D modeling and rendering capabilities seamlessly. This method is widely used by researchers and students who use CIR to simulate architectural spaces. For example, they can project a 1:1-scale simulation of building design on the screen by producing panoramic visual content, which can facilitate visualization and judgments of the design [23, 24]. While mastering these tools requires technical proficiency, online 3D databases like the Library for Universal Virtual Reality Experiments (LuVre) offer accessible alternatives for users with time or knowledge constraints [25].

Figure 5: On the screen, the overlaid navigation interface allows users to select Street View imagery for display on the panoramic screen. While some locations offer indoor views, like the Bavarian State Opera in Munich, Germany, projected in this photo, outdoor views are also accessible for all locations with Street View content, as indicated by the highlighted blue streets.

3.2. Spatial Audio Content Creation

Spatial audio plays a crucial role in immersive environments, yet its capture or generation can be challenging for architects or designers without specialized knowledge in sound engineering. This section introduces alternative methods for creating aural cues for subjective assessment experiments, including ambisonic recordings and room acoustics simulations.

3.2.1. Ambisonic Recordings

Ambisonic recordings represent one effective method involving the capture of field recordings using ambisonic microphones. Ambisonics is the leading recording technique for interactive spatial audio reproduction in soundscape studies [8]. Ambisonic microphones are typically equipped with four cardioid capsules arranged in a tetrahedral configuration, enabling full 360° capture. The resulting recordings yield four directional channels, known as A-format, which are then converted to B-format comprising an omnidirectional signal and three figure-8 signals corresponding to the X, Y, and Z Cartesian axes. Some commercially available plug-ins streamline this conversion process, ensuring intuitive handling. When played back on a multichannel sound system, such as the one in the Collaborative Immersive Room (CIR), participants perceive sounds emanating from precisely recorded directions in the room, enhancing immersiveness. For use in the laboratory, ambisonic microphones combined with a microphone preamplifier capture high-quality audio in the A-forma FOA (See Figure 6. Subsequently, ambisonic recordings can be edited and converted using digital audio workstation software, such as Reaper, equipped with an A-B format plug-in. This workflow presents numerous advantages: the equipment is portable, making it suitable for fieldwork; it provides a robust spatialization option for immersive experiences; and it is relatively straightforward to execute, even for individuals without specialized sound expertise, such as designers and architects.

Figure 6: Setup example to ambisonic recordings - H3-VR (Zoom) and Sennheiser (AMBEO) microphones combined with a Mix Pre 6 sound device. This approach offers efficiency and portability, enabling the reproduction of comprehensive spatial audio information with relatively compact equipment. Note that in this image, both microphones were together, but just one is necessary to do the recordings; in this case, we used both to compare the results.

3.2.2. *Room Acoustics Simulations*

Room acoustics simulations offer another approach to creating aural cues for presentation in the CIR. This research will discuss two methods for generating Room Impulse Responses (RIRs): the 2D walkable auralization method [18] and room acoustic software, such as CATT-Acoustic.

Developed by Jonas Braasch and Samuel Chabot (2022) [18] for collaborative virtual reality systems, the 2D walkable auralization method is particularly focused on the acoustic learning process for students, offering a rapid method for understanding the acoustic consequences of room acoustic design. The basis is a 2D-floor plan with scale or known overall dimensions of the building. The floor plan is annotated using a bitmap editor (e.g., GIMP or Photoshop), with corners noted with red dots and columns with green dots. The original scale is also noted with two scale points. After annotation, the auralization program automatically creates the wall elements and generates the RIR, which can be convolved with a sufficiently dry record for presentation on the screen. The workflow is easily accessible to architects and designers. It has been utilized in the Aural Architecture course for students to create their walkable auralizations and display them on the screen.

Alternatively, virtual auralization for architectural simulations using software for room acoustic consulting and audio virtual reality offers a more time-consuming process that requires a certain level of acoustic knowledge to use the software and run the simulation. This method typically involves modeling the space in 3D and exporting it for simulation software, such as CATT-Acoustic, to sign the surface materials, specify source and receiver positions, measure the RIRs, and create the auralizations. Acoustic consultants often employ this method to present clients with the experience of listening to their as-yet unbuilt spaces [26, 27].

