Spatially Immersive Sound in the Bird Song Diamond Project

March 2019

BRUMLEY John

Spatially Immersive Sound in the Bird Song Diamond Project

School of Integrative and Global Majors
Ph.D. Program in Empowerment Informatics
University of Tsukuba

March 2019

BRUMLEY John

I, John Brumley confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

The aim of this thesis is to consider how the manifold research and material outcomes associated with the Bird Song Diamond project have repercussions for the field of Human Informatics. Specifically, how the development of a spatially immersive sound system for simulating the complexities of human-environmental interaction has unlocked new applications for projecting sound into space; in situations both artistic as well as practical. Given the potential for these novel applications for immersive and localized sound, this thesis addresses the research question of how spatial, immersive sound can produce empowerment for individuals and groups. Empowerment, specifically, within the context of the Bird Song Diamond Project for artists, researchers, and the public, and generally, as an effective system for compact, relatively low-cost, spatialized sound.

Beginning with an introduction to the Bird Song Diamond project, including the primary aim of the bird song research and all affiliated research fields, and covering the entire history of the Bird Song Diamond installation and how it has evolved over the course of each subsequent exhibition. The importance of the Bird Song Diamond project for human informatics will be explained by considering aspects of the Bird Song Diamond project can inform and supplement the field. Briefly, rather than adopt a human-centric position for considering human activity and well-being, the Bird Song Diamond project, and the resulting spatial immersive sound system, demands a holistic awareness of interspecies and environmental relationships in order to ultimately improve the human subject. Additionally, the continuation of aggressively cross-disciplinary collaboration, which is one of the defining features of the project, is an important and difficult to maintain strategy,

and such an expanded view of this activity should be a continuous goal of Human Informatics.

After focusing heavily for the introduction on the Bird Song Diamond project, is a change in topic to clarify for the reader, the notions of immersion, immersive sound, spatial sound, and spatially immersive sound. Included in this section are works that relate to, or involve aspects of, spatially immersive sound. A distinction is made between the understanding of spatial immersion throughout history, musical and acoustic techniques for spatial immersion, and contemporary technologies used to produce such effects. From ancient architectural acoustic tricks, to canonized orchestral and theatrical methods of immersion, to self-driven perceptual reorientation, a very broad base of material on sound, immersion, and spatialization will be touched upon prior to arriving at spatially immersive sound. The importance of spatially immersive sound in the context of the Bird Song Diamond project's installation and conceptual framework is then explained.

Once the main concepts and tools involving spatially immersive sound are introduced, a series of related projects, which precede the author's own spatially immersive sound system, are discussed. Each project which is covered establishes a different aspect of spatial immersion, yet by themselves, the projects do not meet the full requirements of a spatially immersive system. For example, all make use of a moving ultrasonic speaker to achieve dynamic positioning of a localized sound source, but have not developed the projects to facilitate immersion (or are only capable of a limited immersive experience). These shortcomings lead directly into the next section in which the author's system is presented.

A thorough account of the system used in the Bird Song Diamond project gives the reader a full understanding of the underlying technical aspects of the project, which are more apparent after having established spatially immersive sound in the previous section. This includes details regarding early versions, development, improvement, fabrication, and calibration. Comparisons are made between two major shifts in system architecture, which corresponds to an older wired system and a newer wireless one. Hardware tests to determine the most stable microcontroller firmware are recounted, as well as other short experiments to measure the system's efficacy outside of human perception.

But of course, humans literally come first in Human Informatics, so two experiments, involving live, human participants, are presented. Each experiment aims to measure a particular aspect of the spatially immersive sound system, with the first focused on perceived direction of moving sounds, and the second on spatial separation of bimodal sources. It is shown by the first experiment that participants can successfully estimate sound direction when given a set a possible directions, though as movement complexity increases, the overall predictive ability decreases. The bimodal experiment is less conclusive, showing that for some participants, in some situations, an estimation of the point of subjective equality can be made, however it requires further experiments to eliminate problematic environmental errors. Finally, installations of the system in two international exhibitions of the Bird Song Diamond project, the LargeSpace in Tsukuba and the CCD Memorial in Mexico, provide an opportunity for comparing a fixed and moving ultrasonic speaker systems. While attempts using video tracking and behavior annotation were made to find comparable data, the differences between the two installations, and multitude of other simultaneous media in bird song diamond, was too great for the author to measure the impact of sound in isolation. Fortunately for the exhibition in Mexico, user surveys and interviews were conducted, with questions aimed to gain insight into how visitors felt about the sound. The responses confirmed the importance of sound for the Bird Song Diamond Project and that people were quite aware of the dynamically changing locations of sound.

Based on the user study and participant interviews, the spatially immersive sound system has been shown to empower users through stronger engagement with complex networks of bird song and movement at the heart of the Bird Song Diamond project. Additional applications for spatially immersive sound outside of the Bird Song Diamond project can provide future research possibilities. Two additional projects, using smart phones for the creation of spatially immersive sound, have been developed. While not yet fully implemented, these two projects constitute future work.

Finally, only through the collaboration of artists, scientists, and engineers with the aim of co-developing interconnected systems and artworks has this system emerged, and the significance for inter-disciplinary, inter-field, inter-cultural partnerships cannot be underestimated.

Acknowledgements

I would like to take a moment to express my gratitude to those who's support and patience have allowed this thesis to materialize. For those involved in the Birds and Bird Song Diamond Project, it is the primary source of material and impetus for the majority of the research contained herein. The connections established between Hiroo Iwata and Victoria Vesna to bring BSD to Japan were essential for the further development of the project, and,ultimately, my own journey to Tsukuba. Charles Taylor's wisdom and guidance during the time spent in his laboratory greatly prepared me in forms of methodology and protocol that I would face during my current research path. All three have continued to give support throughout my studies in Empowerment Informatics, aiding with journal papers, conference talks, and other lessons along the way. Additionally, Professor Hiroaki Yano and Professor Kazuya Inoue have patiently listened to my questions, no matter how incoherent, and provided valuable answers. Visiting speakers, especially, Erkki Huhtamo and Eddo Stern, offered wonderful advice on bridging artistic and engineering practices. Living abroad for years at a time was not an easy experience, but working alongside a cohort of helpful and friendly people helped a lot. Aki Yamada has been a lifeline here, willing to provide so much assistance and always available to give advice, listen, or share a snack. Maša, Aisen, Takeshi, and especially Alberto have all had to endure sharing a space with me, and, due to their good-nature, lots of supportive, thought provoking, and humorous conversations took place.

Table of Contents

List of figures

List of tables

Abbreviations

API	**A**pplication **P**rogramming **I**nterface
SIS	**S**patially **I**mmersive **S**ound
HI	**H**uman **I**nformatics
BSD	**B**ird **S**ong **D**iamond
ALife	**A**rtificial life
MUS	**M**oving **U**ltrasonic **S**peaker
UDP	**U**ser **D**atagram **P**rotocol
GPS	**G**lobal **P**ositioning **S**ystem
PWM	**P**ulse **W**idth **M**odulation
PSE	**P**oint of **S**ubjective **E**quality
JSON	**J**avaScript **O**bject **N**otation

Chapter 1

Introduction

> "If the Soul be consider'd as it enjoyes a life more noble then the natural, namely, that of Grace; the Sense of Hearing seemes the Author of this Life." - III. Conference XXIV.I. Which of the five Senses is most Noble (Renaudot et al. 1664)

The medium and perceptive capacity of sound is often overshadowed by its visual cousin. Historically, the dominance of vision has led to a hierarchical categorization in philosophy, metaphysics, and aesthetics, with audition generally deemed secondary to that of vision. (Aristotle. & Ross 1981; Plato. & Allen 2008; Galilei & Drake 1990) More recently, psychophysical experiments have confirmed a visual dominance at play in bimodal (hearing and vision) tasks (Colavita 1974). Perhaps due to this philosophical, and ultimately physical, bias for the visual, the creative potential for audio information transmission has not received the attention paid to visual modalities.

In the field of immersion, and consequently the fields of virtual and augmented reality, creating a convincing simulation or simulacrum requires additional attention paid to these non-visual modalities. The bouncing of a ball in a room will feel more grounded in reality when it also produces sound, even better if the resultant sound is filtered and modulated by environmental factors; the dusty parquet floor of the empty, unheated office produced a hollow thud that echoed off the long glass windows with each bounce. Just as a convincing work of fiction constructs a reality through details, well-done and carefully considered sound design provides sensory details that can drastically improve the effectiveness of an immersive experience.

Advances in sound producing technologies have spawned alternative approaches for immersive sound. Far in the past, sound localization required a cadre of musicians, while more recently, it required an improbable number of speakers (Desantos et al. 1997) or a fixed listener in an anechoic chamber *and* an improbable number of speakers (J. Berkhout et al. 1993; A Gerzon 1973). With the emergence of parametric directional speakers (Pompei 2019) sounds can be beamed to particular locations, silent until arriving at the destination and able to reflect off surfaces. Parametric speakers, due to the unfamiliar ways in which sounds produced by these speakers behave, offer a radically different method for the creation of localized sound. Rather than offer a replacement for existing sound localization techniques, parametric speakers inhabit a specific use-case depending on the applications.

Artworks, in particular, in their attempt to evoke or explore new modes of experience, consciousness, and embodiment, are naturally open to technologies that offer forms of novel perception. Specific to and intertwined with the work done in this thesis, the enhancement of parametric speakers for more flexible applications of sound localization, is the Bird Song Diamond Project.

The aim of this thesis is to consider how the manifold research and material outcomes associated with the Bird Song Diamond project have repercussions for the field of Human Informatics. Specifically, how the development of a spatially immersive sound system for simulating the complexities of human-environmental interactions has unlocked new applications for sound projection in space; in artistic as well as practical situations. Given these novel applications, this thesis addresses the research question of how spatial, immersive sound can produce empowerment for individuals and groups.

1.1 The Bird Song Diamond Project

The Bird Song Diamond (BSD) project is a series of multifaceted and multidisciplinary installations with the aim of bringing contemporary research on bird communication to a large public audience. As will be seen in the following section, even an isolated point in the constellation of research underlying BSD is esoteric enough that a broader public would not readily engage with the material. Combining the various combinations of research into a comprehensive and engrossing, yet digestible, outcome evades most forms of media approaches. With this in mind, the approach of developing an experiential installation regarding the behavioral aspects of birds while also directly involving members of the multidisciplinary research team became an elegant solution to a complex problem. Using art and technology to create immersive experiences, BSD allows large audiences to embody bird communication rather than passively observe. Active, embodied participation with the subject matter proved to be an efficient conduit of knowledge. What began as a mandatory provision from the grant-giving organization, the National Science Foundation, became successful enough to merit a renewal of the funding in order to continue research on bird song.(Charles Taylor et al. 2011)

1.1.1 Goals

Bird communication extends beyond the handful of functional categories we generally attribute to them: mating, territory, warnings. Recent techniques in collection and analysis have allowed us to dissect bird songs to reveal grammar and syntactic structures that were previously unobservable. Bird Song Diamond (BSD) is an ongoing project that seeks to disseminate this scientific research on birdsong (Arriaga et al. 2014), as well as emerging techniques in collection using sensory networks (Yu et al. 2016) and analysis based on artificial life simulations (Ikegami et al. 2017). The project has developed over the past five years as a public outreach extension of a larger research initiative entitled “Acoustic Sensor Arrays for Understanding Bird Communication.” Our solution to reaching large audiences while providing active involvement with the subject matter has been the development of immersive and interactive art installations. Employing insights from the above research and integrating concerns regarding the collision and fusion of anthropogenic activity with biological and geological systems (Chertow 2000; Crist 2013; Harraway 2016; Malm 2018; Stengers & Goffey 2015; Wainwright & Mann 2018; Wark 2016), BSD allows participants to approach new understandings of bird communication and perception in a more holistic and contemporary sense by actively participating in these systems from the standpoint of a bird. Promoting public awareness of scientific research in this participatory and immersive setting allows intuitive understanding of abstract ideas, and continues in the tradition of aesthetically informed scientific outreach (Hein 1990; Tuchman 1971). Through multisensory stimuli and participation, participants' awareness is focused to contain the audio and visual elements which are of primary interest to birds (Martin 2011). Therefore, both audio and visual content is selected based on proximity of concern for a bird in order to limit, or channel, the participant's viewpoint to that of a bird's perception.

The past has seen numerous depictions of birds and bird song in art. In music (Head 1997; Kraft 2013), in painting (Bugler 2012), sculpture (Brâncuși 1928; Calder 1971; Sosolimited et al. 2016) and installation (Simon 2014; Boursier-Mougenot 1999), birds have generally been represented in artwork as separate entities from the viewer. More recently, interactive and immersive artwork which engages the viewer as a participant within the work has dissolved the distance between the depiction of a bird and the becoming a bird (Kac et al. 1996; Legrady et al. 2013; Milk et al. 2012). The BSD project builds on methods used in interactive and immersive artworks to further close the experiential distance between participant and subject.

The difficulty in subjectively understanding the qualities of bird perception is managed through human participation within environments modulated to the concerns of birds (Gibson 1986; Nagel 1974). The necessary technology involved depends on the sensation that BSD is presenting to the participant. However, for participants to experience the complex dynamics focused on by the associated research, comparably complex methods for audio and visual production and immersion must be used.

Mirroring artificial life techniques of simulated natural systems for exploring the grammar of birdsong, the creation of complex multi-modal immersive installations brings these concepts to the public by having audience members actively participate as components of an emergent system. Involvement within the installation, while surrounded by bird-centric stimuli, enables an embodied understanding of both the larger social processes of birds and individual bird perception (Massumi et al. 2002; Thrift 2008).

In order to create such multisensory-media-saturated environments, the technical systems used to drive the experience have had to evolve with each iteration, adapting to the unique physical demands and cultural context of a space. For example, creating immersion in an outdoor space like Times Square requires strategies and technologies that augment and integrate the surrounding environment into the experience, while controlled, indoor settings have more freedom but also demand more effort to construct the entire experience. Striving to recreate bird-like perception involves the development of not only software simulations of bird behavior, but of hardware which can effectively reproduce these perceptions; often physically replicating behavioral characteristics. The trajectory from the early BSD installations to the most recent has seen a steady increase in scale, software complexity, and sculptural elements. As each iteration has gradually added more components to our overall inventory of materials, the later installations have been able to adapt more readily. Additionally, over the series of BSD installations, multiple collaborators with different specializations have shaped the technical and iterative design of the project to dovetail with his or her existing research. From audio feedback systems to sensor networks to machine learning classifiers to stereoscopic projection and robotic speaker systems, the combinations of novel forms of mediated experience have allowed BSD to explore various methods for constructing embodied experiences.

