SoundHolo: Sonically Augmenting Everyday Objects and the Space Around Them

Daniel Rosenberg Munoz
School of Design
Carnegie Mellon University
Pittsburgh, PA, USA
drosenb2@andrew.cmu.edu

Miles Scharff
School of the Arts
Columbia University
New York, NY, USA
mjs2436@columbia.edu

Brandt YouRan Li
School of Design
Carnegie Mellon University
Pittsburgh, PA, USA
youranl@andrew.cmu.edu

Kimberly Blacutt
School of Design
Carnegie Mellon University
Pittsburgh, PA, USA
kblacutt@andrew.cmu.edu

ABSTRACT
SoundHolo is a room-scale platform for sonically augmenting everyday objects by dynamically attaching 3D sounds to them and their surrounding space. By integrating wave field synthesis and motion tracking within a 20x30x12 foot room, SoundHolo explores how 3D sounds can augment any object without relying on embedded electronics or personal devices. Augmenting everyday objects with interactive output capabilities is a long-standing topic in HCI research. We contribute by enabling any physical object to emit sound, regardless of shape or material constraints, and allowing public, device-free experiences. We implemented SoundHolo along with a matrix demonstrating eighteen output variations, offering a palette of design options for HCI researchers developing sonic augmentations. This paper integrates visuals and texts detailing the platform implementation, application scenarios, and preliminary evaluation.

AUTHORS KEY WORDS Tangible Interaction, Wave Field Synthesis, Sonic Interaction Design

TEI '25, March 4–7, 2025, Bordeaux / Talence, France

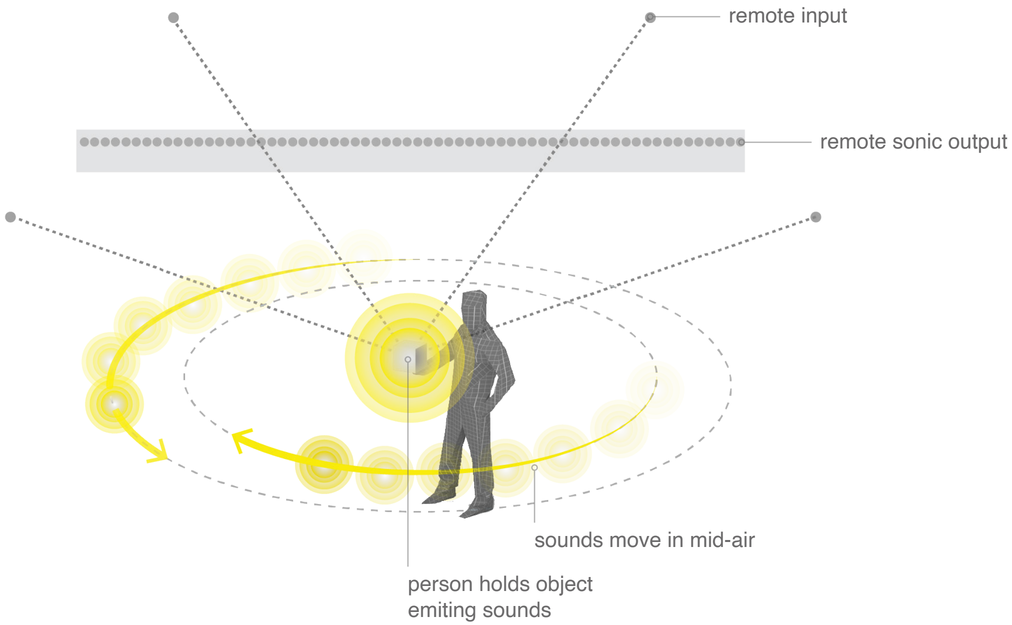
ACM ISBN 979-8-4007-1197-8/25/03.
https://doi.org/10.1145/3689050.3704416

CSS CONCEPTS
Human-centered computing, Interaction design

INTRODUCTION
Augmenting the physical world with interactive output capabilities has been a central focus of HCI research, particularly in ubiquitous computing [26, 6] and TUI work [12, 24]. Most applications augment everyday objects by embedding electronic components and speakers within their material surfaces [23]. For certain objects, however, their geometry or visual and tactile features constitute a constraint for the seamless integration of the required electronics [10, 31]. Embedding hardware within a curved and transparent plastic bottle filled with water, for example, could be challenging considering its curved geometry, transparency, flexibility, and drinkable content. This constraint has limited the form and materiality of interactive everyday objects [10, 31]. Alternatively, other applications use wearable headsets or headphones to augment everyday objects sonically. In Audio Augmented Reality, for example, sounds can be spatially co-located with objects in the physical world [4, 30]. This sonic augmentation is delivered directly through the user's wearable devices rather than the speakers embedded in the objects. As a result, only the person wearing the device experiences the augmentation. In public settings, the difference in user experience between the person using the device and spectators, who observe without a device, may raise social acceptability issues [2, 15].

We introduce SoundHolo, a room-scale platform that integrates wave field synthesis (WFS) and motion-tracking to dynamically attach sound sources to physical objects in motion or locate them in mid-air. Using SoundHolo, we investigate how sound can augment everyday objects and the space between and around them without relying on embedded electronics or personal devices. With SoundHolo, any graspable thing can emit sound, including crafted objects, organic materials, and even edible items, expanding the repertoire of materials that HCI researchers can use in sonically augmented applications. With SoundHolo, all users in a room can perceive the same sonic augmentations without personal devices, mitigating potential social acceptability issues in public spaces.

We designed and implemented a 64-channel speaker array integrated with ten infrared motion-tracking cameras in an empty 20x30x12-foot room. To our knowledge, SoundHolo is the first interactive WFS system to employ a high-resolution speaker array in a large space, enabling multiple users to engage simultaneously. To control this system, we developed the workflow and software integration for users to specify the number, configuration, and motion of sound sources attached to objects or in mid-air around them. Using this workflow, we developed the SoundHolo Matrix, demonstrating 18 combinations visually. With this matrix, we aim to inspire HCI researchers to explore the proposed combinations in future interactive applications.



In short, we contribute:

- A concept for providing a room-scale sonic augmentation to physical objects through interactive wave field synthesis.
- A working implementation integrating a custom-made high-resolution 64-channel array with a motion tracking system.
- A visual matrix demonstrating 18 possible sonic augmentations to inspire future interactive explorations in HCI.
- Application scenarios demonstrating potential real-world opportunities for SoundHolo.
- Results from preliminary evaluation discussing the perceptual efficacy of the system.

We propose three user scenarios to show SoundHolo's potential real-world applications, including (1) immersive soundscapes in branded experiences, (2) storytelling merchandise in retail stores, and (3) spatial auditory displays in learning environments. In a preliminary evaluation, we tested the matrix's perceptual efficacy. Our findings indicate that while all combinations can be technically generated, those that augment physical objects and their surrounding space are perceived more effectively. We conclude by reflecting on the work and insights to discuss future research directions.

RELATED WORK

Sonic TUI

SoundHolo builds on HCI research on tangible interactions. Our aim of sonically augmenting everyday objects aligns with the HCI goal of weaving computational capabilities seamlessly with the objects of everyday life [26, 27]. Since the beginning of research on tangible interactions, the use of sound as a responsive medium has been explored. Paradigmatic examples, such as the Marble Answering Machine [24] and the MusicBottles [13], demonstrate how sounds can be activated using objects in the physical world. For instance, placing a marble on a machine releases a voice message, and opening a bottle triggers a musical sound. In these examples, however, the sound is emitted from supporting surfaces–a machine and a table–not the objects themselves. With SoundHolo, not only could the sounds be emitted directly from marbles and bottles, but they could also move through space between and around them dynamically.

Other HCI research has explored embedding electronics in physical materials to create sonic experiences. ListenTree [23], for example, turns a tree into a living speaker by attaching an audio exciter to its trunk underground. SounHolo allows smaller organic objects like tree branches, rocks, fruits, and flowers to emit sounds without electronic attachments. Alternatively, directional audio has been used to augment objects by projecting sounds onto them from a distance. In Digital Ventriloquism, an actuated ultrasonic array directs audio signals into a space, making them perceptible only when they hit physical objects [11]. SoundHolo achieves a similar sonic augmentation for multiple objects simultaneously without mechanical actuation. Finally, the physical world has been sonically augmented through wearable devices. For example, the Hearing is Believing system uses headphones and helmets to enable users to hear sounds emanating from objects in their environment [28]. SoundHolo removes the need for personal devices, facilitating reality-based interactions [14] seamlessly integrated into the physical world.