Several considerations apply to both room acoustic simulation methods. First, one RIR should be measured for each loudspeaker position in the immersive environment. For example, as the CIR has an 8-channel system, 8 virtual receivers will be used to measure the RIRs for each sound-source pair. Second, considering the aural sweet spot phenomenon is crucial, as the aural scene will be fidelity reproduced in one position within the screen by a finite number of loudspeakers with a fixed position. Finally, as computer-generated spaces are used to estimate the RIR measurements, these methods are particularly valuable for buildings in the design phase or where on-site RIR measurements are not feasible.

3.3. Controller Program

To streamline subjective assessment experiments in the Collaborative Immersive Room (CIR), a custom patch was developed using Max 8, a versatile programming environment by Cycling ’74. Originally designed for musical composition, Max has evolved to offer robust audio (MSP) and video processing tools (Jitter), making it an ideal platform for immersive environments like the CIR. The controller program oversees all steps of the experiment, from visual and spatial audio content manipulation to questionnaire development and final cue presentation.

3.3.1. *Audio and Visual Presentation*

Max is an open platform that allows for the integration of additional function objects and attributes independently developed to enhance the patch. Tools such as Virtual Microphone Control (ViMiC), facilitating real-time spatialization synthesis [28], and IRCAM Spat5, a dedicated real-time audio engine for spatialization, artificial reverberation, and sound diffusion [29], are readily accessible and manipulable through the integrated Open Sound Control protocol (OSC).

Spatialization plays a crucial role in aligning audio with on-screen visuals in immersive environments. However, not all users possess the time or expertise to create customized audio and spatialization solutions for their applications. To streamline the process of presenting audio across the array, an application programming interface (API) called Spatial Audio Worker (SAW) was developed for human-scale immersive environments [22]. SAW, built within the Max programming environment using the IRCAM Spat5 spatialization library [29], orchestrates the placement of virtual sound sources and loudspeakers within a virtual room using a coordinating position (See Figure 7).

The integration of SAW into the controller simplifies the spatialization process, particularly for users unfamiliar with spatialization techniques. It also enables developers to produce audio aligned with on-screen visuals without requiring in-depth knowledge of the system hardware. Users typically interact primarily with the general patch and deposit audio files in the same folder as the controller patch. Adjustments may be necessary depending on the method used to record or simulate the audio content.

Figure 7: The loudspeaker array configuration of the CIR within the Spat5 Viewer interface. Black boxes represent fixed loudspeakers, while green circles denote movable sound sources.

The visual presentation aspect has also been integrated seamlessly into the controller to facilitate the panoramic screen workflow. Simplifying the process, users are only required to create a designated folder for the experiment. Within this folder, they will find the pre-developed patcher controller alongside individual folders for each scenario they aim to simulate. In each scenario folder, users should drop the visual and sound stimuli.

For example, the controller's Max Patcher file, already developed to facilitate the process, is located within the "Experiment X" folder, accompanied by a "Scenario Y" folder. Within this scenario folder, users must include a .png file representing the image intended for projection on the screen alongside the sound file that corresponds to this scenario. If the researcher wants to test more than one scenario, they can create a folder for each scenario and add the images and sound files for each one into their respective folder. This straightforward organization streamlines the setup process, ensuring smooth navigation and execution of congruent sound-visual presentations within the system.

3.3.2. Questionnaire

Questionnaires serve as indispensable tools in subjective experiments across various domains, facilitating the systematic collection, quantification, and analysis of participants' subjective responses. They offer valuable insights into human behavior, cognition, and perception. Particularly in soundscape studies, questionnaires play a pivotal role in capturing and quantifying individuals' subjective perceptions, preferences, and experiences. Therefore, a key component of the controller program discussed in this paper is the integration of a questionnaire application.

The controller has a user interface that allows participants to interact with a wireless device, such as a tablet. This device will be used throughout the experiment to answer the questions and control the experiment. While the foundational programming aspects of the controller have already been established (See Figure 8, customization is necessary based on the experimental design and the specific questions researchers wish to pose.