1.1.2 Multidisciplinary Collaboration

Bird Song Diamond (BSD) is a project we have been pursuing since 2012 and integrates the research of the involved scientists and engineers with the practices of artists and designers. We have taken scientific research performed in the laboratory and adapted it to function in an installation setting with interactivity. The chronological order of past events and installations are shown in Table 1.1.

The origins of this project with research into the grammar and meaning of bird songs have brought together computational linguists, electrical engineers, and artificial life specialists. For example, novel methods of data collection, soundscape partitioning, and automated annotation have been developed (Suzuki et al. 2017). More sophisticated analysis of bird song structure (Kasahara et al. 2012) was developed, and subsequently further insight into the syntactic and grammatical capacity of various species (Taylor et al. 2017). Work on soundscape capture and analysis has led directly to the development of the BSD Mimic system.

The sound and visual development of the project is also directly related to natural systems and how such systems react in the presence of humans (anthropogenic products). Work has been done to provide a more detailed explanation of the sonic aspects of landscape ecology resulting in a "soundscape ecology" (Pijanowski et al. 2011) that takes into account the collective biological, geological, and anthropogenic sounds (the biophony, geophony, and anthrophony, respectively), which comprise the entire soundscape (Krause 1987). Our research has contributed to the simulation of natural processes and environments in order to evaluate emergent behavior and real-world effects (Ikegami et al. 2017; Suzuki et al. 2018).

Using BSD as a platform to create such artificial systems in real world situations, as interactive installations, provides both unique audience engagement as well as opportunities for researchers to witness and experiment with artificial life interactions with the real world. The first BSD installation in 2014 took place at the sculpture garden on the UCLA campus (Figure 1.1) and employed artificial life techniques. This audio-only iteration, incorporated into the trunk and branches of a large Coral tree, acted as an autonomous sensor network (ASN), in which interconnected sensors took in environmental information (light and humidity), processed incoming data based on a chemical reaction model, and then modified its own sonic output based on these changes (Maruyama et al. 2013). The multi-channel audio was output

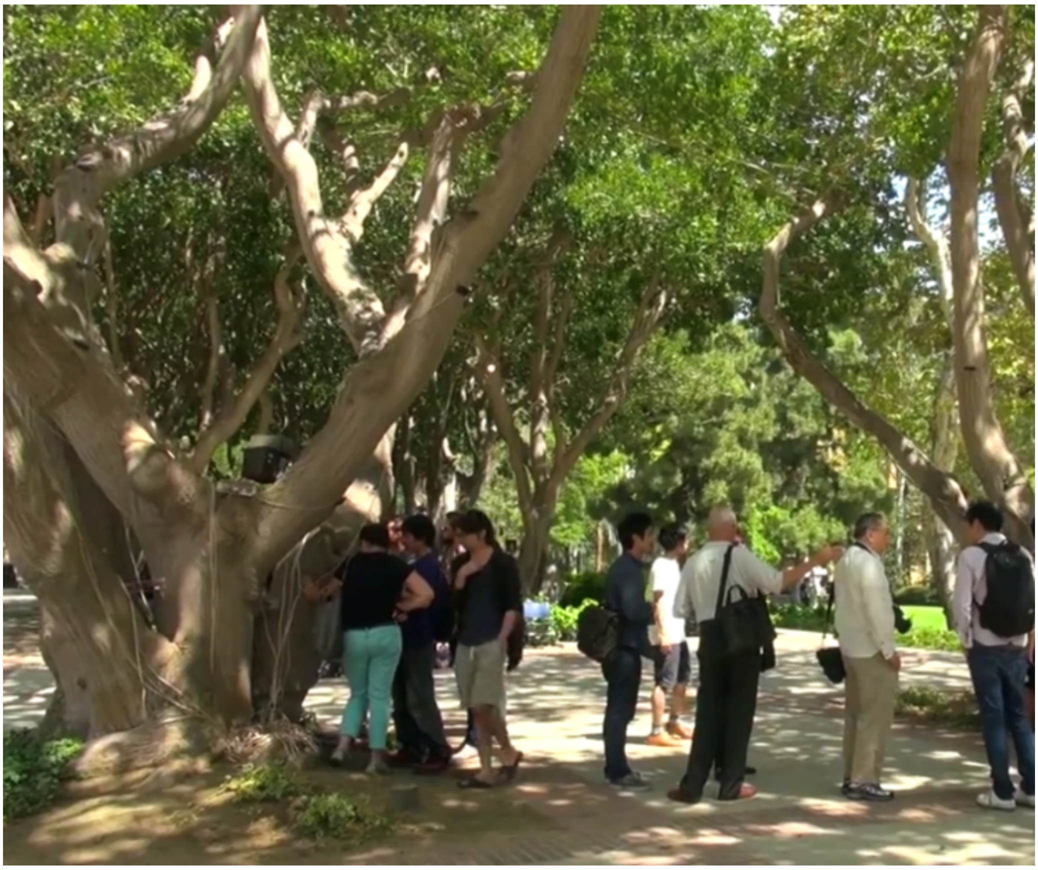


Figure 1.1: UCLA Sculpture Garden with the first iteration of BSD installed in the trees

through 20 parametric speakers, which are specialized speakers that produce highly directional “beams” of sound (Shi & Gan 2010) , giving the impression of many individual sound sources emanating from the branches of the tree. Due to the unique qualities of the speakers, the sonic effect was that of artificial birds inhabiting the branches of the tree. Throughout the installation the artificial sounds of the speakers and natural sounds of birds and wind and people created a complex soundscape underneath the branches of the Coral tree. As the day progressed and the environment shifted, the ASN in turn reacted and shifted its sounds.

Artist involvement was essential to the realization of the primary goal of the BSD project: public outreach. We have been able to successfully present BSD in festivals and public spaces thanks to the contributions of artists, allowing thousands of people to experience the project. Installation work, and especially interactive installation, requires historical and experiential knowledge in order to place the content of the research into contemporary cultural discourse. The artists developed the narrative structure and interaction design for each installation, and evaluated these aspects in preparing for successive installations. Additionally, the artists often participated in the installation by aiding audience members, explaining the system, and troubleshooting issues. Artists also consistently worked and communicated with researchers, leading to collaboration throughout the entire development process. They participated with scientists in the laboratory and out in the field, as well as invited the researchers to contribute to the development of the installations. As the director of the installation, the artist must take the role as a facilitator of specialists, working to bring the expertise of the individual researchers into the installation. The artist must also find coherence and connections between the aesthetics of the installation and the research so that the final outcome is unified.

1.1.3 Past Exhibitions

The most recent installation at the Ars Electronica Center in Linz, Austria, during the 2017 Ars Electronica Festival, featured the most developed and integrated BSD Mimic system to date. Taking advantage of gained insights from the previous versions, the interaction design was more streamlined and the underlying software more sophisticated. Audience members were invited to participate by attempting to mimic the birdsong of different species of birds drawn from Europe and Japan. Presented within the unique Deep Space 8k theatre, immersive large scale projected visuals with ten channels of audio supported the thematic aspects of the system while providing visual feedback from the BSD Mimic system.

The installation in Deep Space 8k integrated three diverse research interests for the first time in an art installation: large scale swarm simulation (Ikegami et al. 2017), controllable clusters of fixed-position directional speakers, a birdsong classification system for annotating many localized sound sources using a microphone array (Suzuki et al. 2017). While the first two systems were used for establishing the audiovisual environment, the third system provided a means for evaluating similarities between human and bird vocalizations. By comparison with the other two systems, BSD Mimic required far greater interdisciplinary collaboration between both the researchers studying bird song classification and the artists developing the installation. Even the existing software used in research was integrated with the installation software so that audience members were able to directly interact with the algorithms used in classifying birdsong. This BSD Mimic system became the primary mode of interaction with the audio and visual elements of the installation, while also creating a performative platform for the participants. BSD Mimic also allowed for post-performance analysis of how well individual vocalizations approached a target birdsong; further testing the limit of the birdsong classification software beyond its normal use case. The aggregate results of the performances were stored in order to evaluate overall user performance and quantitatively reveal the success or failure of the interaction.

A further outcome from receiving quantitative information regarding the BSD mimic system was that comparisons between collaborative interaction and individual interaction could be made. Having more concrete evidence regarding the effectiveness of different modes of interaction can assist with future installation planning. Iterations of BSD Mimic had made use of either solo or collaborative interaction design, and while observation had led to the hypothesis that collaborative interaction improved participant scores and overall success of the BSD Mimic experience, it was now possible to more formally evaluate data collected from the exhibition and post exhibition experiments

Table 1.1: Chronology of Major BSD Exhibitions with name and location.

Year	Exhibition
2012	Project Begins with initial stages of planning
2014	Sculpture Garden. UCLA, CA
2015	Art|Sci Gallery, California NanoSystems Institute. UCLA, CA
	New York Electronic Arts Festival. Governor's Island, NY
	Times Square. New York City, NY
	Nanolab, CNSI. UCLA, CA
2016	Large Space. Tsukuba, Japan
	Artificial Life and Robotics 21 (AROB). Beppu, Japan
	Waterbodies. Los Angeles, CA & Harvestworks, NY
2017	Deep Space 8k, Ars Electronica Festival. Linz, Austria
	Speculum Artium. Trbovlje, Slovenia
2018	Beall Center for Art + Technology. Irvine, CA
	CCD Memorial, FACTS Festival. Mexico City, Mexico

BSD has been presented in evolving forms across multiple locations since the first installment in 2014. Discussion of each iteration will be divided based on the type of interaction used in the Mimic system starting with an overview of individual followed by collaborative experiences. Details about other aspects of the installations will be covered to provide context for how BSD Mimic related to the larger BSD project.

As described above, and listed in Table 1.1, the first BSD installation was presented at the UCLA sculpture garden in 2014. The following year, BSD was shown on Governor's Island and Times Square in New York. In 2016, BSD traveled to Japan to show at the LargeSpace immersive display in Tsukuba. The BSD Mimic system was also brought to the 21st Artificial Life and Robotics (AROB) symposium in Beppu, Japan within the same month. September 2017 was the installation in Deep Space 8k at the Ars Electronica Festival in Linz, Austria. Following the festival, the BSD Mimic system traveled to Trbovlje, Slovenia for the Speculum Artium festival. Numerous other exhibits, containing elements of these landmark presentations, were conducted throughout this period.

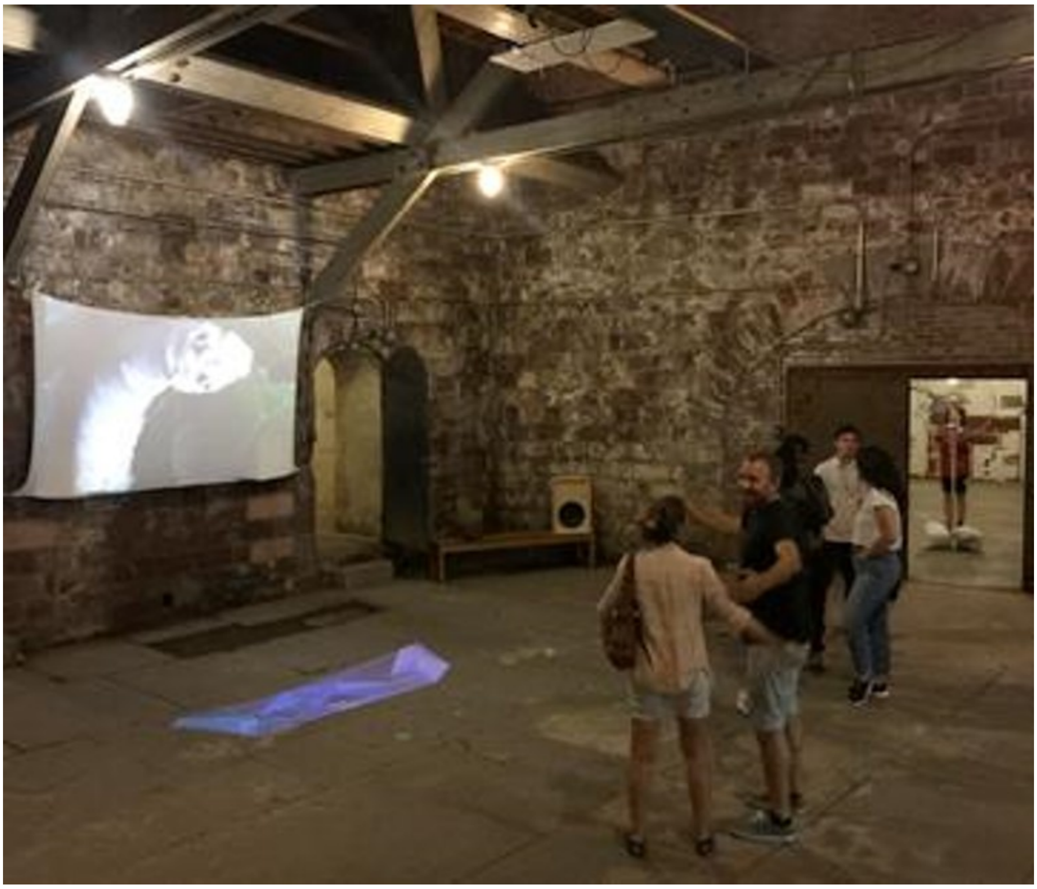


Figure 1.2: The central room of the BSD installation on Governor's Island. The wall and floor projections are visible on the left side. On the right side of the image, one of the BSD Mimic stations is visible through a doorway.

1.1.3.1 Governor's Island

As part of the 2015 New York Electronic Arts festival, BSD was to be installed on Governor's Island in the magazine of a former US Army post in the 18th and 19th century, now national park, known as Fort Jay. The installation was in a partially subterranean set of rooms with a distinctive "steampunk" aura (Figure 1.2). One section of the installation consisted of parametric speakers, used in the 2014 installation , mounted around the ceiling of the central room with a projection aimed at the wall opposite to the entrance. The unique layout and acoustics of the underground space allowed for a sharp contrast between the cold, military surroundings and the sounds of local bird song coming from the speakers. Based on the amount of movement within the space, the projection would cycle between different vantage points shot from a bird's perspective ranging from ground level to far above.