Interactive WFS

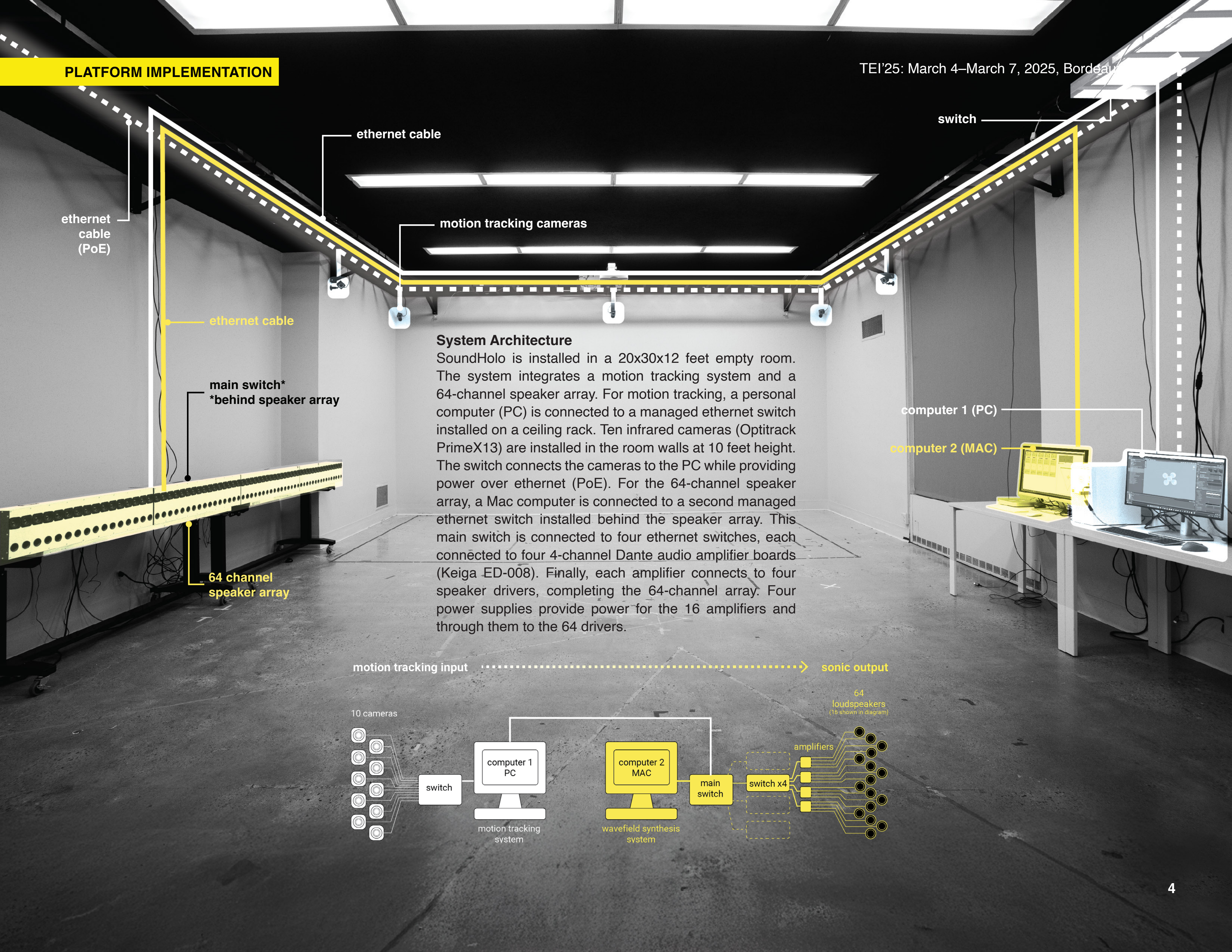
SoundHolo utilizes wave field synthesis (WFS), a spatial audio technique that creates point sources of sound in free space using an array of speakers [25]. This technique is based on Huygens' principle, which states that any wavefront can be generated from the superposition of multiple wavefronts [3]. By adjusting the amplitude and phase of each speaker, the generated wavefronts can collectively produce a sound source positioned in front of the array. What distinguishes WFS from other spatialization techniques, like ambisonics, is that the perceived location of the constructed sound source is the same as its physical position in the room. As a result, every listener will perceive the sound source coming from the same location, providing a naturalistic auditory experience [25].

Although WFS is relatively established within acoustics and experimental music, this technique has not yet seen widespread adoption in HCI research. Most research within acoustics focuses on spatial audio optimization through theoretical principles and technical development [5, 21]. In experimental music, composers have applied WFS to create spatialized concerts and immersive performances [9]. However, most acoustic and musical applications focus on non-interactive sonic experiences, where the location and motion of the sounds do not change in response to physical objects or people's actions in a space.

SoundHolo advances HCI research on interactive WFS. A few HCI examples utilize computer vision to track objects and augment them with 3D sounds using speaker arrays. For instance, in WFS-AR [20], WFS and augmented reality (AR) are combined to synthesize sounds onto a physical paddle tracked using visual markers. To track objects, this system requires video-see-through AR glasses. SoundHolo does not require users to wear personal devices. The BoomRoom integrates IR cameras with WFS to create a music-mixing experience [22]. Using gestures, a single user can "touch" virtual sounds by extracting them from bottles and placing them in stationary locations in mid-air.

The system features a 56-channel array with speakers spaced at 17 cm, arranged in a circular space of 3 m diameter. In contrast, SoundHolo explores the motion of multiple sound sources as they emerge from objects and fly around them inside large spaces. In the interactive installation GrainStick [16, 17], users move motion-tracked controllers to produce the sound of rolling grains between them. The installation explores a single sound point following a specific motion path. SoundHolo explores multiple sound configurations along various motion paths.

In sum, SoundHolo is the only interactive WFS application that investigates how multiple sounds can be attached to objects or moved around and between them in space. Our system also integrates motion tracking with a high-resolution array, using 64 speakers spaced at 59 mm within a large 20x30x12 feet room. To our knowledge, no other interactive WFS has been implemented with such high resolution in an environment of these dimensions. This high-resolution is expected to enhance sound localization and improve the overall interactive WFS experience. The room's size allows many users to interact with a WFS system from multiple vantage points and distances.