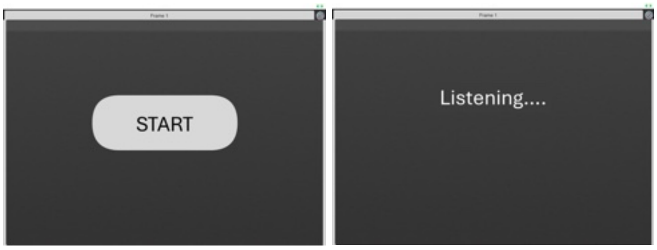
Nowadays, the participants press a "Start" button when they are ready to begin. From there, the first visual stimulus is presented on the screen simultaneously with the soundscape that is being analyzed for that environment. The participant experiences this scenario for the predetermined time set by the researcher; after this time, questions automatically begin to appear on the device screen. The participant answers the questions with a simple touch on the screen, and the next question appears. When all questions for that scenario are asked, the controller automatically moves on to the next visual and sound stimulus. Upon completion of the experiment, the controller generates a comprehensive report with the responses for subsequent data analysis.

One advantage of this application is that it enables researchers to allow participants to engage independently in an immersive environment without intervention. Furthermore, unlike research utilizing virtual reality headsets, where participants must frequently remove the equipment to answer questionnaires, this method ensures uninterrupted immersion throughout, fostering a more fluid and dynamic experience. Additionally, it streamlines the comparison between different scenarios, facilitating the detection of subtle changes in the acoustic and visual environment with greater efficiency. For example, when examining soundscape preference for a certain environment, the program allows the participant to replay different soundscapes to choose which one they prefer for that scenario.

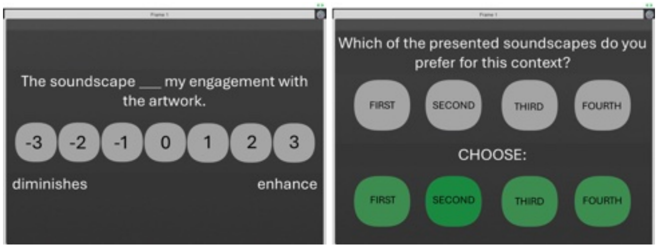
4. PRACTICAL APPLICATION

Previous efforts to streamline audio-visual experiments within Rensselaer Polytechnic Institute's (RPI) immersive spaces, such as the Spatial Audio Worker (SAW) [30] and the Pin-Up system [18], have significantly reduced the operational complexity in the Collaborative Immersive Room (CIR). However, no program has yet integrated all the existing systems with questionnaire administration and data report generation for perceptual and psychoacoustic experiments. Some advantages of these environments for psychoacoustic experiments have long been known, such as the quick and effective ABX comparisons between auditory scenes compared to traditional methods.

This advantage was demonstrated in previous museum acoustics research at RPI, where participants navigated visually and son-



(a) “Start” button (b) ‘Listening . . . ” message


The soundscape ___ my engagement with the artwork.

Which of the presented soundscapes do you prefer for this context?

(c) Question example (d) Soundscape comparation

Figure 8: Example of a questionnaire applied for subjective assessment in soundscape preference.

ically between New York museum soundscapes via Ambisonic recordings from 18 locations across 5 museums (See Figure 9, aiming to analyze soundscape preferences in museum environments [20]. In this project, the Controller Program described in Section 3.3 orchestrated all stages after the visual and audio content creation, including ambisonic manipulation, spatialization, questionnaire design, and final presentation.

Figure 9: Participant engaged in the audio-visual experiment, interacting with the controller via a wireless user interface (Apple, iPad) while answering the questionnaire.

Visual stimuli were generated using 360° Photography and HDR Imaging (Section 3.1.1), while audio content was captured through Ambisonic Recordings (Section 3.2.1). After this content creation, all the steps were performed within the controller program in Max. Ambisonic recordings in A-format were converted to B-format FOA using an A-B converter plug-in (Sennheiser, AMBEO A-B format Converter), further processed using the SAW, facilitating multichannel loudspeaker playback in the 8-channel CIR setup (See Figure 3. SPL measurements taken during each recording ensured consistent playback levels later in the laboratory environment, enabling users to discern changes not only in texture and spatial composition but also in absolute dynamic range across museum rooms.

The original ambisonic recordings of the real auditory environment in each museum were posteriorly augmented with additional sounds to compose different soundscapes. Four different soundscape scenarios were presented for each museum room, including the original recorded soundscape, the original recording combined with sounds in congruency with each exhibition, such as music or natural sounds, the original recording combined with conventional sound masker signals (white noise), and “silence” (just the background noise in the CIR room). Therefore, this project intersects with augmented reality (AR) audio, expanding real sound environments with virtual auditory elements to create different scenarios, adding to a real environment rather than constructing a wholly synthetic sound environment.