In two long, reverberant rooms attached to the central room, the BSD Mimic system was installed. This was the first publicly shown iteration of the mimic system. Participants received instructions from speakers placed in their respective rooms before singing into the microphone. The two stations on opposite sides of the main room were set up so that each participant could see the other from across the central room. Both participants could hear the other as they attempted to mimic each bird song creating a collaborative and slightly competitive dynamic. To indicate the state of the competition, the diamond was projected on the floor of the central room and would grow in the direction of the participant who received a better score. When the diamond reached the edge of the projection area, the diamond would be replaced by a new starting diamond in the center of the room.

Most attendees came in groups of two or more, usually on family outings or small day trips to the island. This being the case, there were rarely problems in finding multiple users to participate in the mimic system. The unique sound properties of the mimic rooms added another incentive for participants to vocalize, as the strong echo was pleasant to hear. While the separation of the rooms provided a direct view of the second participant, the visual feedback on the floor was not easy to view making the connection for the participants unclear at times.

1.1.3.2 Times Square

While presenting BSD on Governor's island, which included a collaborative version of the mimic system, a temporary mimic performance was presented in Times Square. Participants stood on a platform to elevate themselves above the crowds while wearing noise canceling headphones to further separate themselves from the sounds of the crowded street (Figure 1.3). The sudden shift to quiet, natural sounds of birdsong from the artificial soundscape of Times Square highlights the lack of integration of nature within urban landscapes. Attached to the microphone stand, a tablet computer was used

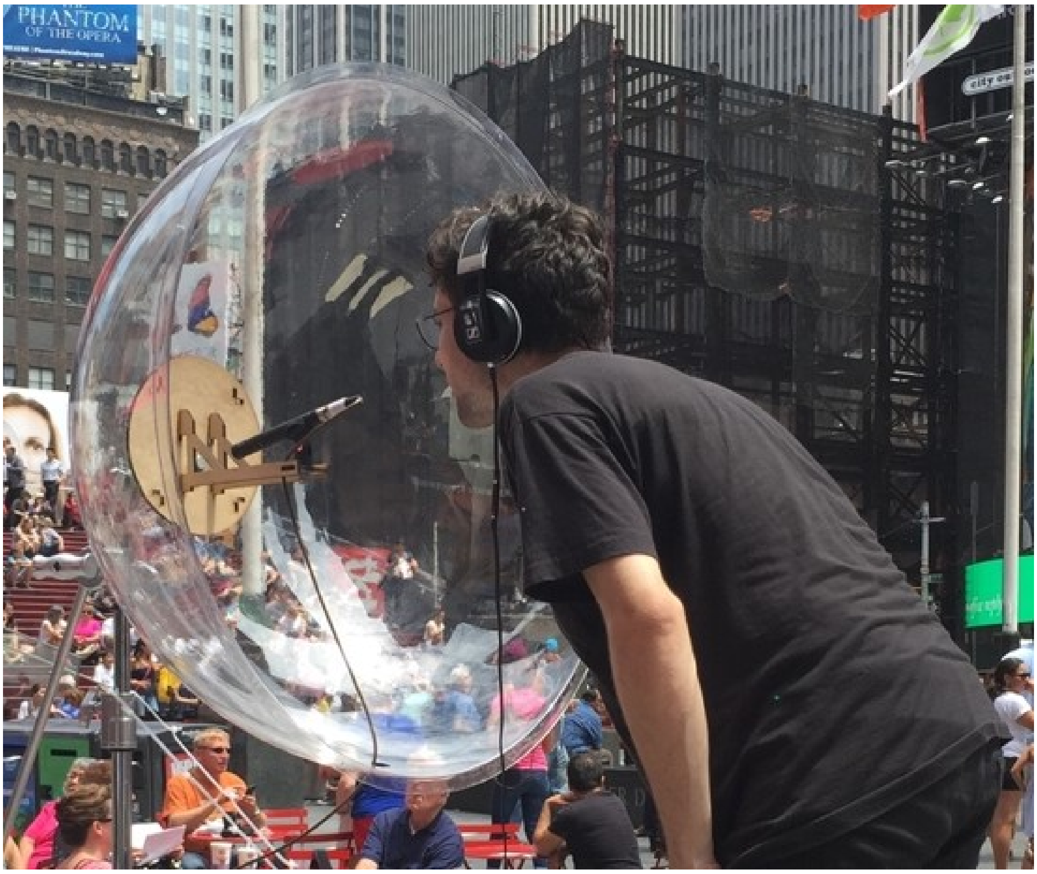


Figure 1.3: BSD Mimic in Times Square.

to display results in the form of a growing diamond generated using the Processing programming language (Reas & Fry 2014) and controlled from the Supercollider audio program using the Open Sound Control (OSC) communication protocol (Schmeder et al. 2010). The diamond would increase in size based on how accurately the participant mimicked the bird song. The elevation of the participant and audible separation from the chaotic surroundings of Times Square proved to be a successful effect in focusing attention on bird song. At the same time, the busy nature of the location and the "spotlight" nature of being on a pedestal made it difficult to find participants even with large numbers of people in the immediate area.

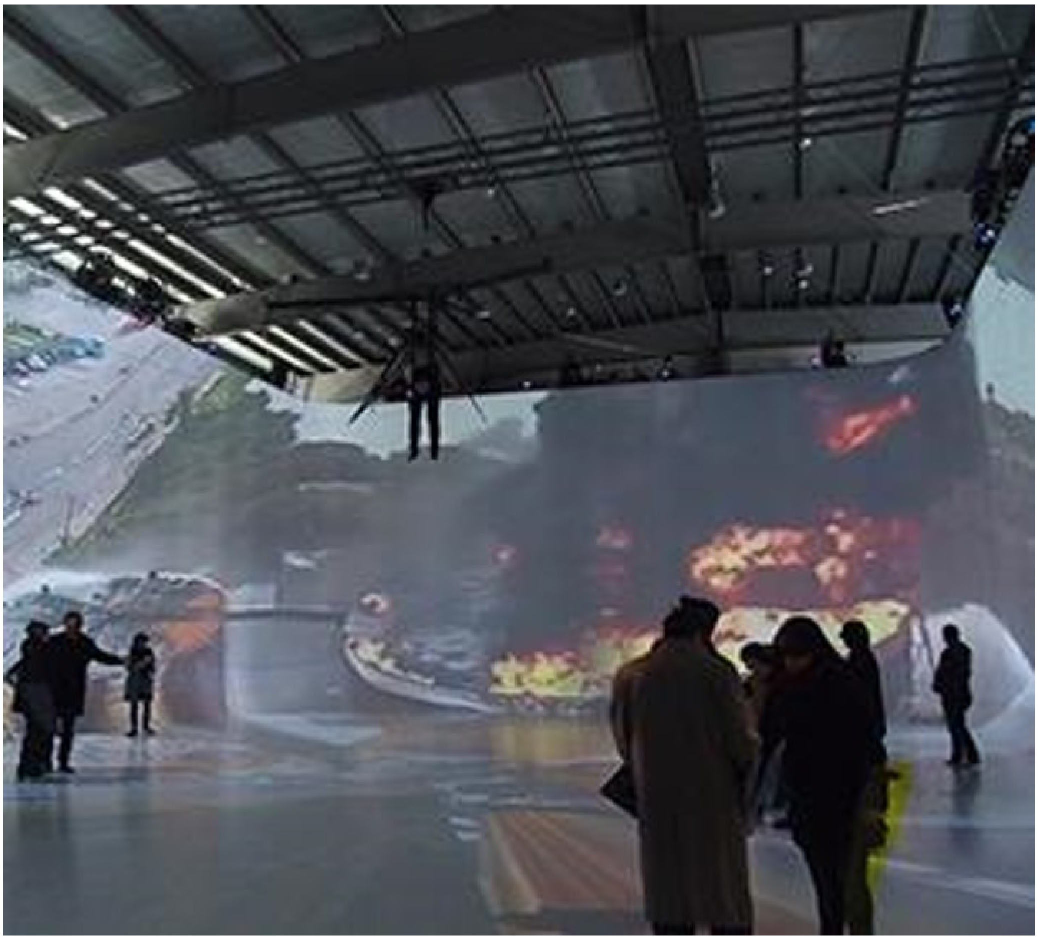


Figure 1.4: A section of the BSD LargeSpace performance in which a participant wearing a bird costume floats above projections of natural disasters.

1.1.3.3 Large Space

In the Winter of 2016, BSD was presented at the University of Tsukuba through an immersive display called LargeSpace (Figure 1.4). This display extends the paradigm of the Cave Automatic Virtual Environment (CAVE) system as a one-to-many format display (DeFanti et al. 1993) by dramatically increasing the footprint of the display area. The LargeSpace measures 25m x 15m, with a height of 7.7m enabling large groups of people to enter the display. Stereoscopic projections, created by 12 projectors around the upper perimeter of the space, cover the walls and floor while also accounting for the curvature of the projection surface. In addition to the display, the LargeSpace also allows for tracking a high number of participants using a

20 unit OptiTrack system. LargeSpace also has a wire-driven motion base to provide a single participant the ability to move vertically in the space (Takatori et al. 2016).

Prior iterations of BSD took a more traditional interaction design approach to installations in that the works were constantly running and able to be engaged by audience members at any time. The LargeSpace version of BSD, due to limitations on access to the space and the necessity for technicians to be involved at all times, shifted to be more performance oriented. With scheduled performance days and times, audiences would arrive together and all experience the project simultaneously. Audience members all wore active shutter glasses for viewing the stereo imagery as well as custom designed bird wings embedded with tracking system markers to determine each person's location in LargeSpace. Each performance lasted 15 minutes and was followed by a short discussion about the LargeSpace, the performance elements, and the underlying research.

The performance was composed of four distinct phases which presented a narrative portraying birds and bird perspectives in both natural and artificial forms. A single audience member, wearing an origami-inspired bird costume and attached to the LargeSpace motion base, would move through the space to indicate shifts from one phase to the next. Unique sounds, making use of the directional parametric speakers from previous installations, were created from a mixture of natural bird songs and composed artificial sounds (Maruyama et al. 2014). Additionally, interactive flocking simulations were used in different phases to juxtapose natural elements, represented by paper cranes, and artificial elements, represented by drones. The narrative culminated in a real drone flying into the space and influencing the flocking simulation with the use of a tracking marker on the body of the drone, and the origami bird audience member descending back to the floor to initiate a growing diamond.

BSD Mimic was set up outside of the LargeSpace projection room, and audience members waiting to experience the LargeSpace performance could test their abilities to mimic bird song. The instructions for BSD Mimic were translated to Japanese, and the collection of bird songs was changed to native

Japanese birds to reference the local soundscapes that would be familiar to those living in Japan. Adjusting the system to use native bird songs tested a person's existing familiarity with their surroundings, and additionally sought to have participants reflect on their future engagement with environmental sound. To make this version of BSD Mimic more portable, the pedestal to change the height of the participant was removed. A growing diamond shape was again used for visual feedback, but was projected on a nearby wall rather than displayed on a screen.

While many users were able to participate with the installation, it was clear that the scale and the spectacle of BSD in the LargeSpace drew much attention away from the BSD Mimic system. Even while not participating in the LargeSpace performance, audience members frequently watched the performance from outside the display area rather than attempting BSD Mimic itself.

1.1.3.4 Beppu

Participating in the AROB conference during the same month as the LargeSpace performance allowed for a second installation of BSD Mimic as an entirely standalone, individual interaction experience (Chacin et al. 2016). This installation tested the portability of the system as the entire piece had to be carried by the exhibitors in standard sized luggage. The system was shown outside of the conference room in a slightly darkened, yet open, area of the convention center. In the individual configuration, the audio of the piece was entirely contained within the participant's headphones with the audience able to hear only the participant's mimic attempts. This version tested the use of a semi-transparent, white plastic hemisphere as a container for the microphone which could double as a projection surface for the visual feedback of the growing diamond (Figure 1.5). Additionally, this was the first version of the BSD Mimic system in which the stand holding the dome and microphone was covered in a mirror-finish acrylic diamond shape. This shape added a physical analogue to the projected diamond, and also doubled as a storage space for the computer running the system.

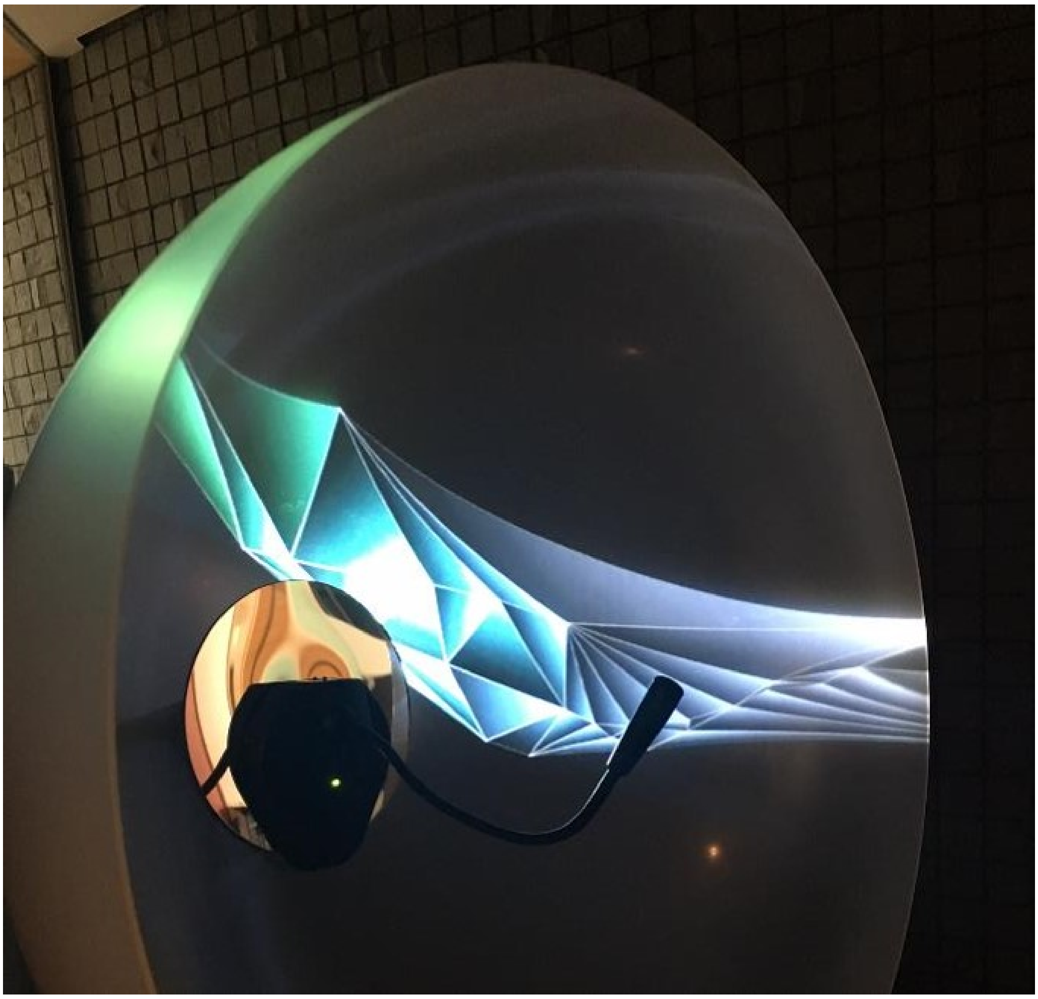


Figure 1.5: The BSD Mimic installation in Beppu projected user feedback on to a dome.