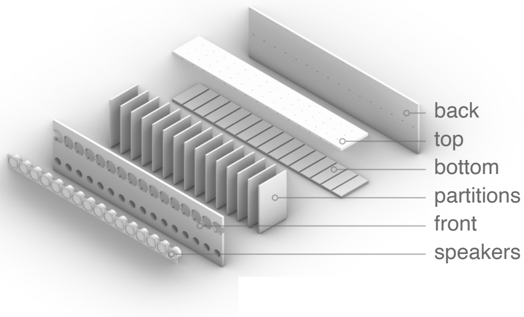


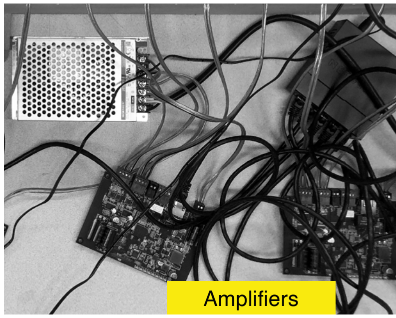
System Architecture

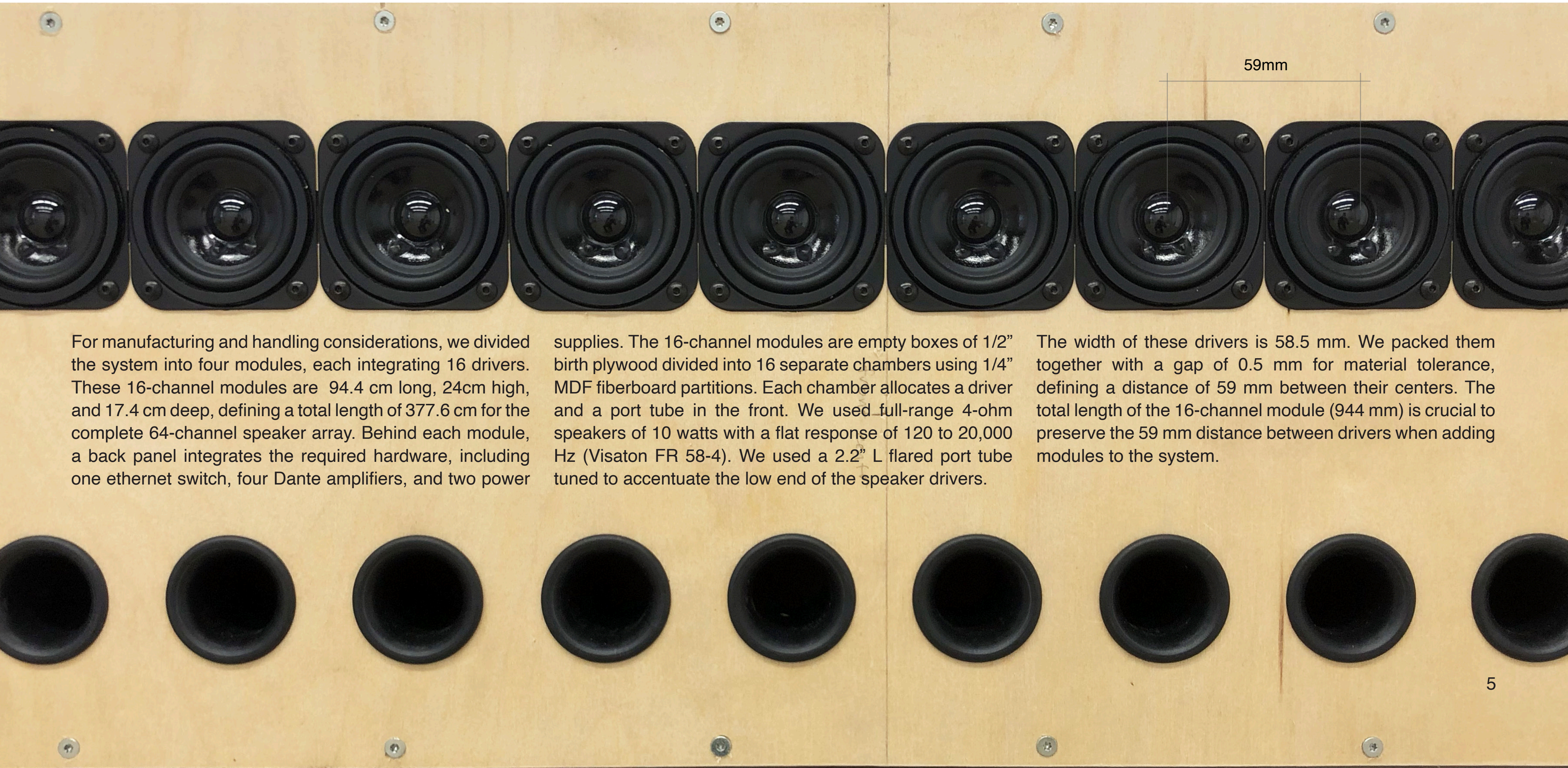
SoundHolo is installed in a 20x30x12 feet empty room. The system integrates a motion tracking system and a 64-channel speaker array. For motion tracking, a personal computer (PC) is connected to a managed ethernet switch installed on a ceiling rack. Ten infrared cameras (Optitrack PrimeX13) are installed in the room walls at 10 feet height. The switch connects the cameras to the PC while providing power over ethernet (PoE). For the 64-channel speaker array, a Mac computer is connected to a second managed ethernet switch installed behind the speaker array. This main switch is connected to four ethernet switches, each connected to four 4-channel Dante audio amplifier boards (Keiga ED-008). Finally, each amplifier connects to four speaker drivers, completing the 64-channel array. Four power supplies provide power for the 16 amplifiers and through them to the 64 drivers.

Physical Design

We designed and implemented a custom-made 64-channel speaker array for WFS. The amount, size, and arrangement between speakers are critical to generating an effective WFS effect. Based on the Huygens' principle, the more speakers in a WFS array and the closer they are together, the better 3D sound source reconstruction. Our system provides a high-resolution array with 64 speakers spaced 59 mm apart, generating sound sources between 250-2500 Hz without aliasing.







For manufacturing and handling considerations, we divided the system into four modules, each integrating 16 drivers. These 16-channel modules are 94.4 cm long, 24cm high, and 17.4 cm deep, defining a total length of 377.6 cm for the complete 64-channel speaker array. Behind each module, a back panel integrates the required hardware, including one ethernet switch, four Dante amplifiers, and two power supplies. The 16-channel modules are empty boxes of 1/2" birth plywood divided into 16 separate chambers using 1/4" MDF fiberboard partitions. Each chamber allocates a driver and a port tube in the front. We used full-range 4-ohm speakers of 10 watts with a flat response of 120 to 20,000 Hz (Visaton FR 58-4). We used a 2.2" L flared port tube tuned to accentuate the low end of the speaker drivers.

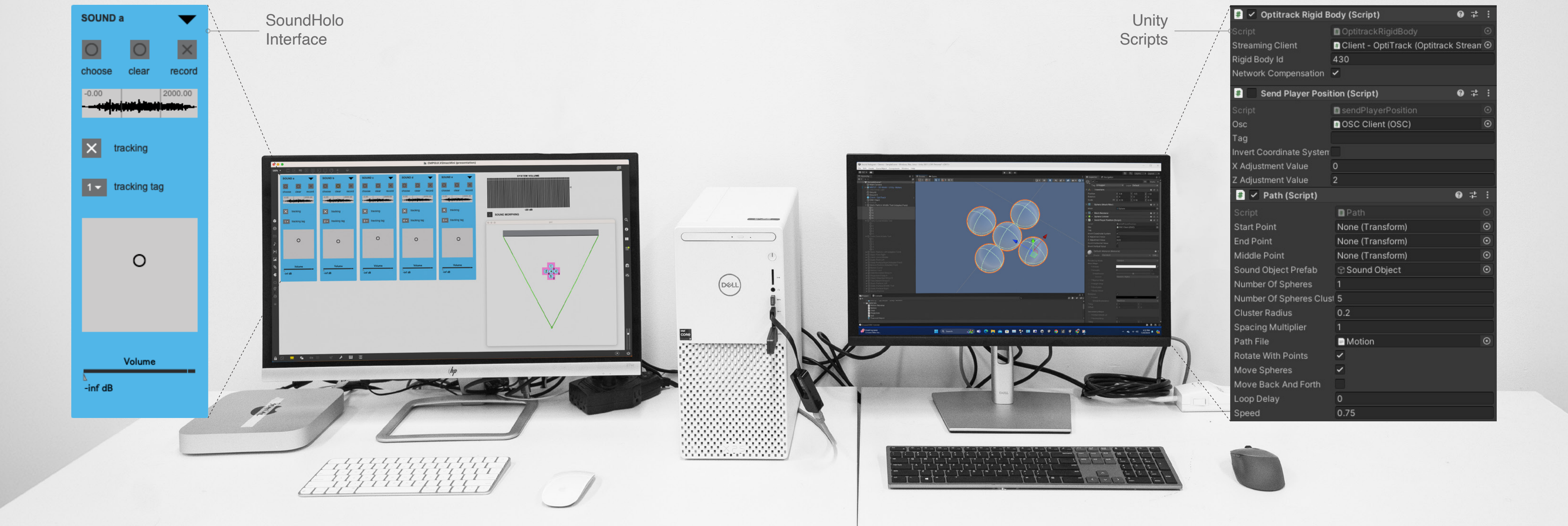
The width of these drivers is 58.5 mm. We packed them together with a gap of 0.5 mm for material tolerance, defining a distance of 59 mm between their centers. The total length of the 16-channel module (944 mm) is crucial to preserve the 59 mm distance between drivers when adding modules to the system.