The controller program combined all four sound stimuli scenarios with visual stimuli to streamline the experimental procedure. Participants interacted with the controller using a wireless user interface (Apple iPad), experiencing soundscapes and visual stimuli and answering questions simultaneously. Random presentation of audio-visual stimuli mitigated memory bias, and upon completion, the controller generated a comprehensive data report for statistical analysis.

The study produced statistically significant results, emphasizing that combining congruent sounds in the museum soundscape has the potential to improve engagement and immersiveness and reduce distractions. It was possible to prove that the majority of participants preferred soundscapes aligned with museum exhibitions, demonstrating the potential of immersive environments and the developed controller program in psychoacoustic experiences. After the experiments, participants’ primary comments were that they appreciated their ability to interact with the experiment and answer all questions interactively, only using the user interface, appreciating the interactive and straightforward nature of the experiment.

This pilot study lays the groundwork for future psychoacoustic experiments in the CIR, offering a versatile platform for facilitating quick and direct comparisons between various visual and auditory scenarios. Beyond museum contexts, the controller system allows researchers to conduct subjective evaluations while presenting simultaneously visually and sonically different environments. Additionally, it opens avenues for indoor and outdoor soundscape research across diverse spaces, further expanding the utility of immersive environments in academic and practical settings.

5. ACKNOWLEDGMENT

This research was supported by the National Science Foundation under Grant No. HCC-1909229.

6. REFERENCES

[1] F. Ruotolo, L. Maffei, M. Di Gabriele, T. Iachini, M. Masullo, G. Ruggiero, and V. P. Senese, “Immersive virtual reality and environmental noise assessment: An innovative audio–visual approach,” *Environmental impact assessment review*, vol. 41, pp. 10–20, 2013.

[2] M. Luigi, M. Massimiliano, P. Aniello, R. Gennaro, and P. R. Virginia, “On the validity of immersive virtual reality as tool for multisensory evaluation of urban spaces,” *Energy procedia*, vol. 78, pp. 471–476, 2015.

[3] L. Maffei, M. Masullo, A. Pascale, G. Ruggiero, and V. P. Romero, “Immersive virtual reality in community planning: Acoustic and visual congruence of simulated vs real world,” *Sustainable Cities and Society*, vol. 27, pp. 338–345, 2016.

[4] J. Lacey, “Site-specific soundscape design for the creation of sonic architectures and the emergent voices of buildings,” *Buildings*, vol. 4, no. 1, pp. 1–24, 2014.

[5] R. Ranjan and W.-S. Gan, “Natural listening over headphones in augmented reality using adaptive filtering techniques,” *IEEE/ACM transactions on audio, speech, and language processing*, vol. 23, no. 11, pp. 1988–2002, 2015.

[6] B. Boren, M. Musick, J. Grossman, and A. Roginska, “I hear ny4d: Hybrid acoustic and augmented auditory display for urban soundscapes,” in *20th Int. Conf. on Auditory Display (ICAD2014)*, New York, USA (June 22–25, 2014).

[7] I. ISO 12913-2, “Iso/ts 12913-2: 2018 acoustics—soundscape—part 2: Data collection and reporting requirements,” 2018.

[8] J. Y. Hong, J. He, B. Lam, R. Gupta, and W.-S. Gan, “Spatial audio for soundscape design: Recording and reproduction,” *Applied sciences*, vol. 7, no. 6, p. 627, 2017.

[9] S. Chabot and J. Braasch, “An immersive virtual environment for congruent audio-visual spatialized data sonifications,” in *23rd Int. Conf. on Auditory Display (ICAD2017)*, Pennsylvania, USA (June 20-23, 2017).

[10] A. D. Blumlein, “British patent specification 394,325 (improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems),” *Journal of the Audio Engineering Society*, vol. 6, no. 2, pp. 91–130, 1958.

[11] C. Guastavino and B. F. Katz, “Perceptual evaluation of multi-dimensional spatial audio reproduction,” *The Journal of the Acoustical Society of America*, vol. 116, no. 2, pp. 1105–1115, 2004.