As a logistical experiment, the difficulty in transporting the hemisphere, which was smaller than hemispheres used in prior installations, proved that if portability was a primary goal for the system, then the current hemisphere material would not be sufficient. This led to later tests using different materials and collapsible constructions while still attempting to maintain the same hemispherical shape. Further, observation of the participants showed that many were hesitant to use the system. While many showed interest, the number of participants remained low. The context of the work in a scientific academic conference with few demonstrations may have contributed to this hesitancy, though social dynamics and feeling of becoming the focus of attention was also evident.

1.1.3.5 Deep Space

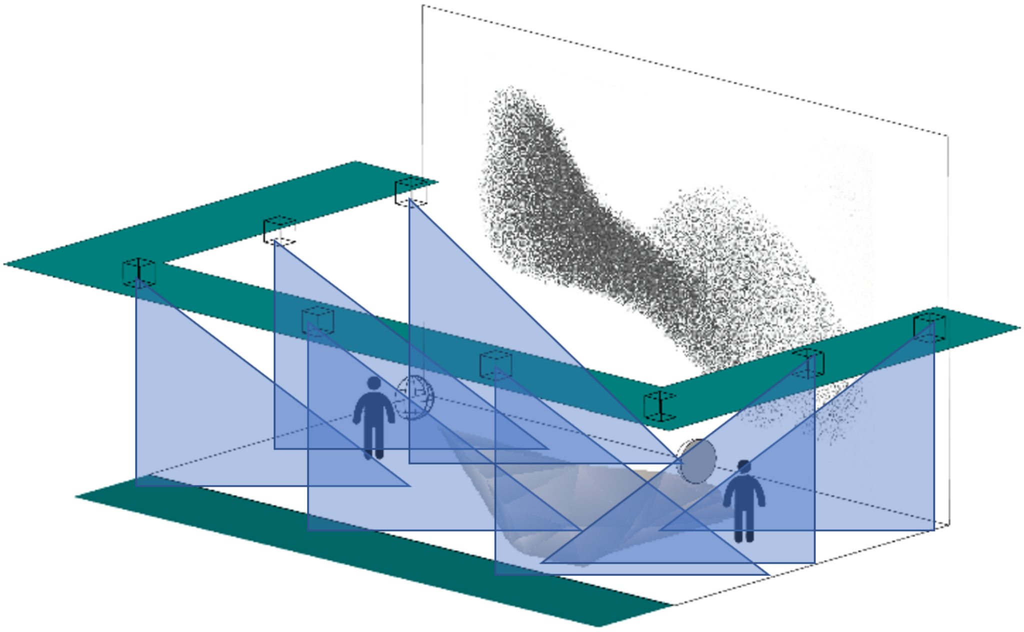


Figure 1.6: Diagram of the Deep Space 8k installation. Wall projection features boid simulation. Floor projection features diamond. There were 8 parametric speakers around top and two BSD mimic stations on floor.

Presenting BSD in the Deep Space 8k at the Ars Electronica Center in Linz during the 2017 Ars Electronica Festival offered the opportunity to expose a very large and diverse audience to the project. The Ars Electronica Festival is one of the premier festivals in the world to showcase projects that engage in art and technology. The average attendance to the festival for the five festivals leading up to 2017 was 82,500 people with a wide demographic of festival attendees. The Deep Space 8k is a large scale immersive display, in a similar category to that of the LargeSpace, and was recently upgraded from the original version of the system to handle both 8k resolution on the wall and floor surfaces (Kuka et al. 2009). With similar limitations as the LargeSpace regarding length and frequency of performance, it was initially decided to expand upon the systems from the prior LargeSpace performance, while adapting to the differences in display, sound, and tracking capabilities of the Deep Space 8k system. Given these circumstances, many major updates and modifications to different aspects of the system were developed.

One of the main goals for this iteration of the project was to fully integrate the BSD Mimic system with two other aspects from the previous installations: the large-scale visualization of swarms of birds with audience interaction; and multi-channel directional audio produced by parametric speakers. Incorporating these visual and audio aspects was further supported through the creation of a robotic positioning system allowing the directional speakers to follow the positions of the swarms of birds. By combining the three systems into a cohesive experience, it was expected that the increase in stimuli would further immerse the participants into the sensory experience of a bird. It was important to consider the connection between all aspects of the installation to construct an experience in which the audience feels responsible for the changes in the installation. Interactivity, in this respect, is meant to give a sense of ownership to the work, leading to longer engagement with the source material. Providing further cohesion amongst all elements, the BSD Mimic microphone stands were sculpturally enhanced to reflect the surrounding visual elements. Placed directly into the projection area of the Deep Space 8k, the stands became a focal point, in both visual and interaction, for the performance (Figure 1.6).

Much of the BSD Mimic system had to be re-evaluated and rebuilt in order to integrate it as a component along with the visual and audio aspects of BSD in Deep Space 8k. Formerly, the BSD Mimic system had been used as a standalone system with the primary goal of having a portable and easy to implement counterpart to the large scale and highly involved flagship BSD installations. Having some variability in potential implementations, the core of the system consisted of one or more microphones mounted to a hemisphere(s) and connected to a computer. Output from the computer was usually achieved through headphones attached to the computer which facilitated the audio-based interaction. Depending on the setting, visual feedback could also be created on a screen or projection surface.

On the software side, while the program was relatively straightforward to set up, and made use of a graphical user interface (GUI) to simplify the specification of input and output ports, applying adjustments to the difficulty level, and adjusting levels for noise reduction. However, even when the system had been set up and calibrated for a specific location, the output of the system, consisting of a percentage representing a user's accuracy at reproducing bird vocalizations, fluctuated in unpredictable ways with dramatic variations in percentages that often created more confusion and less time that each user spent with the system. In developing an updated version of the BSD Mimic system for the Deep Space 8k performance, both the physical and software shortcomings were considered while also working to integrate the mimic interaction with the larger performance.

In order for the hemisphere structures of the BSD Mimic system to match with the visual projections planned for the Deep Space 8k, a faceted shell, which borrowed visually from the diamond structures projected on the floor of the space (Figure 1.6), was to cover the functional stand structure. These shells expanded on the earlier reflective shells used in the Beppu installation described above (3.1.3). An artist created a 3D model of the structure, which was split into individual panels using the Autodesk computer aided design (CAD) software Fusion 360 (Autodesk 2018) and exported as a series of scalable vector graphics (SVG) files for use with a laser cutter. After cutting sheets of mirror-finish acrylic plastic into each individual face, the shell was constructed into segments and then attached around the outside of the metal support structure. Because the stands could not be in Deep Space 8k before or after the performance, the metal stand was attached to a custom designed wheeled platform. Wireless microphones were used to simplify the movement between spaces and to prevent the visual distraction and potential trip hazard of audio cables traversing the floor between the microphones and audio interface. The entire process was repeated for a second hemisphere. During the performance, the hemispheres were placed over the two diamond structures with platforms to allow users to reach the microphones. By using a mirror finish on the outer shell of the stands, broken reflections from both the projected visuals as well as the audience and participants within Deep Space 8k could be seen across the surfaces.

The software side of the BSD Mimic system was rebuilt to replace the former version for improved accuracy and consistency for a range of users in varying environments. These improvements required the separation of the software into two separate components:

1. The underlying program for analyzing and producing predictions

2. The front end which served as the user interface for interacting with the installation, and additionally acted to relay data (audio recordings, scores, triggers) between the first component and the software running the audio and visual aspects of the installation.

As mentioned before, the original BSD Mimic system, while easy to install and maintain, yielded inconsistent predictions and was difficult to protect against small changes in environmental noise, microphone quality, and variations in user vocalizations (vocal range, loudness, breath, and production means: whistling, humming, singing). To make improvements would require a more sophisticated approach, so work was done to adapt techniques used in scientific research on soundscape capture and analysis to be used for interspecies vocalization comparisons between humans and birds.

The resulting software made use of machine learning techniques to create a trained model of birdsong based on hundreds of field recordings for twelve different species of birds. Using this model, a recording from a participant could be fed into the system to produce a feature space distance between the human vocalization and the nearest bird vocalization. This distance represents the "closeness" in which the participant's song came to the song they were attempting to mimic. As with previous versions of the system, this value is then mapped to a percentage, from 0-100, to provide the user with a more familiar evaluation accuracy to the target birdsong.

The second, frontend component of the redesigned BSD Mimic system was built as an encapsulated object in the visual programming language Max (previously known as Max/MSP) (Cycling74 2018). This Max object managed the sequential aspect of the mimic interaction. After recording a participant's vocalization to a specified directory, the object would request an evaluation from the backend by sending an OSC message specifying the target bird song and which of the two participants was being evaluated. Upon receiving a reply message containing the mapped percentage value, additional OSC messages were sent to the software managing the visuals instructing the visual elements to shift accordingly. The percentage was relayed to the participant with an additional message and then the sequence waited for the next participant to approach the microphone. Because the interaction medium was audio based and Max was already used to drive all of the sound during the performance, including the mimic audio component as a Max object within the larger sound control software reduced the operational complexity of having to juggle an additional program.

Instructions were given verbally through pre-recorded text-to-speech (TTS) synthesis using stock Apple System Voices in two different languages, “Karen” for English and “Otoya” for Japanese, to accommodate international users. For an event in Slovenia after the Deep Space 8k performance, an additional voice needed to be recorded. Because there were no available TTS voices in the Slovenian language that matched the quality of the other two languages, verbal instructions were recorded by a human. The inclusion of languages and bird song from different parts of the world reflects both the backgrounds of the researchers and artists working on the project as well as the locations in which the project was previously shown.

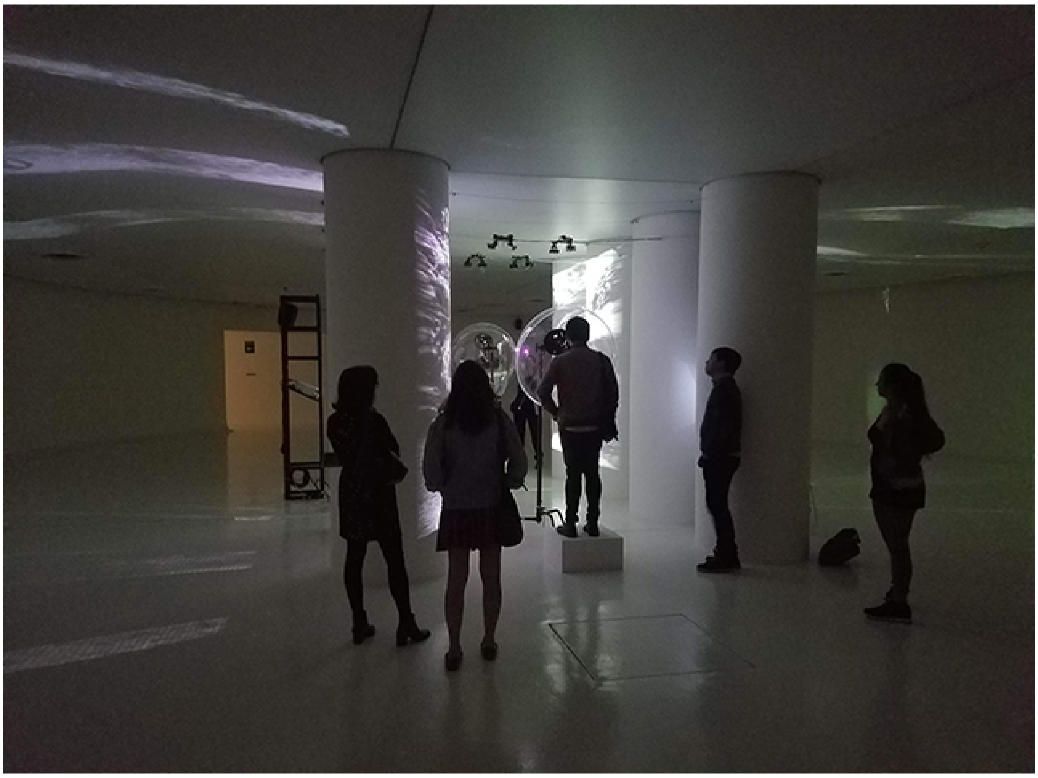


Figure 1.7: The central corridor of columns in the CCD Memorial space. The Spatially Immersive Sound system can be seen above the BSD Mimic participants.

1.1.3.6 CCD Mexico Memorial Space

Invited to exhibit BSD in Mexico City for the exhibition “Espacios de Especies” at the Memorial Space of the Centro de Cultura Digital, located underneath the controversial “Estela de Luz” monument commemorating Mexico’s independence, became an opportunity to fully test the updated version of the spatially immersive sound system. The complexity of the space as well as difficulties created by restrictions for light, mounting of the speaker mechanisms, and unique acoustic properties all led to a number of modifications of the previous elements of the work. Memorial is an immersive space, but not a CAVE-like or with projection capabilities as LargeSpace or Deep Space had available. The space, painted fully white, had an array of colored lights in the ceiling, the colors of which could be manually or programmatically controlled to change the color of the space. Unfortunately, the combination of the light with projections was not a possibility, so the importance of the projected elements remained and the lights of the Memorial were off for the duration (Figure 1.7).

Due to the duration of the exhibition, November 2018 through February 2019, the installation needed to withstand daily use for a number of months. This meant that the deep learning Mimic system developed for the Deep Space performances could not be used because of difficulties in setup and maintenance, so the software was reverted to the SuperCollider based program which had already been used in the long term Governor’s Island exhibition and had proved its stability. The projected visual elements also shared similarities to the Governor’s Island installation in that they could modulate between different bird-related perspectives: abstract boid-simulation from the Deep Space performance, close-up worm and earth views, and drone-views of immigrant caravans traveling across Mexico. The birds songs used in the Mimic system were sourced from local species to create the local connection between the participants and their surroundings.