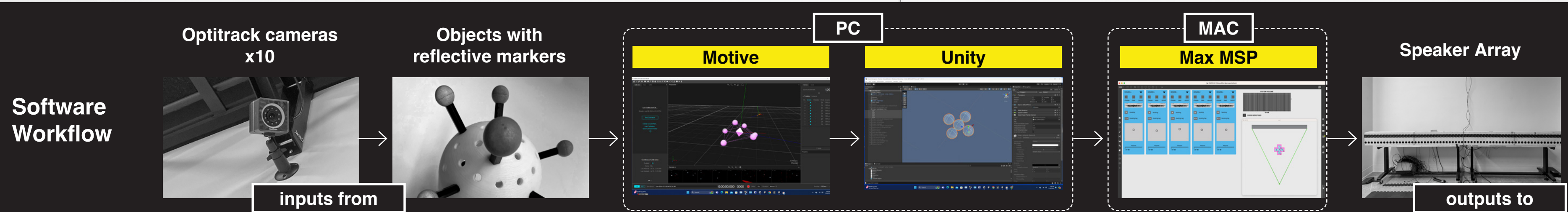
Software Implementation

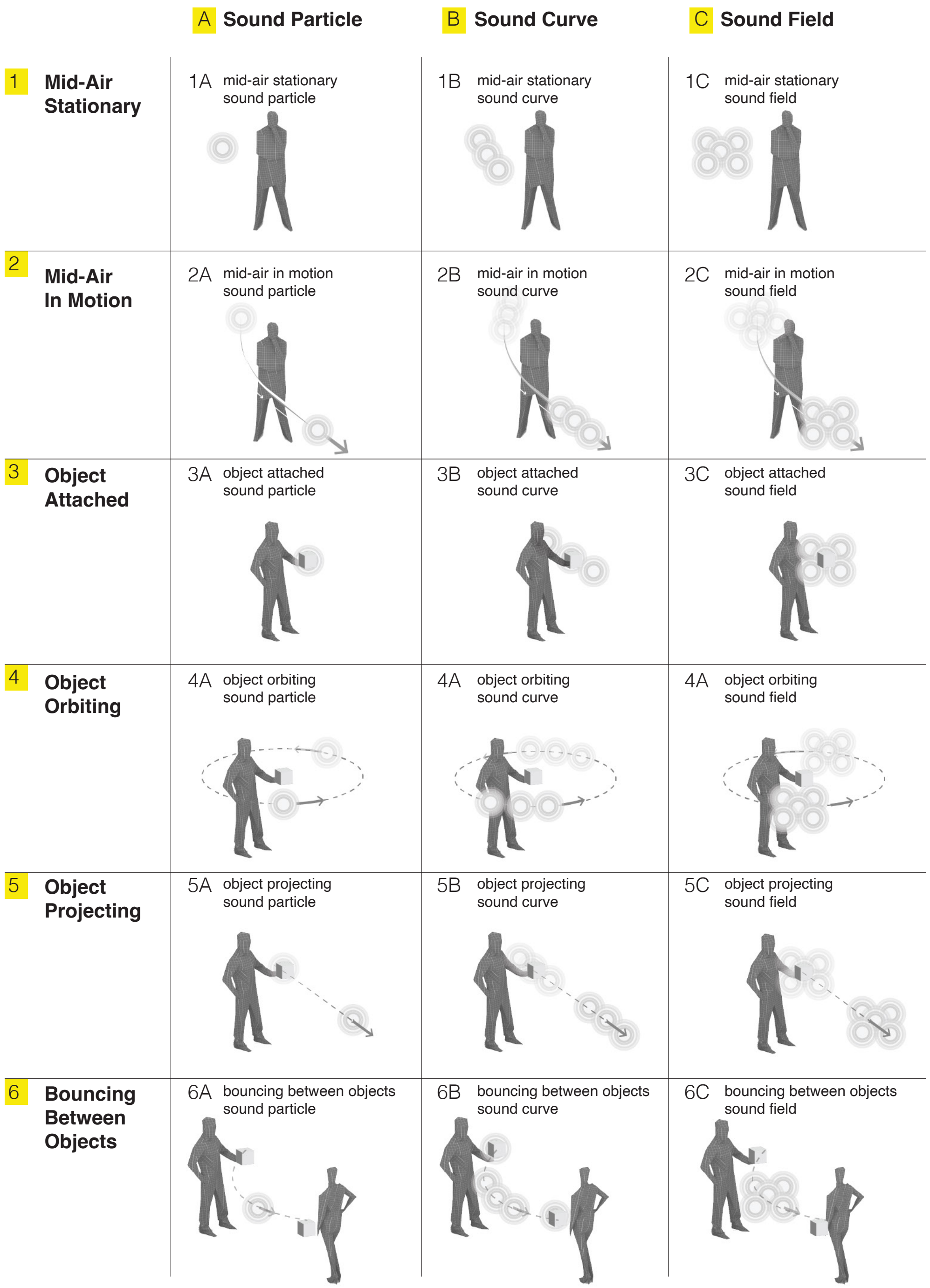
SoundHolo, uses a suite of software platforms for motion tracking and WFS control and integration. Specifically, Motive, Unity, and Max MSP/Jitter platforms are used to generate 3D sounds attached to physical objects and moved around them in mid-air. For SoundHolo to work, these three platforms need to work synchronously and be connected as part of the same network.

Motive is used to track physical objects as rigid bodies, assigning IDs and motion data in real-time. By pairing with infrared cameras, Motive captures the position of small retroreflective markers, delivering real-time motion data [32]. For SoundHolo, a set of reflective precision spheres with a radius of 6.4 mm to 12.7 mm is used. Unity receives the motion data from Motive and visualizes the physical objects in motion within a digital replica of the room in 3D. We have developed a set of Unity scripts that represent sound sources as digital objects for users to move freely–without reference to a physical source–or in response to tracked physical objects. The organization and position of these digital objects are then translated into a new set of motion data and associated IDs.

Max MSP is used as an audio processing platform. Position data from tracked objects is sent from Unity to Max through the Open Sound Control (OSC) protocol. That position data is then translated into a separate audio stream for each of the 64 speakers in the array in order to construct a sound source in the desired location. This translation is done by a WFS algorithm from Ircam's Spat5 external library [33]. Audio is sent from Max MSP to the speakers through Dante Virtual Soundcard. We created an interface for users to upload audio files, control local and global volume, and visualize source position through a Jitter render of the listening area.







SoundHolo Matrix

While implementing SoundHolo, we explored multiple sound configurations along various motion paths. We identified eighteen combinations, which we organized into the Soundholo Matrix. We propose this matrix as a palette of design options for HCI researchers looking to incorporate novel sonic augmentations into their interactive applications.

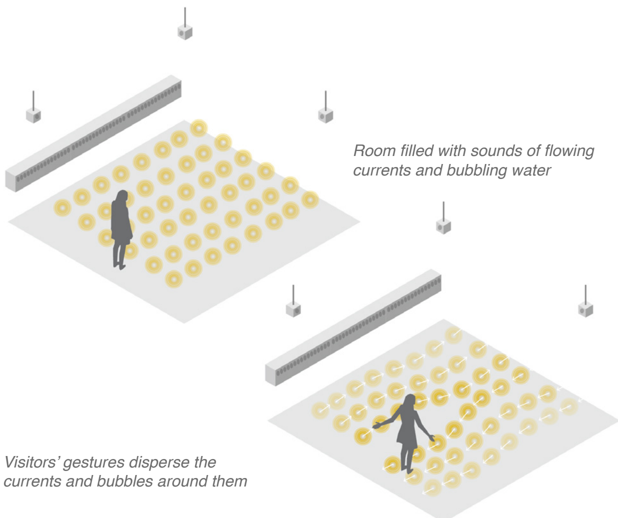
SoundHolo can generate multiple 3D sounds simultaneously to produce diverse sonic experiences. By dynamically choreographing the location of these 3D sounds, the platform can create a range of configurations, including (A) particles, (B) curves, and (C) fields. A sound particle corresponds to an individual 3D sound source. For example, a human voice or a bird tweet. A sound curve corresponds to a one-dimensional array of sound particles organized next to each other in a curved or straight line. For example, the sound of a waterfall or stream of water. A sound field corresponds to a two-dimensional array of sound particles grouped on a regular or irregular polygon. For example, the sound of rain on a surface or wind through trees. These three configurations are represented on the matrix's horizontal axis.

SoundHolo can also dynamically position these configurations–particles, curves, and fields– in space or co-located with physical objects. By anchoring a location in space, the system can create 3D sounds floating in mid-air without reference to a physical source. These sounds placed in the space can be (1) stationary or (2) in motion. Using motion tracking, the system can also generate 3D sounds moving in response to the location of physical objects. 3D sounds can be (3) attached to, (4) orbited around, and (5) projected from a physical object moving in space. For orbit and projection, the motion of the sounds can follow rectilinear or curvilinear paths. When tracking multiple physical objects, the 3D sound configurations can also (6) bounce between them, following a path defined by the objects' dynamic location. These six placements are represented on the matrix's vertical axis.

We used this matrix to ideate three user scenarios by combining three distinct sound configurations (horizontal axis) with three different motion paths (vertical axis). As a result, each scenario explores a unique sonic augmentation: a field moving in mid-air (2C), a particle attached to an object (3A), a curve bouncing between objects (6B).

Although all sound sources produced with our system occupy space and propagate sound waves in 3D dimensions, the location of these sources is limited to the plane defined by the height of the speaker array, i.e., if the speaker array is placed at 4 feet from the ground, the sound sources will be produced in a plane at that height.