[12] C. Guastavino, B. F. Katz, J.-D. Polack, D. J. Levitin, and D. Dubois, “Ecological validity of soundscape reproduction,” *Acta Acustica united with Acustica*, vol. 91, no. 2, pp. 333–341, 2005.

[13] W. J. Davies, N. S. Bruce, and J. E. Murphy, “Soundscape reproduction and synthesis,” *Acta Acustica United with Acustica*, vol. 100, no. 2, pp. 285–292, 2014.

[14] J. Y. Hong, B. Lam, Z.-T. Ong, K. Ooi, W.-S. Gan, J. Kang, J. Feng, and S.-T. Tan, “Quality assessment of acoustic environment reproduction methods for cinematic virtual reality in soundscape applications,” *Building and environment*, vol. 149, pp. 1–14, 2019.

[15] H. Alamirah, M. Schweiker, and E. Azar, “Immersive virtual environments for occupant comfort and adaptive behavior research–a comprehensive review of tools and applications,” *Building and Environment*, vol. 207, p. 108396, 2022.

[16] C. Xu, T. Oberman, F. Aletta, H. Tong, and J. Kang, “Ecological validity of immersive virtual reality (ivr) techniques for the perception of urban sound environments,” in *Acoustics*, vol. 3, no. 1, 2020, pp. 11–24.

[17] T. Höllerer, J. Kuchera-Morin, and X. Amatriain, “The allosphere: a large-scale immersive surround-view instrument,” in *Proceedings of the 2007 workshop on Emerging displays technologies: images and beyond: the future of displays and interacton*, California, USA (August 4th, 2007).

[18] S. Chabot and J. Braasch, “Walkable auralizations for experiential learning in an immersive classroom,” *The Journal of the Acoustical Society of America*, vol. 152, no. 2, pp. 899–910, 2022.

[19] J. Braasch, J. Goebel, and T. Vos, “A cinematic spatial sound display for panorama video applications,” *Organised Sound*, vol. 15, no. 3, pp. 260–270, 2010.

[20] M. J. Bem, “Effects of sounds on the visitors’ experience in museums,” Master’s thesis, Rensselaer Polytechnic Institute, Troy, NY, 2023.

[21] C. M. Schwitter and D. Gallagher, “Designed for performance,” *Civil Engineering Magazine Archive*, vol. 79, no. 9, pp. 60–71, 2009.

[22] S. R. V. Chabot, “Augmenting collaborative immersive systems through cognitive technologies for educational scenarios,” Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy, NY, 2022.

[23] R. L. Elder, “Influence of immersive human scale architectural representation on design judgment,” Master’s thesis, Rensselaer Polytechnic Institute, Troy, NY, 2017.

[24] M. Huang, S. Chabot, C. Leitão, T. Krueger, and J. Braasch, “Spatially-aware group interaction design framework for collaborative room-oriented immersive systems,” *Applied Ergonomics*, vol. 113, p. 104076, 2023.

[25] B. Schöne, J. Kisker, R. S. Sylvester, E. L. Radtke, and T. Gruber, “Library for universal virtual reality experiments (luvre): A standardized immersive 3d/360 picture and video database for vr based research,” *Current Psychology*, vol. 42, no. 7, pp. 5366–5384, 2023.

[26] M. Azevedo and J. Sacks, “Auralization as an architectural design tool,” in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, 2014, pp. 162–168.

[27] D. Thery, V. Boccara, and B. F. Katz, “Auralization uses in acoustical design: A survey study of acoustical consultants,” *The Journal of the Acoustical Society of America*, vol. 145, no. 6, pp. 3446–3456, 2019.

[28] N. Peters, T. Matthews, J. Braasch, and S. McAdams, “Spatial sound rendering in max/msp with vimic,” in *Proceedings of the 2008 International Computer Music Conference*, Belfast, Ireland (August 24-29, 2008).

[29] T. Carpentier, “A new implementation of spat in max,” in *15th Sound and Music Computing Conference (SMC2018)*, Cyprus, Chipre (July 4-7, 2018), pp. 184–191.

[30] S. Chabot and J. Braasch, “Interactive application to control and rapid-prototype in a collaborative immersive environment,” in *Audio Engineering Society Convention 151*, Online (October, 2021).