An updated, wireless version of the spatially immersive sound system was installed on wires in the center of the room amongst eight structural columns (supporting the above ground monument). The speakers were designed to follow the a flock of virtual birds as they flew throughout the Memorial space, and while invisible to the participants, the movement of the invisible boids could be detected by the movement of the sound in the space. Because the spatially immersive sound system was not designed for long term use, it was only installed during the first few days of the exhibition in order to record participant behavior and receive survey feedback regarding participant experiences which will be discussed later.

1.2 Human Informatics

The field of Human Informatics, as the name implies, combines psychophysiological aspects of human beings (cognition, brain function, sensation, motion, physiology) with the rapidly expanding and increasingly ubiquitous and indispensable (at times to a worrying amount) realm of information technologies, systems, and data-driven techniques. As a subfield contained within informatics, it takes the human subject as a system which produces data. While this approach has been critiqued in light of governmental and private data collection and surveillance using large scale data collection of digital traces and behavior, the aim Human Informatics approach takes a different strategy.

1.2.1 Goals of Human Informatics

Generally concerned with quality of life for humans, Human Informatics deviates from the more egregious forms of data collection mentioned above. Further, Human Informatics, at least where it begins, is not necessarily seeking to understand more generalized social or economic issues related to large scale human behavior, but rather to explore and investigate human functions through deliberately placed physical sensors to monitor and aid in the functional approach to an individual's environment. Goals here revolve around this human-environment relationship. What are methods to enable a person to engage more effectively with devices in the home? How can workers discover more creative solutions to existing problems and how can they access and organize information related to the problem at hand? In emergency scenarios, what are the problems facing first responders as well as those affected, and what technologies can be developed that alleviate physical and psychological tolls on these individuals?

1.2.2 Approach, Methods, and Techniques

The combination of sensory and perceptual measurements and data repositories, existing and speculative theories related to human brain and cognitive functions, and social relationships with new forms of actuators, robotics, displays, interfaces, network connected devices, and mixed reality environments for the application within designs, products, services , and systems is the primary approach taken by Human Informatics researchers. It is often only possible to achieve this multifaceted combination through interdisciplinary research, with specialists in each of these areas contributing key components along the developmental pipeline to eventually create material outcomes for individuals and society.

The ability to alter the focus of an existing research field is a common strategy. For example, the field of Human-Computer Interaction relies on an understanding of the psychophysical limitations of humans for perceiving and reacting to changes in traditional interfaces and interaction paradigms. The Human Informaticist considers similar background information while also questioning the existing frameworks for interaction, often proposing entirely different modes of interaction and interface design based on advances in technology. These methods must take into account the historical developments of psychology, neuroscience, and engineering while considering contemporary contexts, environments, and living conditions as well as projecting these considerations toward speculative and future scenarios. The Human Informaticist attempts to alter the *Angelus Novus* (Figure 1.8) interpretation of historical progress:

> "A Klee painting named Angelus Novus shows an angel looking as though he is about to move away from something he is fixedly contemplating. His eyes are staring, his mouth is open, his wings are spread. This is how one pictures the angel of history. His face

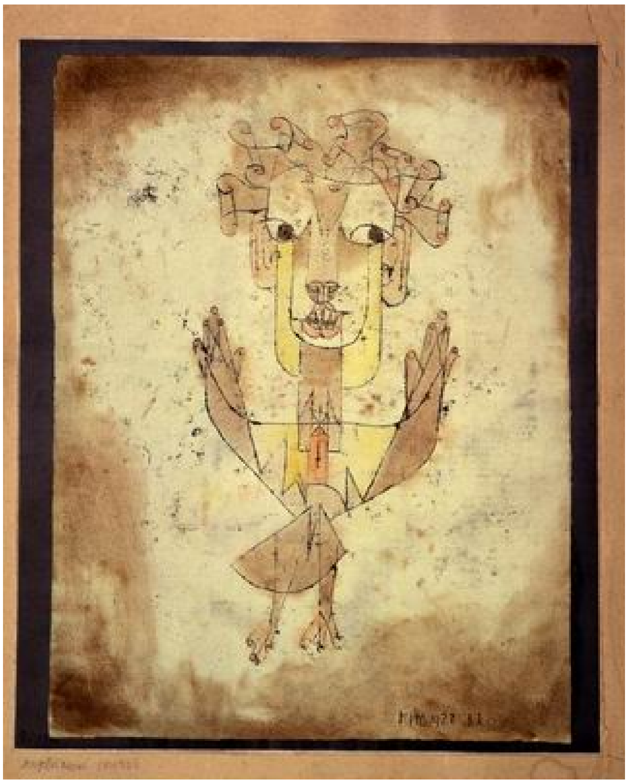


Figure 1.8: *Angelus Novus*. Paul Klee. 1920

> is turned toward the past. Where we perceive a chain of events, he sees one single catastrophe which keeps piling wreckage upon wreckage and hurls it in front of his feet. The angel would like to stay, awaken the dead, and make whole what has been smashed. But a storm is blowing from Paradise; it has got caught in his wings with such violence that the angel can no longer close them. The storm irresistibly propels him into the future to which his back is turned, while the pile of debris before him grows skyward. This storm is what we call progress"(Benjamin et al. 1968)

Rather than uncontrolled propulsion into future, with eyes fixed on past failures, the Human Informaticist angel takes on a Janus-like visage, a dual gaze while sitting at the pivot of the contemporary. From this perspective, especially for researchers who's fields are often solely concerned with technological advancement without consideration of humanistic and anthropological meaning, it is necessary to remember and to remain engaged with the Human qualities of Human Informatics. On the other side of the coin, emerging scholarship such as Digital Humanities (Burdick et al. 2016) consists of researchers who, previously unconcerned with the role of technology outside of word processors, are adopting methods and techniques from Informatics, Engineering, and Neuroscience. Some argue that, for the most part, research in Informatics is nearly always biased toward human advancement, making the addition and distinction of *Human* Informatics a redundancy. However, it is the placement of *Human* before *Informatics* that acts as a reminder for those who adopt this research moniker to be cognizant of the anthropological nature of their research.

1.2.3 Beyond Human Centricity, an ecological perspective

The claimed holistic nature of Human Informatics is admirable for considering not only the mechanical, but the cognitive and social nature of human functionality as well (Sato 2018). The combination of humanistic endeavors with psychophysical aspects mentioned above truly encompasses most of the concerns that face contemporary persons. Yet the whole of humanity is not isolated. This is apparent in much of the HI research done on disaster preparedness and reaction. Human-centered Design (Cooley 1996) plays a major role as one of the primary methodologies of HI, but as its name implies, the human is at the center and the environment is a secondary concern. In Empowerment Informatics, one of the major tenets is *Harmonizaton*, for alleviating and smoothing the experience between humans and engineered systems (Iwata 2015). To reframe harmonization for Human Informatics as a whole, it is necessary to move beyond human-centered design, to the surrounding environment at localized and larger geological scales, to the organic and inorganic; animal, vegetable, and mineral. Ultimately, these concerns loop back, at varying timescales, to affect the human in Human Informatics.

1.3 Bird Song Diamond and Human Informatics

The focus of the bird in Bird Song Diamond comes from the backgrounds of the behavioral ecologists forming the underlying research project that BSD draws from. This research, as mentioned above, has shown that the complexity of bird song can approach the complexity of human language. To achieve these early findings, much research has been done on bird physiology, sensing techniques, and data-driven models of bird song. This approach shares many similarities to the approach which Human Informatics takes in regards to human subjects. In this respect, the commonalities shared between the two fields is quite striking and illustrates that the techniques in Human Informatics can be applied to research outside the sphere of anthropocentric concerns. A strong example of the crossover from bird research to HI related applications is the development of networked sensor arrays for automatic spatialized recording, recognition, and classification of bird, bird location, and bird song, which alone could inhabit similar research questions in the field of HI and specifically in the field of Computer Supported Cooperative Work (CSCW).

Additionally, by bringing research in behavioral ecology, computational linguistics, and artificial life to the public, BSD illuminates the connections and relations between human and bird subjects. It is through the experience of being a bird, becoming an *other* or non-human, which is made possible through the use of immersive spaces and particularly through the use of spatially immersive sound, that alternate avenues of thought based on unearthed bird-human relationships can be developed. The necessity of bird song and its potential loss due to a detachment from nature (urban living) and detrimental anthropogenic effects on the environment (Carson & Darling 1962) is highlighted in the BSD experience. The emergence of participants into the world outside of the immersive space of BSD brings an altered perception of hearing related to the recognition and appreciation of bird song, which is instilled in participants throughout the installation.

1.4 Summary of Chapters

Chapter 1 begins with an introduction to the Bird Song Diamond project, including the primary aim of the bird song research and all affiliated research fields, the importance of the Bird Song Diamond project for human informatics will be explained. Briefly, rather than adopt a human-centric position for considering human activity and well-being, the Bird Song Diamond project, and the resulting spatial immersive sound system, demands a holistic awareness of interspecies and environmental relationships in order to ultimately improve the human subject. Following this broader introduction is a clarification for the reader, of sound, spatial sound, and spatially immersive sound. Included in **Chapter 2** are works that relate to, or involve aspects of, spatially immersive sound. A distinction is made between the understanding of spatial immersion throughout history, musical and acoustic techniques for spatial immersion, and contemporary technologies used to produce such effects. The importance of spatially immersive sound in the context of the Bird Song Diamond project's installation and conceptual framework is then explained.

After establishing spatially immersive sound in the previous section, **Chapter 3** provides a thorough account of the system used in the Bird Song Diamond project and gives the reader a detailed understanding of the underlying technical aspects. This includes details regarding early versions, development, improvement, fabrication, and calibration. Finally, examples of installations of the system in two international exhibitions, Austria and Mexico, provides background for **Chapter 4** in which methods for testing and evaluating the effectiveness of spatially immersive sound in exhibition settings, as well as methods for evaluating basic concerns about the effectiveness of directional sound using parametric speakers. From multiple prior installations to the user study and participant interviews, the spatially immersive sound system has been shown to empower users through stronger engagement with complex networks of bird song and movement at the heart of the Bird Song Diamond project. **Chapter 5** outlines the findings from the tests outlined in Chapter 4, while placing these findings into the broader concerns of Empowerment and Human Informatics.

Additional applications for spatially immersive sound outside of the Bird Song Diamond project can provide future research possibilities. With **Chapter 6**, possible developments are considered before a deeper look into an alternative method for creating spatially immersive sound using headphones connected to a standard smartphone. While not yet fully implemented, the project has had much development and constitutes future work in this burgeoning research field. Finally, only through the collaboration of artists, scientists, and engineers with the aim of co-developing interconnected systems and artworks has this system emerged, and the significance for inter-disciplinary, inter-field, inter-cultural partnerships cannot be underestimated.

Chapter 2

Related Works

Spatially Immersive Sound (SIS) is distinct from most familiar forms of sound used in immersion and locative techniques. Perhaps the name will cause more confusion for those already working within the fields of spatialized sound and immersive sound, who make use of refined techniques that given the proper setting and equipment provide highly precise and convincing virtual sound sources or appropriately support a larger immersive setting. SIS, however, occupies a place that is more unique, somewhere between the precision of some systems and the supportive nature of others (these systems will be covered later on in this chapter), but also stepping outside of the spectrum defined by these two conditions. The SIS system proposed here and developed in support of the BSD project, relinquishes many of the requirements of existing sound immersion techniques in order to explore previously unreachable sound interactions; notably: multiple localized virtual sound sources in the company of freely moving listeners and flexibility for adapting to a variety of spaces. Issues regarding the free listener movement, multiple virtual sound sources, and location agnostic capabilities will be contrasted with existing techniques and technologies which have historically supported immersive sound. Additionally, the fundamental technology driving SIS is clarified while projects that share technical characteristics with the proposed SIS system are described in more detail. Prior to these comparisons, the concepts of immersive and spatial sound will need to be unpacked.

2.1 Immersive Sound and Spatially Immersive Sound

For users in a virtual environment, fully addressing all physical senses is a necessary beginning for achieving total immersion. Following this virtual dominance over participant perception comes interaction with the synthetic environment to drive home the "reality" of the virtual. Immersion and sound, however, have a historic relationship that is related to, but not always associated with convincing a listener of a novel environment that is generally the goal of virtual reality immersion (Apollinaire 1984). Here, for the sake of simplicity sound will be conflated with music, but the reader should be aware that this is a complicated issue and varies depending on the time period, culture, and author (Kahn 1999; Hegarty 2007; Cage 1961; Schaeffer et al. 2012).

One form of immersive sound is program music, here meaning music following a program or narrative, which seeks to aid in the feeling and atmosphere of an existing story, supporting actors through instrumentation, tempo, dynamics, and eventually through melodic and harmonic motives. With sound acting as a supporting character to the a narrative, maintaining audience immersion in a story meant that the sound should not be distracting or overpowering. Contemporary film scores continue in a similar tradition with sound, both in the conception of music as well as sound effects. Possibly in traditional forms of music, attempts at immersion through "imitative" musical sounds, which are often seen as anathema to serious forms of composition, have been in support of recreating bucolic settings within the concert hall; a trombone imitating a cow, a flute imitating *bird song*. In a concert setting where sound, or more often music, is the focus, what defines immersion becomes less clear. Rather than using the term "immersion", one might instead look to descriptions of music as entrancing, evoking a sort of ecstasy. Becoming "lost", "drunk", or overwhelmed by sound is an alternative conception of immersion that connects sound with the imagination, mysticism, and the unknown. For both of these characterizations of immersion with sound, various methods, techniques, and technologies have been utilized to achieve such an effect.

Another factor, necessary for returning back to the topic of SIS, is the addition of spatialization to these existing mechanisms of immersion. With the advent of audio recording and playback through inscription and electronic transmission, much of the spatialized effects of live music and performance were lost, at first by technological limitations, later by standardization; monophonic audio dominated the majority of the world's playback soundscape (movies, radio, phonography, etc.) until stereophonic playback became the de facto standard in the second half of the 20th century (Beardsley & Leech-Wilkinson 2009). Even with movie theatres and "home entertainment systems" adopting digital surround sound and other multichannel audio-panning technologies (discussed in the next section), audio experiences that utilized spatial immersive sound were uncommon outside of specialized theatres and one-off performances. The following sections will attempt to describe developments in SIS and the various components that comprise methods for achieving sound immersion and spatialization.