Immersive Soundscapes

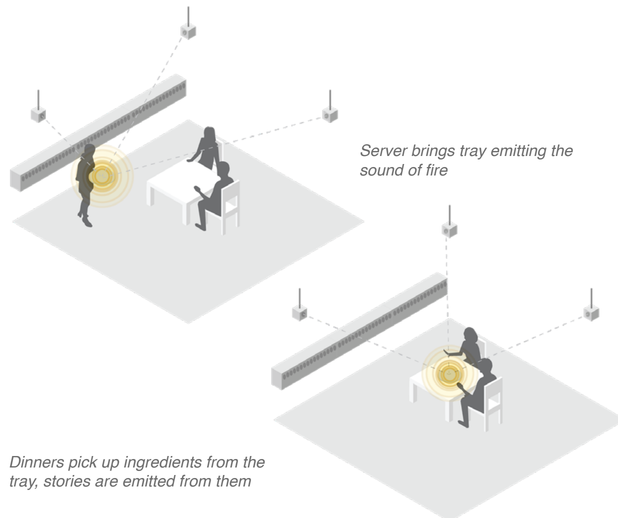


Sound fields are moved in mid-air, corresponding to 2C on the SoundHolo Matrix.

Soundscapes are experiences where multiple sounds are listened to simultaneously within an environment [7]. In immersive art, therapeutic, or branded experiences, SoundHolo can generate mid-air vibrations, creating fully immersive sonic experiences. By clustering sound sources together, SoundHolo can fill the space with vibration fields, allowing people to not only hear the sounds with their ears but also feel them with their entire bodies. At the Venice Biennale, for example, a pavilion could simulate the experience of underwater diving in an ocean territory. Visitors could enter a blue room where an underwater environment is simulated through 3D sounds arranged in dynamic fields that change their audio quality, shape, motion, and vibration in response to the visitor's full body movements.

A challenge in implementing this scenario is the potential obstruction of wavefronts when users are positioned between the listener and the speaker array, which could affect the accurate reconstruction of the sound fields.

Storytelling Merchandise

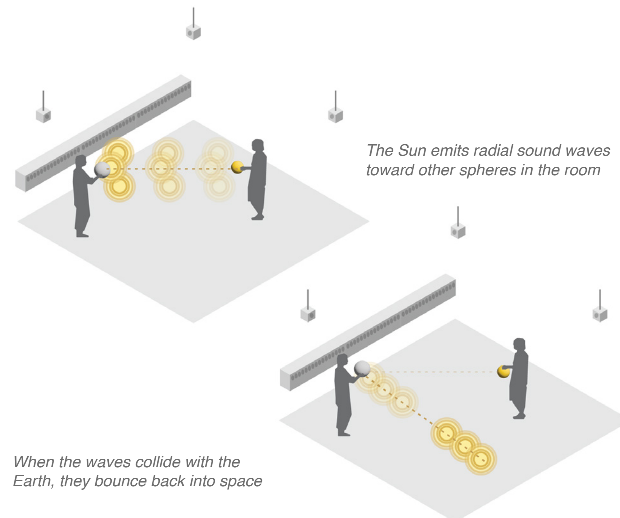


Sound particles are attached to physical objects, corresponding to 3A on the SoundHolo Matrix.

In tangible storytelling, physical objects become vessels for digital narratives [19]. In clothing stores, grocery stores, or bookstores, SoundHolo can transform any merchandise item–such as shoes, vegetables, or books– into a medium for tangible storytelling. By tracking these objects and dynamically attaching 3D sound to them, SoundHolo can provide stories and product information to consumers as they shop in physical retail spaces. For instance, diners in a restaurant could hear interactive narratives emerging from their meals as they are served, learning about the history and cultural origin of the dishes. Depending on how diners interact with their meals, they could trigger slightly different stories tailored to the unique background of the dish.

A challenge in implementing this scenario is the potential for camera occlusion when users' bodies partially obscure the tracked objects, which may impact the tracking system's performance.

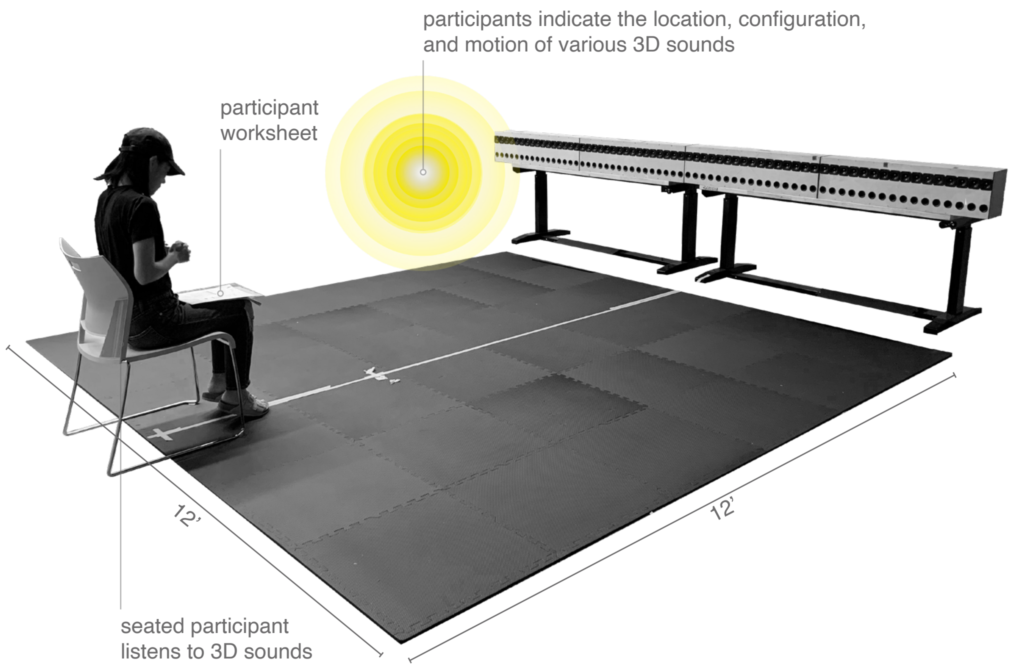
Spatial Auditory Displays

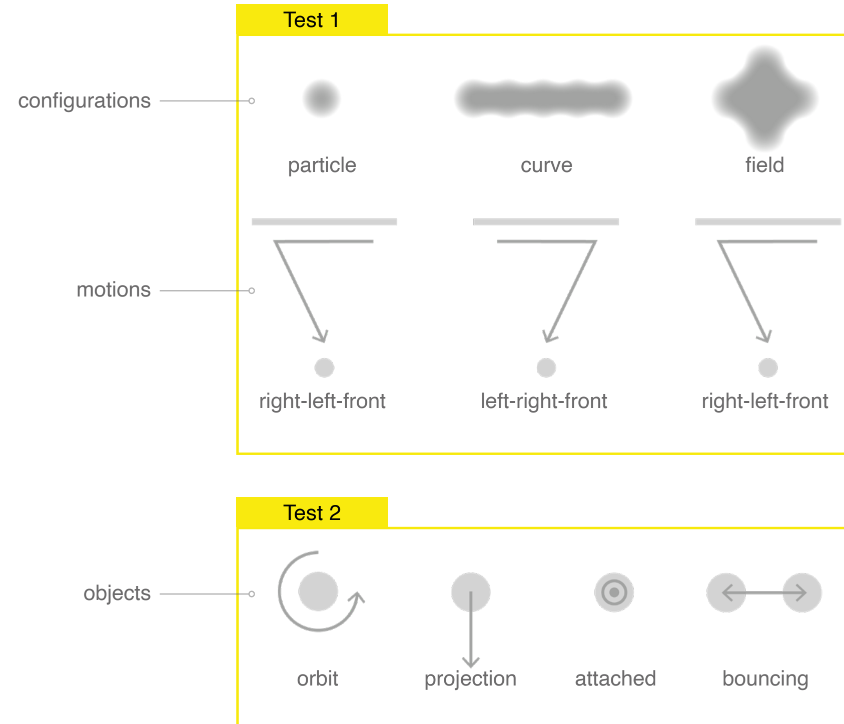


Sound curves are generated between multiple physical objects, corresponding to 6B on the SoundHolo Matrix.