2.2 Early Examples

For as long as the acoustic properties of sound have been studied, attempts at harnessing, focusing, and diminishing the effects of sounds have been explored. Through architectural spaces (Vitruvius Pollio. & Morgan 1960) and instruments designed to focus or block sound waves to resonating elements that sustained and amplified plucked, blown, and struck objects, sound has been manipulated for harnessing desirable properties while eschewing "noise" and other undesirable elements. Whether for clarity, aesthetic beauty, or scientific inquiry, as the ability to direct and manipulate sound became more advanced, the notion of immersiveness of sound could be explored through a variety of techniques. The term *immersive*, literally meaning to be surrounded, enveloped, or plunged underwater, was not described metaphorically until the emergence of virtual reality technology in the 1980s (Lanier & Heilbrun 1988), however media theorists and technologists have described the use of equipment for such purposes since the middle of the 20th century (Bush 1945; Sutherland 1965). It could be surmised that not until the material possibility of a totally immersive experience, in the metaphorical sense, that this usage was introduced into the collective lexicon.

2.2.1 Immersive Spaces

The earliest architectural acoustics, especially due to a lack of amplification, were concerned with how to effectively carry the human voice to large audiences. This can be seen in the open air theatres developed by the romans, and an even earlier reference by Vitruvius to a certain “sounding vase” or *echeia* which, when placed in particular locations in the theatre, resonated with the speaker’s voice to enhance the effects of the words:

> “In theatres, also, are copper vases and these are placed in chambers under the rows of seats in accordance with mathematical reckoning. The Greeks call them Echeia. The differences of the sounds which arise are combined into musical symphonies...it becomes fuller, and reaches the audience with a richer and sweeter note.”(Godman 2007)

Becoming immersed in the resonating words of an orator is not an uncommon experience, and one can imagine that by resonating the entire theatre with tones derived from the speaker’s own voice, perhaps having an oblique connection with Robert Ashley’s “She was a Visitor” (Ashley 1967), that the experience was super effective. Architectural features for intensifying sound, whether spoken word, a singer’s voice, indications of time and prayer (usually through bells or singing), or emphasizing the sublime and heavenly in religious ceremony, continued to be expanded throughout history; often a literal expansion, as the size of venues from temples to cathedrals relied on open, interior spaces for cavernous echoic sound.

In 19th century Europe, Wagner, fueled with extravagant funds from his patron King Ludwig II of Bavaria, constructed an opera house in Bayreuth specifically for the performance of his own works. Known as the “Bayreuth Festspielhaus” and completed in 1876, the theatre epitomized the sort of domineering control that Wagner commanded as the sole creative force behind his work; pushing Romanticism to absurd proportions. This space extended Wagner’s “total artwork” (Wagner et al. 1964) into the material context by exercising designed control over the audience’s experience of the work. The interior construction from the angled wedge of the "continental" seating layout, the separation of the proscenium from the audience, and the hidden orchestra pit all intended to strengthen the content as well as the reception of the work. It is clear that through such tight control over the audience experience, that Wagner sought to guarantee a level of immersion in the performed work.

While aimed at realistic rather than mythical goals, another architectural feat offered the experience of traveling the world without setting foot outside. Arguably at the apex of the 19th century form of mass media, the panorama, the Maréorama represented a confluence of perceptual sensations for creating an immersive experiences for the visiting crowds. Presented at the 1900 Paris Exhibition, the Maréorama was a moving panorama that took passengers on a simulated trip from Marseille to Yokohama. The imagery, in the form of a dual moving panoramas on either side of the ship was designed by Hugo d'Alesi and painted on 750 x 13 m rolls of canvas that unraveled over time to simulate the passage of a boat. Spectators stood on a 70m long platform mounted on a hydraulically controlled gimbal, dressed to appear as the deck of a steamship and included boat-like motion, smokestacks that produced smoke, the sounds of the boat's propeller and steam whistle, and "deck hands" that moved about the platform. Fans created the feel of ocean breeze replete with the smell of seaweed, and lights emulated the passage from day to night with the occasional flash of lightning. With all of these painstaking details, up to 700 people could experience a year-long voyage in the span of thirty minutes. While soon eclipsed by the moving picture show of the Lumière brothers, which had already been established five years prior, the Maréorama provided a kind of full body experience that acts as an early antecedent to the CAVE system.(Huhtamo 2013)

At the middle of the 20th century, and at another world exposition, the Philips Pavilion acts as an early example of a spatialized, immersive audio-visual experience. Presented in 1958 at the World Exposition in Brussels and conceived by architect Le Corbusier as *Poème Électronique*, the entire formal structure and internal experience followed a conceptual plan laid out by Corbusier: "I will not make a façade for Philips, but an electronic poem. Everything will happen inside: sound, light, color, rhythm ... Per-

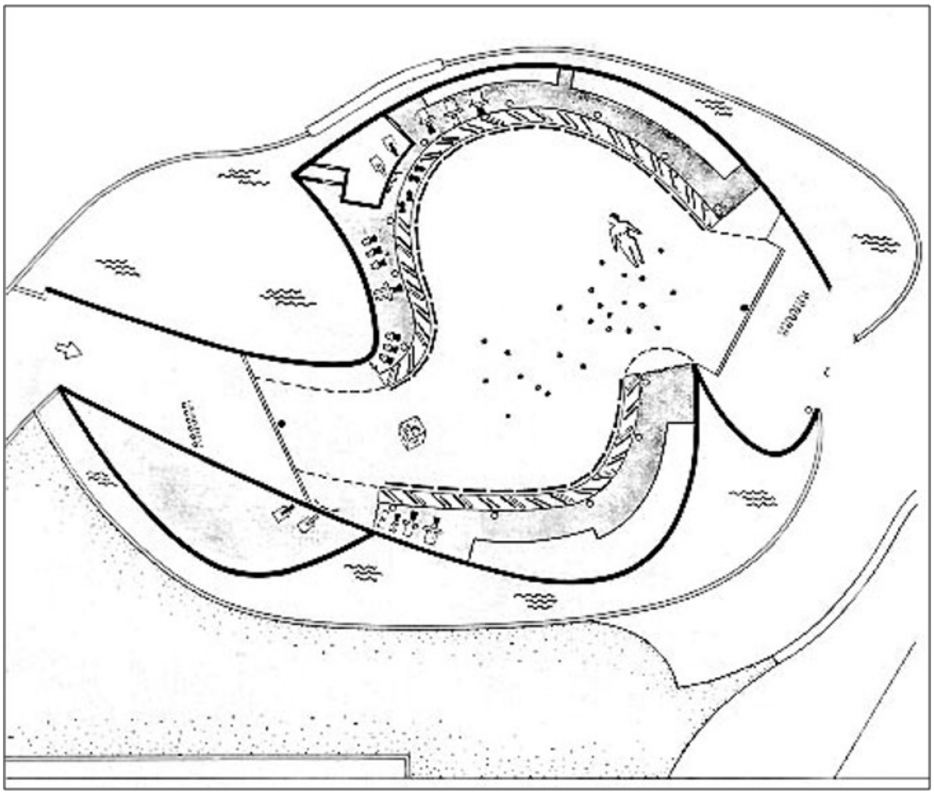


Figure 2.1: Diagram of the Philips Pavilion. Le Corbusier modelled the shape after a stomach to match with his metaphor of transforming the public as they passed through the building.

haps a scaffolding will be the pavilion's only exterior aspect" (Figure 2.1). With the intention of the experience transforming the public as they moved through the pavilion, the interior of *Poème Électronique* contained a barrage of media including projected film, superimposed lighting elements, illuminated sculptural objects, and most significantly music, composed by Edgard Varese, which could be spatially controlled across approximately 425 speakers (though the exact number is contested) embedded into the structure of the pavilion. The music, an electronic tape composition for three channels, was clearly not composed directly for the speakers of the building (a 425 channel work), instead the distribution of the sound across the speakers was controlled by "sound projectionists" which used rotary telephones to turn on and off groups of speakers (Lombardo et al. 2009; Valle et al. 2010). Varese had spent his career attempting to create spatialized sound by using traditional compositional techniques and described the pavilion as "a spectacle of light and sound [with]"sound routes" to achieve various effects such as that of music running around the pavilion, as well as coming from different directions, reverberations, etc. for the first time, I heard my music literally projected into space."(Chadabe 1997)

2.2.2 Techniques

There have been a number of musical techniques for increasing the spatial and immersive feeling of a work. By far the most straightforward and direct way to achieve this is through scale. Resizing the work, prior to the prevalence of powered amplification, meant an increase in personnel, increase in number of instruments, and in turn a larger space to hold everything. Additionally, increasing the length of the work also adds to the monumental nature of a work, and surely a meek audience member, awestruck and overpowered by the scale of the ensemble, will feel a certain immersion in the sublime nature of the experience. This is what many composers in the romantic period of European classical music of the 19th and early 20th century attempted to achieve. Berlioz's *Requiem*, Mahler's *Symphony No. 8*, and Schoenberg's *Gurre-Lieder* all represent this strategy of massive scale with each requiring over 400 performers. The composer Scriabin's unfinished *Mysterium*, strongly influenced by theosophic symbolism and the composer's own synesthesia, was to be performed in the Himalayan foothills over a period of seven days and incorporate thousands of performers, colored lights, smells, and touch, all culminating in the transfiguration of humanity itself. Fortunately for humankind, the work was unfinished.

In western music, romanticism and grandiosity subsided in the wake of modernism (generally speaking), yet new conceptions of sound and approaches to music and listening brought to the forefront other means for achieving listener immersion. The Dream Syndicate or Theatre of Eternal Music was formed in the 1960s by a group of musicians and artists influenced by Indian Classical Music, Japanese Gagaku, natural sounds, and electronic hum, who through collective improvisation using long, sustained tones, aimed at creating a deeper focus on the practice and intention of composition, of "getting inside the sound" so that the body is no longer perceivable.(Benjamin Piekut & George E. Lewis 2013) Pauline Oliveros practices the technique of *Deep*

Listening in order to allow the inclusion and consideration of all sounds, her improvisations follow a collective intentionality based on localized and global attention in reference to mandala symbolism(Benjamin Piekut & George E. Lewis 2013; Miles 2008; Oliveros 2005). For many, the roots of this practice are located with John Cage, influenced by the Zen Buddhist teachings of D.T. Suzuki directly incorporated Zen concepts of "unimpededness" and "interpenetration" as well as non-involvement. Cage described unimpededness and interpenetration, respectfully, as:

> "…seeing that in all of space each thing and each human being is at the centre and furthermore that each one being at the centre is the most honored one of all … Interpenetration means that each one of these most honored ones of all is moving out in all directions penetrating and being penetrated by every other one no matter what the time or what the space, [so that] there are an incalculable infinity of causes and effects, that in fact each and every thing in all of time and space is related to each and every other thing in all of time and space."(Nyman 1999)

Collective and sustained engagement with sound, especially welcoming in sounds that would normally be eschewed as non-musical, as techniques for immersing oneself in sound has only emerged in Western music through the inclusion of non-Western performance concepts; often traditional and established. Japanese composer Toshi Ichiyanagi, regarding his piece *Appearance*, said the "piece creates something, but not a whole thing It leaves things open At the same time, outside elements appear … It's like an old Japanese garden design: those outside elements like the moon, the clouds, the trees change all year round You look at the movements of the stars Those things are included in the garden; however. they are not controlled by the creator."(Nyman 1999)

2.2.3 Technologies

Regarding spatialization and localization of audio sources, the potential for controlling these factors had been addressed somewhat by techniques, placement of performers or sound-making devices in theatres to produce sounds in which the source was not visible to the audience, within the audience, or positioned around the audience. However with technologies for recording and audition, and especially with the introduction of stereophonic audio mentioned above, the possibilities of controlling the perceived location of a recorded instrument or synthetically generated sound became more viable. The most basic version of sound spatialization using two or more speakers is to pan the audio between speakers by varying the intensity of a sound across a number of speakers. For example, a sound that occurs to the left of a listener could be represented across stereo speakers by having maximum intensity from the left speaker and minimum intensity from the right. Surround sound is, for the most part, an extension of this panning technique, with multiple sounds mixed across a larger number of speakers. Beyond panning techniques are more complex technologies that attempt to model different aspects of sound perception to convincingly locate sounds around a listener.

2.2.3.1 Binaural Audio

Recreating the realism of a performance, or environment, for a human listener relies very much on reproducing the physiological factors that influence the ways in which humans perceive sound. Using a single microphone to record a performance featuring many performers, Indonesian gamelan for example, would not effectively capture the ways in which an audience member would experience the sound. The primary factor being that many humans have two ears, which are placed on the periphery of the head and thus have separation in space. Binaural audio is a technique used in recording, and more recently with synthesized audio for conditions with virtual or synthesized sound environments, which tries to effectively capture audio so that a listener using headphones would hear sound as though she was in a particular location.

A basic binaural recording setup might consist of only two microphones set apart at roughly the width of a human head as shown in Figure 2.2. While this somewhat resolves the issue of sound arriving to each microphone at

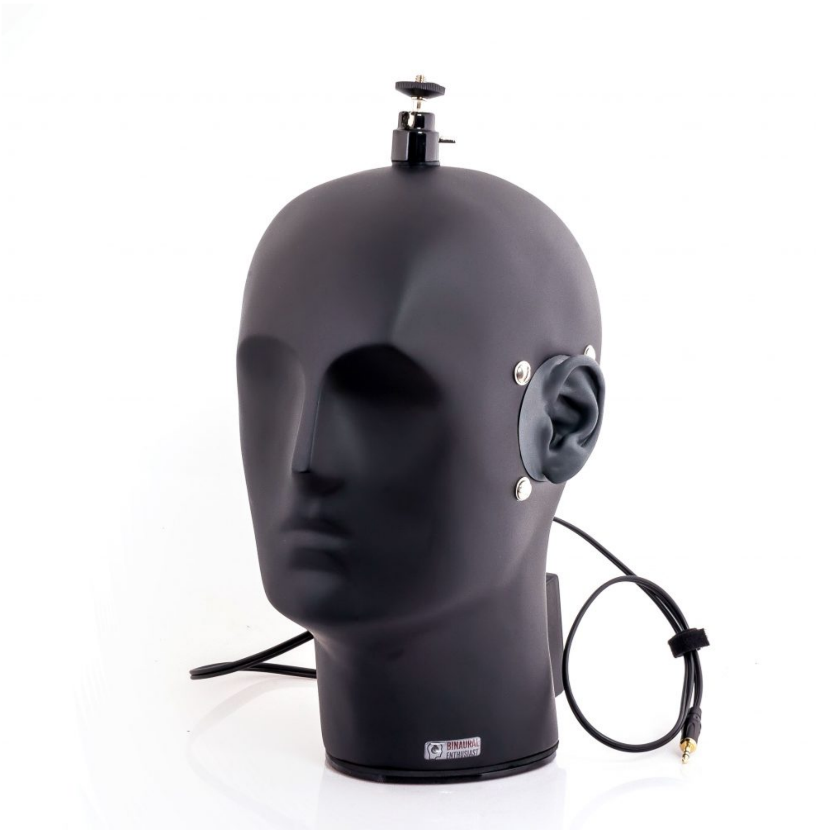


Figure 2.2: An example of a binaural recording unit. Microphones are embedded in the cavity of the ears.

different times just as sounds arrive at our ears at slightly different times (Interaural Time Difference, ITD), it does not take into account the physical construction of the ear, the shape and angle of the external auricle, the resonant frequencies of the skull and inner ear, the differences in how stereocilia within the cochlea respond to that of the microphones used to record, as well as the embodied audition that might involve non-auditory perceptions for adding meaning to an experience. The external physical hurdles can be approached through the use of sculpted auricles and the placement of a skull-like, in shape and density, obstruction between the microphones to emulate the ways that incoming audio is filtered and modulated (Interaural Level Difference, ILD), even using various sizes and shapes to simulate a different listener. Of course this cannot achieve all of the intricacies that each individual listener might experience, and especially not the embodied experience of witnessing a live performance, it does greatly increase a sense of presence when not physically in a space.(Hermann et al. 2011)

For synthetically generated sound sources in a virtual environment, a physical apparatus for emulating human physiological hearing conditions is not an option. Therefore, it is necessary to make use of Head Related Transfer Functions (HRTFs) for creating binaural audio that approaches natural hearing conditions. The above factors related to ITD, ILD, and forward position of a listener are used as parameters for this function, with the potential to model individual HRFTs based on each listener for best results, however an averaged or generalized HRTF is more commonly used for larger audiences. As mentioned before, binaural audio is most effective when used in conjunction with headphones, however it is also possible to play back binaurally recorded, or generated, sound using speaker systems. one of the most common areas to find the use of binaural audio playback using HRTFs is in video games that emphasize spatialized sound for added realism.

2.2.3.2 Ambisonics

When creating virtual experiences that avoid the use of headphones, the problem of modeling the physiological characteristics of a listener is compounded by having to create the impression of a localized sound source across a larger space. While binaural audio attempts to “move” the listener’s ears to an novel environment through the use of headphones, non-headphone techniques must account for the characteristics of the physical space (size, acoustics, etc.) that surrounds the listener. Ambisonics takes the idea of audio panning mentioned above and enables it to distribute in three-dimensions across varying multichannel speaker setups, by decoupling the “encoding” of sound sources into a virtual sound field, or *full-sphere*, from the physical number and placement of speakers. While a many channel panning system, found in *Poème Électronique* or later on the *Acousmonium*, act more as panning instruments in which sound is manually controlled across a large, spatially distributed set of speakers, in ambisonics, a composer or operator places a virtual sound source around a listener and the sound is automatically allocated across spatially distributed speakers so that the performer does not have to “learn” the sound-to-space process with every new setup.

The resolution and ideal listener location of an ambisonic system relies on the number of directional components used to calculate the sound, with 4 channels (sound pressure, x-axis, y-axis- and z-axis) representing the minimal number of channels or *first-order ambisonics*. High-order ambisonics (HOA) expand potential localization resolution, and given the correct number and placement of speakers can achieve effective spatialization of multiple sound sources. Using an ambisonic order of l, a planar setup would require $2l + 1$ speakers, while a spherical arrangement requires an exponential increase in speakers with $(l + 1)^2$ components. This results in high costs to achieve accurate spatialization, and the ability for the listener(s) to move around the space is also limited depending on the resolution of the system.

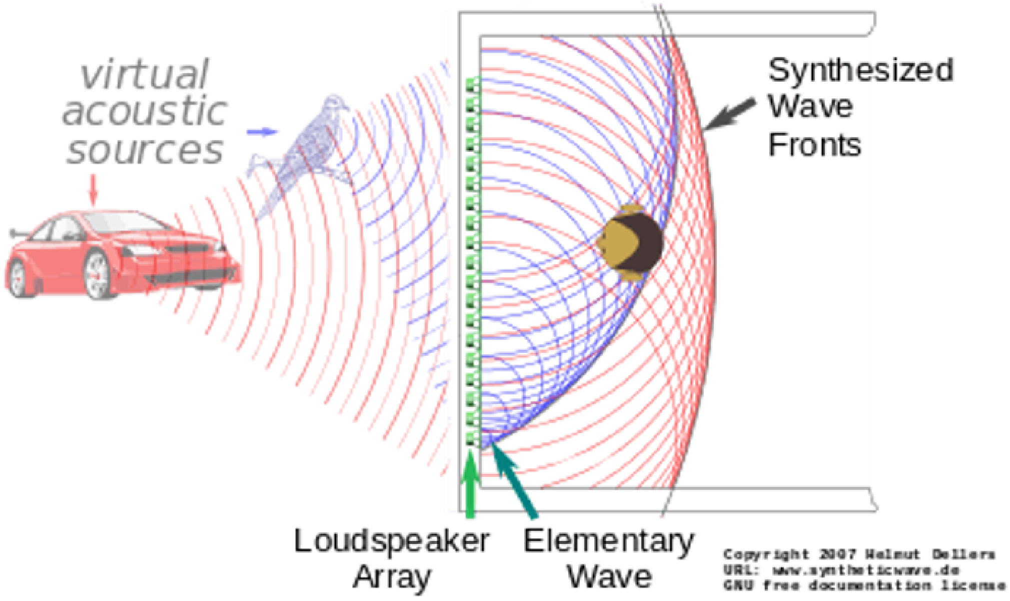

virtual acoustic sources
Synthesized Wave Fronts
Loudspeaker Array
Elementary Wave
Copyright 2007 Helmut Oellers
URL: www.syntheticwave.de
GNU free documentation license

Figure 2.3: Diagram showing how wave field synthesis produces virtual sources through the use of a loudspeaker array.

2.2.3.3 Wave Field Synthesis

This problem of listener position having a large impact on the effectiveness of sound localization, is addressed by another strategy in spatialization known as Wave Field Synthesis (WFS). WFS makes use of a principle that a complex sound can be broken down into its elementary component waves, so by using a large number of closely spaced speakers, a virtualized sound source can be generated by having each speaker activate at the moment the virtual sound would pass through it (Figure 2.3). Although this sidesteps the issue of sound localization for a listener or audience moving about the space, there are still a number of limiting factors related to this audio rendering technique. First, while the number of speakers for a HOA system is already often prohibitive for most spaces, WFS requires an even greater number of speakers, roughly covering the entire surface of the acoustic space. Second, WFS is currently only intended for a planar setup, so that fully surrounding a listener on all sides is not feasible. For the use-case scenario that SIS attempts, it would not be the best option.(J. Berkhout et al. 1993)

The prior technologies for rendering virtual sound sources spatially all made use of traditional loudspeakers. However, for creating spatially immersive sound, a different type of speaker, known as a parametric speaker, was chosen due to a variety of factors. Parametric speakers make use of ultrasonic frequencies to "beam" sound with a much smaller amount of spread than a standard loudspeaker (Yoneyama et al. 1983; Pompei 2019). These ultrasonic carrier frequencies are modulated by sound within the frequency range of human hearing (20-20,000hz) and are broadcast by an array of small ultrasonic transducers. Traveling through the air, the sound is inaudible, but when meeting a solid object (a wall, ceiling, person, etc.) the signal demodulates to emit the audible signal creating the effect that the solid object is the sound source. With multiple speakers, and not nearly as many as HOA or WFS, it is possible to generate highly localized sound sources at different locations in a space; each spatialized sound controlled by a separate speaker. Depending on the number of spatialized audio sources required, parametric speakers can be an effective method for creating a SIS experience; either acting on their own, or supporting non-spatialized audio. Adding directional control over speakers in a system can create moving, localized sound sources, examples of which are described in the following section, which are vital for SIS.

2.3 Prior Work in Spatially Immersive Audio Systems

In the past decade, an increasing number of projects employing parametric speakers have been developed (Ishii et al. 2007; Kuutti et al. 2014; Ochiai et al. 2017). The following section will explore three unique projects that move towards spatially immersive sound by implementing parametric speakers with the ability to change direction.

2.3.1 Vibro-Scape

An early example of a moving ultrasonic speaker for uses approaching SIS applications can be found in the “Vibro-scape Design” project (Watanabe et al. 2007; Watanabe et al. 2006). Designed as a feature of a multi-media dance performance called “Living Lens” which was performed at Brisbane Festival in July of 2006. This performance featured dancers, projections onto a curvilinear screen hanging in the performance space, and the aforementioned MUS arrangement. Sounds emitted from the speakers were restricted to the higher end of the frequency spectrum (for a human listener) and were comprised of synthesized sounds and processed fragments of vocals taken from a choral work.

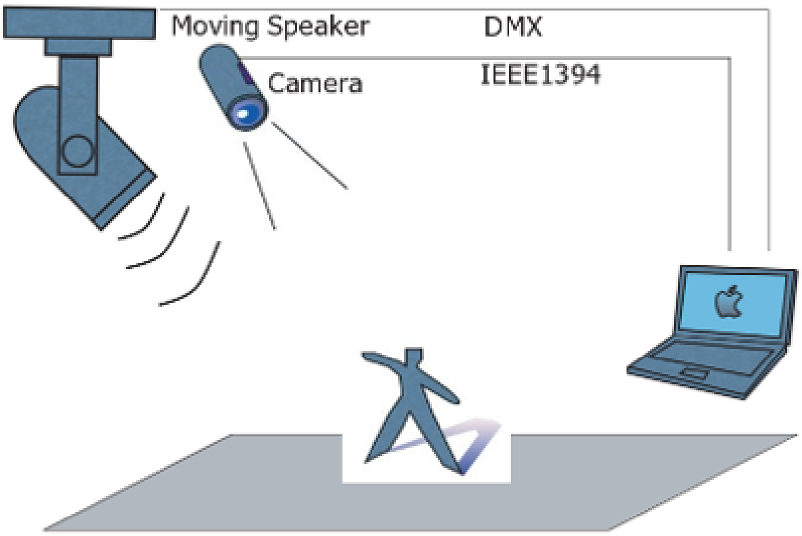


Figure 2.4: A diagram showing the speaker and camera system used for Vibro-scape.

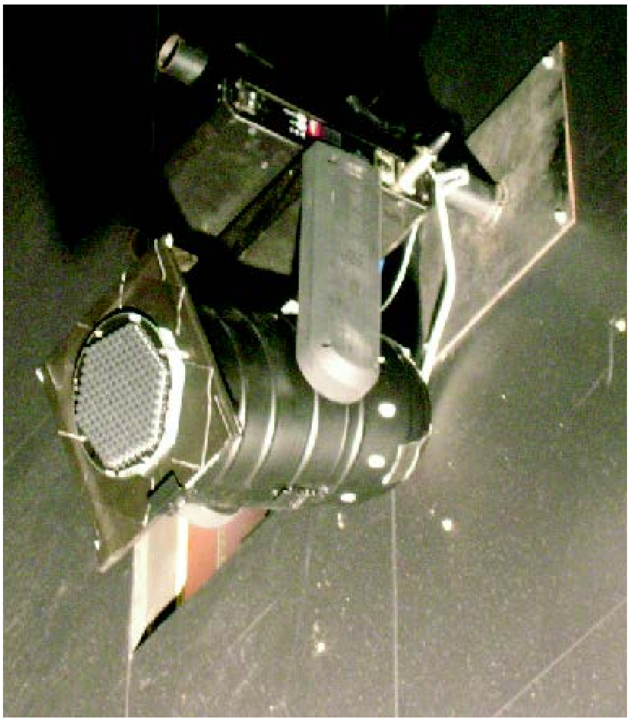


Figure 2.5: The ultrasonic speaker was mounted to the front of a movable lighting system.

The MUS system used an ultrasonic speaker (Mitsubishi Electric Engineering Company Limited Japan, MSP - 10MA) attached to the top of a moving light system shown in figure 2.5 (Visionlight MOVING PAR 201). Mounted on a wall above the performance space, the MUS's position could be controlled through the use of a camera mounted nearby the speaker (figure 2.4). Whether the MUS was manually or automatically positioned based on the live feed from the camera is unclear, however based on written material, there are hints at the possibility of an unspecified tracking system so that the sound could follow the dancers throughout the space (Verdaasdonk 2007). The goal of developing the system was to achieve individualized audience experiences (Watanabe et al. 2006) as well as heightening a sense of immersion by generating analogous audio movement to the dancer's positions in space. The reflection of ultrasonic sound from the curved projection screen or reflection from the dancers creates added sonic effects for movement and visuals.

An additional technique for enhancing immersion was achieved by contrasting low frequency quadraphonic audio with high frequency sounds emitted from the MUS which the artist describes: "As the performers Richard Causer and I-Pin Lin leaped vertically or across the space at alternate moments in time, the sound beams created the illusion of low-flying bats moving in several directions" (Verdaasdonk 2007). The authors explain that through individualized audience experiences, artists can consider more atomic and personalized material in order to create more effective artworks.

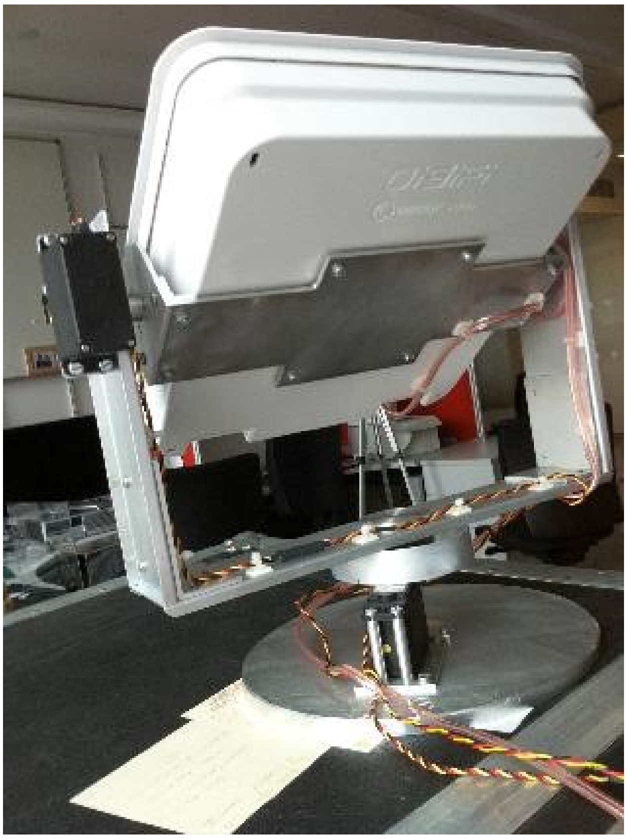


Figure 2.6: The custom built pan tilt gimbal holding the parametric speaker.