Auditory displays use sound as a medium to represent data [8]. In classrooms, libraries, or museums, ShoudHolo can transform entire spaces into auditory displays, allowing sound to emanate from physical objects like pencils, books, or sculptures and even float in mid-air. By tracking these objects and dynamically generating sounds around and between them, SoundHolo can display information in three dimensions. In a science museum, for example, data about solar particle radiation could be represented spatially. By moving spheres representing astronomical objects, visitors could learn how the Sun emits radiation and how the planets use their magnetosphere to repel these particles.

A challenge in implementing this scenario is acoustic interference when multiple sounds cross paths. This issue can be mitigated by choreographing the movements of the sounds to prevent overlapping locations. The other two issues presented in these scenarios could be addressed by increasing the number of cameras and extending the speaker array to surround users across multiple walls.





Experimental Set Up

From a technical implementation, SoundHolo can produce eighteen output combinations, as presented through the matrix. To study how people perceive these variations, we conducted a preliminary system evaluation with 8 participants, including 5 females and 3 males between 20 and 40 years old. We tested participants in two controlled experimental conditions.

The first test investigates the perception of sound configurations–particle, curve, and field–floating in mid-air in two conditions: stationary and in motion. The second test investigates the perception of sound particles moving in reference to physical objects in four paths: attached to, orbited around, and projected from a physical object and bouncing between two objects.

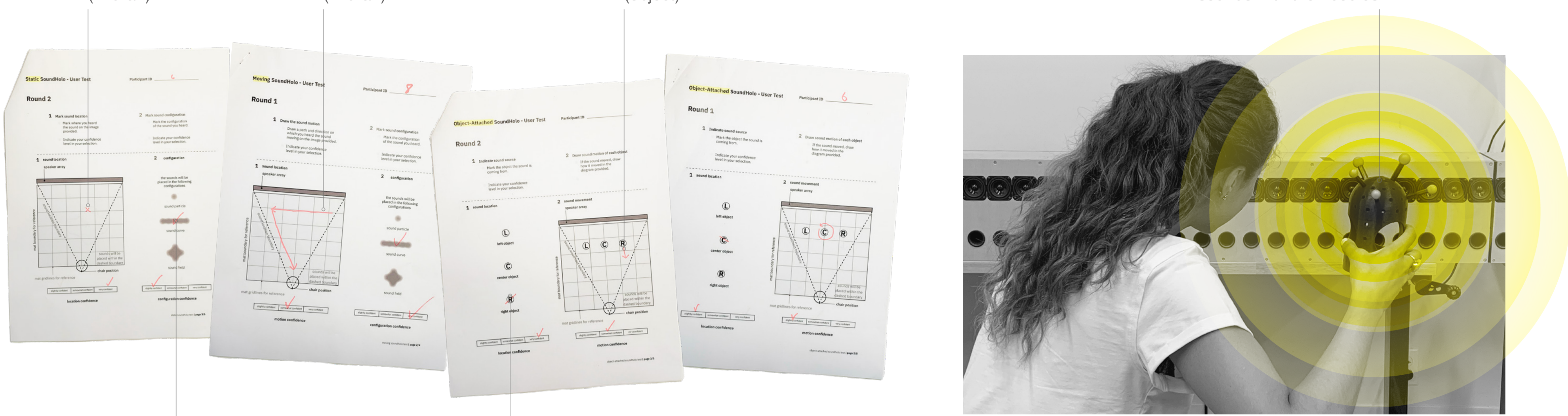
A modular floor outlined a 12x12 foot listening area, with a chair in the center of the room directly in front of the speaker array. In the first test, sound configurations were displayed from 5 stationary locations, followed by 3 motion paths. For the second test, three 2x3 inch objects were placed 3 feet in front of the speaker array, one to the left, one in the center, and one to the right. These objects were equipped with markers for motion tracking. The 3D sounds were presented in 4 conditions: orbiting around the central object, projecting from the right object, emanating from the left object, and bounding between the center object and right objects. In each test, participants experienced the 3D sounds by sitting in the chair (1-3 min) and then moving around the listening area (1-3 min). For the second test, participants also observed the researcher moving the objects (1-3 min) and then grasping them themselves (1-3 min). For all conditions, the various sonic combinations were generated using generic noise-based sound samples, selected to avoid any associations with preconceived auditory experiences.

After each test, the participants used a worksheet to record their responses visually. The worksheet included a floorplan of the listening area next to one multiple-choice question. In the first test, participants drew the location and motion path of the sounds and selected the corresponding configuration. In the second test, participants indicated the objects emitting the sounds and drew the associated motion path. For all responses, a 3-point scale was added to evaluate the confidence level of participants' responses. At the end of the experiment, we asked participants how they felt about the overall experience of listening to sounds in these two experimental conditions: floating in mid-air or moving in reference to physical objects.

Results and Feedback

To evaluate the results, we coded the participants' responses and scored them, assigning positive values for correct answers (+1) and negative for incorrect ones (-1). The values were weighted (1-3) based on the participants' confidence levels. The scores show higher identification results in the second test, in which the sound sources moved in reference to physical objects. Within this test, participants' scores for identifying the source object are higher than for the motion paths emanating from them. Although identifying the source objects is higher when the sound source was attached to an object, the motion path with the most correct answers corresponds to the bouncing condition, in which the sounds are moving between two objects. Projection and orbit motion paths provided similar positive identification results but were lower in comparison to attached and bouncing conditions. Within the first test, the scores are higher for stationary sounds compared to motion conditions. Within static conditions, identifying location scores higher than perceiving the sound configurations. Specifically, in static conditions, participants confused curve and field configurations. In motion conditions, identifying the motion shows higher scores than recognizing the sound configurations.

To analyze the participants' verbal answers, we compared their responses and selected those that were similar among two or more participants. Overall, their verbal feedback corroborated the worksheet results. They indicated that the second testing condition was more accessible: "*For me, the easiest one was to identify the sound source when the sound is coming from an object*" (P8), and "*I think it was more interesting when the sounds were connected to objects [...] it was easier and also more fun to figure out which sound it was and what the interaction was between objects*" (P6). They also confirmed that it was challenging to identify sound configurations and their motion paths: "*Figuring out the shape of the motion path and … at the same time the shape of the configuration was the most challenging part for me*" (P6), and "*I could pick up movement from right to left but when it was front to right I had no idea*" (P7). Participants also provided some unexpected insights about how they used their bodies to explore the sounds spatially: "*It was clear that the sound was going from left to right because I placed both of the objects on the side of my ears*" (P7), and "*it was much easier for me to identify the shape of the sound or the area where the sound was playing when I could like move in and out of the space. But when I was just like sitting in here, it was really hard*" (P5). Finally, participants provided positive feedback about the overall SoundHolo experience: "*We are not really trained to identify sounds, sound motions, and sound shapes… but that made this cool, I liked it. It's interesting*" (P4), and "*for the experience itself, it definitely was cool to see that I could hear sounds in different positions in space. Only that made it neat… so it was fun*" (P7).



Discussion

We developed SoundHolo to generate a range of eighteen sonic combinations. However, while all combinations can be technically generated, those that augment physical objects and their surrounding space were perceived more effectively. Although the scores for identifying sound configurations in mid-air were lower than expected, better results may be achieved when these configurations are connected to physical sources.

Related work has attached sound sources to physical objects by embedding electronics within their material surfaces [23]. SoundHolo's key advantage lies in its ability to dynamically move sound sources both in and out of objects and even between multiple objects. However, for applications where users need to listen through direct contact between their ears and the objects, embedding electronics within the materials may provide a more effective solution.

Additionally, we only investigated the SoundHolo Matrix using noise-based samples for all combinations. Exploring different types of sounds tailored to each combination could enhance the overall results. For example, the sound of a rolling ball might be better perceived in motion, while an alarm clock sound could be more distinct in a stationary position. Similarly, a bird tweet might be better perceived as a particle, a water stream as a curve, and rain as a field.

Related work in WFS for acoustic and experimental music applications has explored synthesizing sounds within non-interactive sonic experiences [9]. While SoundHolo's key advantage is creating interactive WFS experiences, the sounds tested are not currently dynamically synthesized. As a result, while the sound location can change, the sound properties remain constant. Incorporating real-time sound synthesis could further enhance SoundHolo's capabilities.

Lastly, the fact that participants used their bodies to perceive the 3D sounds suggests alternative ways to study the sound configurations produced by the system. Rather than focusing solely on identifying the "shape" of the configurations, it might be more appropriate to explore the experience of being inside and moving in and out of the spaces defined by those shapes. For example, even if the shape of the sound field is not easily recognizable, the experience of being surrounded by sound may be perceived more clearly. Entering a field where sound is all around you might feel different from merely crossing a line of sound.

Compared to other interactive WFS applications [16,17, 22], SoundHolo's primary advantage is its ability to create a variety of sound configurations–particles, curves, and fields. However, while SoundHolo mainly focuses on simple interactions, such as moving objects, related work has investigated gesture-based WFS control for real-time manipulation of sound locations and audio parameters [22].

Limitations and Future Work

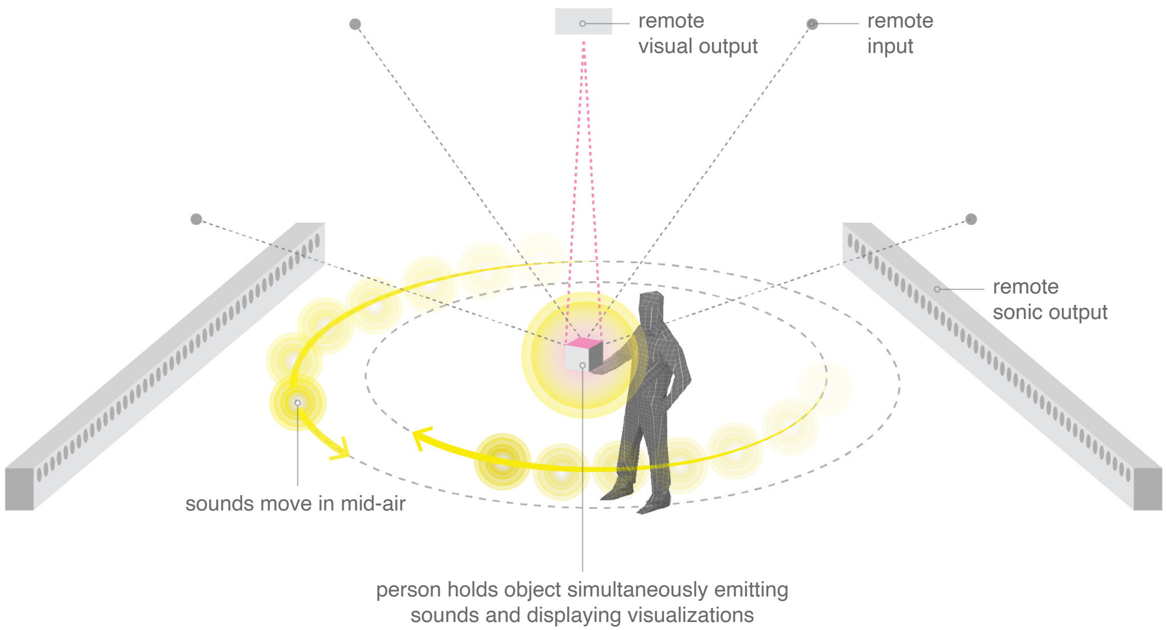
The limitations to our current exploration open opportunities for future work. First, there are opportunities to further improve the WFS experience. One challenge with a linear array is that the sound perception is optimal when facing the array. As users turn around, the spatialization effect may diminish. To address this issue, we are integrating four new modules into a 128-channel array, divided into two sections positioned on opposing walls of the room. Furthermore, small discrepancies in speaker placement and minor manufacturing imperfections might affect the overall WFS performance. Additionally, the room's high ceiling and hard surfaces contribute to reverberation, impacting sound localization. To mitigate these issues, we are implementing an algorithm to calibrate the drivers, compensating for both driver imperfections and room acoustics. We also plan to install absorber panels, curtains, and carpets to decrease the level of reverberated sound.

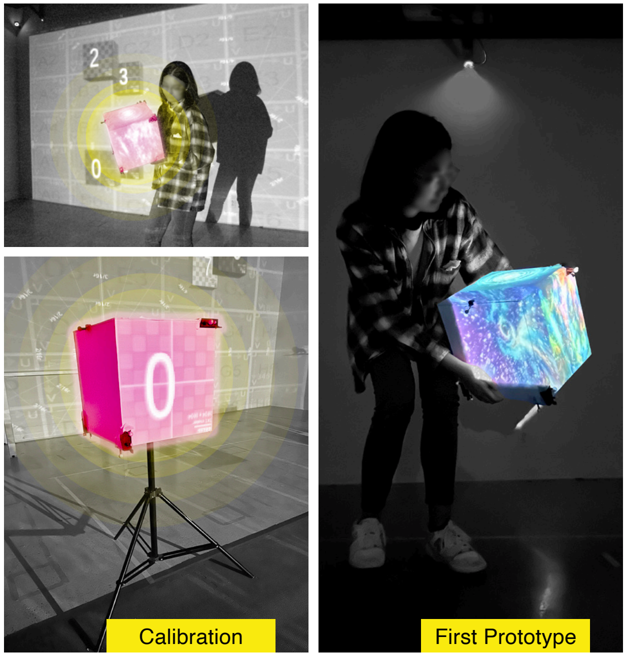
Second, another area for improvement is making the motion tracking less intrusive. Our current setup relies on physical markers to track objects, which can interfere with the seamless experience of interacting with everyday items in their natural, untouched state. We are exploring two alternative approaches to address this issue. One option involves using invisible infrared paint to create fiducials that can be tracked by infrared cameras [28]. The second approach utilizes digital cameras to track objects without the need for markers or fiducials. This method captures numerous photographs and trains a machine-learning system to recognize and track objects based on their unique shapes. This technique shows great promise, particularly for everyday objects with distinct geometries that facilitate object recognition and motion tracking.

Third, the SoundHolo matrix could be expanded by exploring additional combinations. Using our current implementation, we have only explored eighteen sonic combinations. However, more combinations can be investigated by varying the number, configuration, and movement of the sound sources. Incorporating animated properties into the sound configurations, for example, could lead to new hybrid combinations, such as transforming a curve into a field or collapsing that field into a point. Additionally, the three proposed user scenarios derived from the matrix could be refined by developing more concrete commercially viable products/services tailored to specific real-world users.

Fourth, there are opportunities to enhance the perception of SoundHolo's experience by integrating visual and tactile augmentations. Our preliminary evaluation suggests that the presence of physical objects may improve the identification of sound configurations, as well as their location and motion in space. Our working assumption is that projecting digital visualizations onto the floor or physical objects will further enhance this identification. For example, using dynamic projection mapping, cues and animations could be added to the physical objects producing the sounds [1, 18]. We are enhancing SoundHolo with visual augmentations by integrating our speaker array and tracking system with multiple digital projectors. By leveraging our existing tracking system, we can dynamically adjust the digital content of the projectors to map precisely onto the surface geometry of tracked objects. Similarly, tactile augmentations could further expand the SoundHolo experience. We have found that small physical participles, such as sand, can vibrate in response to the 3D sounds produced by our speaker array. Although this early discovery has not yet been fully investigated, it suggests exciting opportunities for future research into haptic feedback actuated remotely using sonic waves.

Future SoundHolo Set-up:
Surrounding speaker arrays + visual augmentations





Finally, our assessment of the SoundHolo experience has been preliminary. Further testing is necessary to comprehensively evaluate the system's capabilities and effectiveness. Including a more diverse demographic in the user study could also enhance the generalizability of our initial findings. Future research could focus on examining the ventriloquist effect in motion and investigating how multiple listeners experience the system simultaneously.

Conclusions

This paper introduces SoundHolo, a room-scale platform designed to sonically augment everyday objects and the space around them. The goal of SoundHolo is to provide new sonic augmentations that will inspire future interactive explorations in HCI. To achieve this goal we developed a working implementation integrating a custom-made high-resolution 64-channel array with a motion tracking system. We explained SoundHolo's system architecture, physical design, and software implementation. We presented 18 sonic augmentations through a matrix that showcases SoundHolo's capabilities visually. We proposed real-world applications for the system, including immersive soundscapes, storytelling merchandise, and spatial auditory displays. Finally, we discussed the results of a preliminary evaluation, which confirms the system's potential to sonically augment both physical objects and the spaces around them.

Acknowledgements

We thank the students, staff, and faculty of the Carnegie Mellon University (CMU) School of Design (SoD) for their support in the development of SoundHolo as a creative technology platform within the Spatial Experiences Lab (SxL). This research was partially funded by the Frank-Ratchye Further Fund (FRFF) from CMU Studio for Creative Inquiry and by the Fund for Research and Creativity (FRC) from CMU College of Fine Arts (CFA).

[1] Eric Akaoka, Tim Ginn, and Roel Vertegaal. 2010. DisplayObjects: prototyping functional physical interfaces on 3d styrofoam, paper or cardboard models. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction (TEI '10)*, January 24, 2010. Association for Computing Machinery, New York, NY, USA, 49–56. https://doi.org/10.1145/1709886.1709897

[2] Fouad Alallah, Ali Neshati, Yumiko Sakamoto, Khalad Hasan, Edward Lank, Andrea Bunt, and Pourang Irani. 2018. Performer vs. observer: whose comfort level should we consider when examining the social acceptability of input modalities for head-worn display? In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*, November 28, 2018. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3281505.3281541

[3] Forrest L. Anderson. 2021. Huygens' Principle geometric derivation and elimination of the wake and backward wave. *Scientific Reports* 11, 1 (October 2021), 20257. https://doi.org/10.1038/s41598-021-99049-7

[4] Benjamin B. Bederson. 1995. Audio augmented reality: a prototype automated tour guide. In *Conference Companion on Human Factors in Computing Systems (CHI '95)*, May 07, 1995. Association for Computing Machinery, New York, NY, USA, 210–211. https://doi.org/10.1145/223355.223526

[5] A. Berkhout, Diemer de Vries, and P. Vogel. 1993. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America* 93, (May 1993), 2764–2778. https://doi.org/10.1121/1.405852

[6] Paul Dourish. 2004. *Where the Action Is: The Foundations of Embodied Interaction* (New Ed edition ed.). MIT Press, Cambridge, MA.

[7] Karmen Franinovic and Stefania Serafin (Eds.). 2013. *Sonic Interaction Design* (Illustrated edition ed.). The MIT Press, Cambridge, MA.

[8] William W. Gaver. 1986. Auditory icons: using sound in computer interfaces. *Human-Computer Interaction* 2, 2 (June 1986), 167–177. https://doi.org/10.1207/s15327051hci0202_3

[9] Johannes Goebel. 2019. The EMPAC high-resolution modular loudspeaker array for wave field synthesis. In *Proceedings of 23rd International Congress on Acoustics (ICA' 19)*, September 13, 2019.

[10] Daniel Groeger and Jürgen Steimle. 2018. ObjectSkin: Augmenting Everyday Objects with Hydroprinted Touch Sensors and Displays. *Proceedings the ACM on Interactive, Mobile, Wearable and Ubiquitous Technololies* 1, 4 (January 2018), 134:1-134:23. https://doi.org/10.1145/3161165

[11] Yasha Iravantchi, Mayank Goel, and Chris Harrison. 2020. Digital Ventriloquism: Giving Voice to Everyday Objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, April 23, 2020. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3313831.3376503

[12] Hiroshi Ishii. 2008. Tangible bits: beyond pixels. In *Proceedings of the 2nd international conference on Tangible and embedded interaction (TEI '08)*, February 18, 2008. Association for Computing Machinery, New York, NY, USA, xv–xxv. https://doi.org/10.1145/1347390.1347392

[13] Hiroshi Ishii, Ali Mazalek, and Jay Lee. 2001. Bottles as a minimal interface to access digital information. In *Proceedings of the Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*, March 31, 2001. Association for Computing Machinery, New York, NY, USA, 187–188. https://doi.org/10.1145/634067.634180

[14] Robert J.K. Jacob, Audrey Girouard, Leanne M. Hirshfield, Michael S. Horn, Orit Shaer, Erin Treacy Solovey, and Jamie Zigelbaum. 2008. Reality-based interaction: a framework for post-WIMP interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, April 06, 2008. Association for Computing Machinery, New York, NY, USA, 201–210. https://doi.org/10.1145/1357054.1357089

[15] Marion Koelle, Matthias Kranz, and Andreas Möller. 2015. Don't look at me that way! Understanding User Attitudes Towards Data Glasses Usage. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*, August 24, 2015. Association for Computing Machinery, New York, NY, USA, 362–372. https://doi.org/10.1145/2785830.2785842

[16] Grace Leslie, Diemo Schwarz, Olivier Warusfel, Frédéric Bevilacqua, and Pierre Jodlowski. 2009. Wavefield Synthesis for Interactive Sound Installations. In *Proceedings of 127th Convention of Audio Engineering Society*, October 01, 2009. Audio Engineering Society. New York, NY.

[17] Grace Leslie, Bruno Zamborlin, Pierre Jodlowski, and Norbert Schnell. 2010. Grainstick: A Collaborative, Interactive Sound Installation. In *Proceedings International Computing Music Conference (*ICMC *'10)*, June, 2010. New York, NY.

[18] Kosuke Maeda and Hideki Koike. 2020. MirAIProjection: Real-time Projection onto High-speed Objects by Predicting Their 3D Position and Pose using DNNs. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '20)*, October 02, 2020. Association for Computing Machinery, New York, NY, USA, 1–5. https://doi.org/10.1145/3399715.3399839

[19] Ali Mazalek, Alison Wood, and Hiroshi Ishii. 2001. genieBottles: An Interactive Narrative in Bottles. In *Abstracts and Applications of the Conference on Computer Graphics and Interaction Techniques (SIGGRAPH' 01)*. Association for Computing Machinery, New York, NY, USA, 189-190.

[20] Frank Melchior, Tobias Laubach, and Diemer de Vries. 2005. Authoring and user interaction for the production of wave field synthesis content in an augmented reality system. In *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '05)*, October, 2005. 48–51. https://doi.org/10.1109/ISMAR.2005.20

[21] Frank Melchior, Christoph Sladeczek, Diemer Vries, and B. Fröhlich. 2008. User-dependent optimization of wave field synthesis reproduction for directive sound fields. In *124th Convention of the Audio Engineering Society (AES '08)*, May, 2008. Audio Engineering Society, Amsterdam, Netherlands, 601–612.

[22] Jörg Müller, Matthias Geier, Christina Dicke, and Sascha Spors. 2014. The boomRoom: mid-air direct interaction with virtual sound sources. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, April 26, 2014. Association for Computing Machinery, New York, NY, USA, 247–256. https://doi.org/10.1145/2556288.2557000

[23] Edwina Portocarrero, Gershon Dublon, Joseph Paradiso, and V. Michael Bove. 2015. ListenTree: Audio-Haptic Display In The Natural Environment. In *Proceedings of the Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)*, April 18, 2015. Association for Computing Machinery, New York, NY, USA, 395–398. https://doi.org/10.1145/2702613.2725437

[24] Brygg Ullmer, Orit Shaer, Ali Mazalek, and Caroline Hummels. 2022. *Weaving Fire into Form: Aspirations for Tangible and Embodied Interaction* (1st ed.). Association for Computing Machinery, New York, NY, USA.

[25] Diemer de Vries. 2009. *Wave Field Synthesis*. Audio Engineering Society. AES Monographs. New York, NY.

[26] Mark Weiser. 1999. The computer for the 21st century. *SIGMOBILE Mobile Computing and Communications Review* 3, 3 (July 1999), 3–11. https://doi.org/10.1145/329124.329126

[27] Mark Weiser and John Seely Brown. 1997. The coming age of calm technology. In *Beyond calculation: the next fifty years*. Copernicus, USA, 75–85.

[28] Karl D.D. Willis, Ivan Poupyrev, Scott E. Hudson, and Moshe Mahler. 2011. SideBySide: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology (UIST '11)*, October 16, 2011. Association for Computing Machinery, New York, NY, USA, 431–440. https://doi.org/10.1145/2047196.2047254

[29] Jing Yang, Yves Frank, and Gábor Sörös. 2019. Hearing Is Believing: Synthesizing Spatial Audio from Everyday Objects to Users. In *Proceedings of the 10th Augmented Human International Conference 2019 (AH2019)*, March 11, 2019. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3311823.3311872

[30] Jing Yang and Friedemann Mattern. 2019. Audio augmented reality for human-object interactions. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*, September 09, 2019. Association for Computing Machinery, New York, NY, USA, 408–412. https://doi.org/10.1145/3341162.3349302

[31] Junyi Zhu, Lotta-Gili Blumberg, Yunyi Zhu, Martin Nisser, Ethan Levi Carlson, Xin Wen, Kevin Shum, Jessica Ayeley Quaye, and Stefanie Mueller. 2020. CurveBoards: Integrating Breadboards into Physical Objects to Prototype Function in the Context of Form. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*, April 23, 2020. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376617

[32] Motive - In Depth. OptiTrack. Retrieved August 1, 2024 from http://optitrack.com/software/motive/index.html

[33] Spat | Ircam Forum. Retrieved August 1, 2024 from https://forum.ircam.fr/projects/detail/spat/