2.3.2 Audio-Visual Projection System

Another project that made use of MUS in a very different context, involved enhancing a mixed reality agent audio-visual system (Ehnes 2010). Citing the issue of cognitive dissonance due to mismatch between a precisely position projection of a virtual character and the amorphous ambient audio of the character's voice when emitted through standard speakers (even with the use of 2 and 3 channel systems), the author created a virtual character system which, through the use of a MUS, could emit audio from the location of the projection. Separate pan-tilt gimbals were used for the projector and the parametric speaker (DigiFi Sonicast S100-50) with the MUS consisting of a custom gimbal built using servos and a microcontroller to send PWM signals (figure 2.6). Custom software for modeling the physical room environment was also built to facilitate proper rendering of the virtual character. This software offered a very simple modeling tool, users manually input vertices of planar surfaces which can then be connected to form lines and triangles, however the primary purpose of the software is to assist with the rendering and movement of the projected virtual character (figure 2.7).

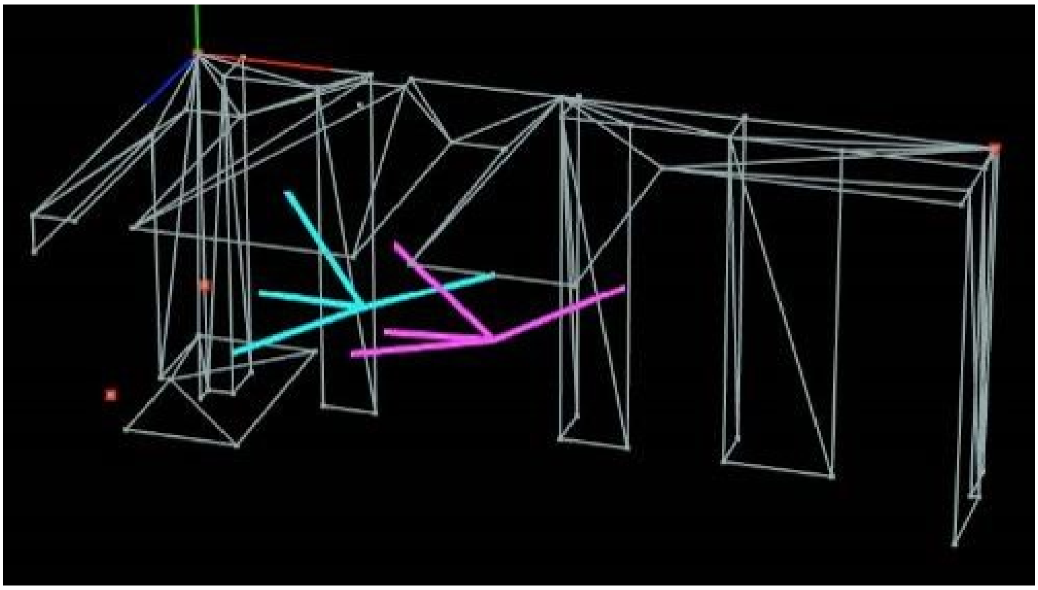


Figure 2.7: Software for modeling the physical room. The speaker and projector are depicted by the purple and blue lines, respectively.

Ultimately, the spatialization of the sound simply followed the calculated direction of the projector rather than move independently, which supported the application of the MUS in the overall system; creating a more realistic dialogue partner with sound localized to the position of the partner. As the project focused more on the visual portion of the simulated character, difficulties regarding sound spatialization still remained. This is especially the case in situations where unintended sound reflections could cause the sound to emanate from an incorrect position. In this project the immersive nature of the experience stems from a more physically accurate portrayal of a virtual character in space. Though the imagery of the character itself purposely avoids realism, the visual accuracy in projection and the spatial accuracy of the sound were important to achieve this goal.

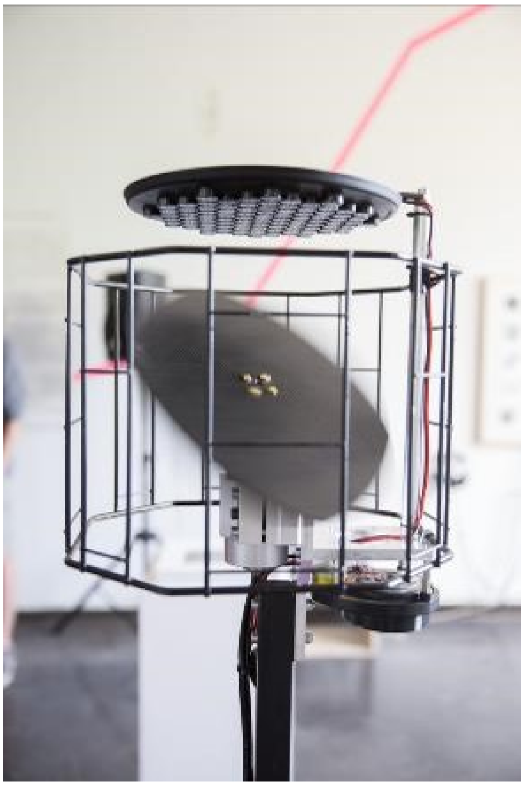


Figure 2.8: The spinning reflector allows audio to be positioned horizontally about the unit.

2.3.3 Pol 2

A third system making us of a MUS differs from the strategy used by the previously described works: using a pan and tilt gimbal to position the speaker in the desired direction of audio projection. *Polyus* (Wernicke n.d.b) and later *POL II* (Wernicke n.d.a) are constructed in a way that the parametric speaker(s) are fixed in place while a moving element positions the sounds through reflection. These devices are mechanically similar to the Leslie speaker, developed by Donald Leslie in the 1930s to enhance the sound of electric organs, which used rotating parts to modify the timbre of sound. Another source of influence could be the rotating, reflective functionality of LIDAR scanning systems. Both projects use a carefully tuned spinning element to reflect sounds produced from the parametric speakers to different positions in a space (figure 2.8). Especially relevant is that by synchronizing pulses of sound and the rotation of the reflection element, the device is able to maintain the position of a sound in space. This enables the positioning of multiple sound points in space all emanating from a single speaker; a kind of spatialized polyphony.

This ingenious technique for extracting multiple spatialized sounds from a single speaker is a clever way to circumvent a number of limitations attached to parametric speakers. An added feature is the ability to dynamically add and position sounds around the device through the use of a touchscreen application. While there is no indication of automatic mapping to different spaces or integration with tracking or automated movement systems, there is support for receiving OSC messages which could be relayed from such systems. Multiple spatialized sound sources improve the immersive nature of the sound device, however the positional placement of the sound in a particular space is limited to the horizontal plane of the rotating reflector. As a sculptural, kinetic object, the optimal position for the function of the device, the center of a room, works aesthetically. This placement limitation could be detrimental for an immersive space where participants are expected to move about, or the device might occlude the visual projections in an audio-visual immersive environment.

2.4 Conclusion

A historical understanding of the immersive qualities of sound and the methods, techniques, and technologies aimed at enhancing or inventing existing and novel techniques for immersive sound has provided context for distinguishing spatially immersive Sound from other forms of immersive sound. Technologies like binaural audio, high order ambisonics, and wavefield synthesis support localized and spatialized sounds while providing, to varying degrees, accurate and convincing virtual audio sources. However they often restrict the movement of the user due to extra requirements in wearables, number of users, space restrictions and acoustic properties of a space, and often prohibitive costs.

Spatially Immersive Sound systems are designed to minimize restrictions on a listener's ability to move about and explore an immersive space, while also providing an immersive sound experience with localized sound sources by using parametric speaker technology. Developments in recent systems have introduced movement to command the direction of a parametric ultrasonic speaker (MUS) with uses in following and enhancing the visual and movement qualities of live performance, accurate virtual character speech spatialization, and a multiple source sound installation device. Each of these works provides different aspects of spatially immersive sound: positioned virtual sound sources, freedom of motion for the listener, and software for managing the location and movement of the sound.

It is the unification and extension of these elements that the spatially immersive sound system created for the BSD project attempts to achieve. The next chapter will give a detailed account of the system, its iterations, and functional effectiveness.

Chapter 3

Spatially Immersive Sound System

Spatially immersive sound enables audio to be beamed to precise positions, the sound itself only manifesting at the moment of collision with a material object. Using ultrasonic speakers to transmit high frequency audio, sound can travel with a minimal amount of dispersion or spread, reflecting off of surfaces to give the impression of a false sound source. Networks of spatially immersive audio devices can create patchworks of localized sound fields within a space; individuals can move between shrouds of private sonic worlds. Here, a system is presented which adds movement to spatially immersive sound devices by combining directional speakers with pan-tilt mechanisms. This addition enables new possibilities for distributed localized sound that can be repositioned based on atmospheric conditions, sounds that constantly move based on swarm or cellular simulations, or personalized sounds that whisper secrets and lies into your ears from afar.

3.1 Prior Iterations

Detailed descriptions of the various iterations of the BSD project have been discussed above, but it is important to isolate the parametric speaker aspect of each version in order to reflect on the changes that the system has undergone over the evolution of the project. Of particular interest is the performance that took place in the Large Space in 2016 as it was the final version of the project to make use of fixed position parametric speakers and spurred the development of a more dynamic system.

Starting with the first outdoor version of BSD, the parametric speakers were mounted outdoors. Making use of the reflective nature of the leaves of the particular species of tree, natural movement of the wind through the branches created more dynamic movement while also connecting conceptually with the source material. Following this initial installation, all further versions of the parametric speaker system have been installed indoors. The next major installation took place on Governor's Island in New York. In the battery of a decommissioned military fort, the speakers were mounted around the upper ceiling of the central room and along the hallway leading into the room. Fixed at key locations that participants would pass through, the effect of transmitting bird song through the directional speakers was to contrast with the surroundings of a cold, empty room.

The Large Space was the first time that an installation of BSD that was done where the intended use of the space was for immersive experiences. Because of this, it was possible to more directly integrate the audio and visual aspects of the project, which, due to constraints in earlier exhibitions was much more difficult. Where the previous exhibitions were in spaces that inevitably became part of the installation, the Large Space offered more of a blank canvas in which all aspects of the project could be accommodated. As the sound design and visual flocking patterns were created by the same group of researchers, it followed that the these two aspects could be connected so that the audio could strengthen the movement of the flocking simulation. While the connection between the two proved successful, the speakers were not able to smoothly follow the flocking simulation. This limitation of the parametric speakers was a direct inspiration for the SIS system developed for the next two iterations of the project.

One other variation on the moving parametric speaker that was introduced between BSD installations, was the construction of a field research tool. Needed for field research on the disruptive effects of bird song in nesting areas, a hand held parametric speaker was developed for use with portable audio players. The battery powered device could be brought on research trips in order to project bird song from a distance without having to mount equipment in the canopy (figure 3.1).

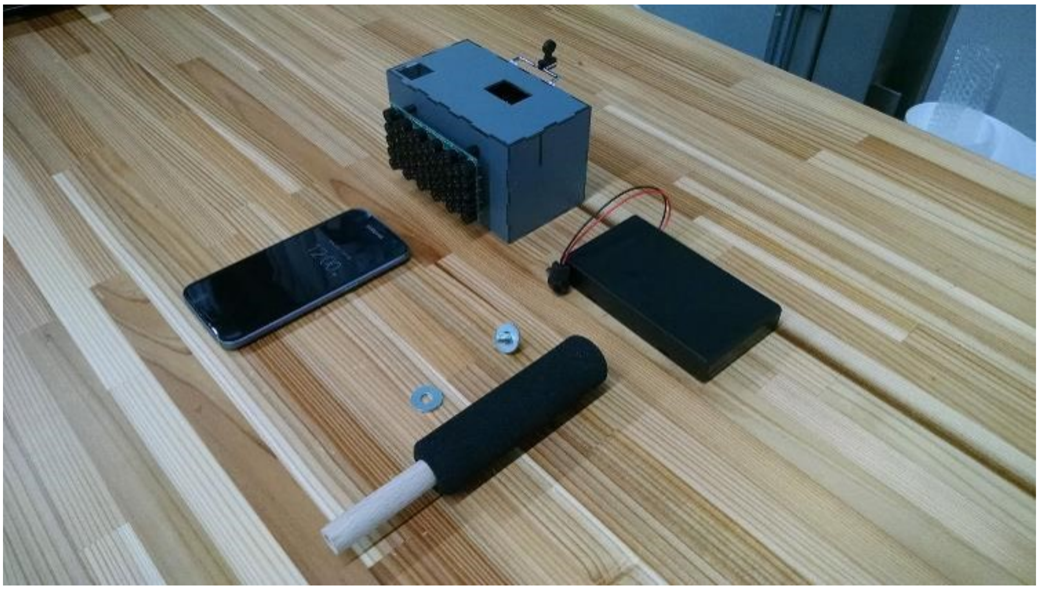


Figure 3.1: A disassembled image of the handheld unit showing the speaker and amplifier box, battery pack, handle, and smartphone for playing audio.

3.2 System Overview

The audio elements for Deep Space 8k were adapted from those used during the LargeSpace installation. Translating the sound design and hardware from the LargeSpace involved scaling down the number of sound sources as well as drastically improving the flexibility of the hardware. The reduction in size of the Deep Space 8k meant that the number of directional speakers could be reduced while still covering the majority of the space. Additionally, the acoustic properties of the Deep Space 8k, especially the projection surfaces, enabled the sound to reflect about the room to a much greater extent. A further difference was the added ability to control the direction of the speakers by mounting each individual speaker to a dedicated pan-tilt mechanism, allowing custom static or dynamic speaker configurations. This enabled fewer speakers to match the sonic capabilities of many speakers in a fixed configuration. With these factors in mind, the number of directional speakers was reduced from twenty to eight. Deep Space 8k also provided a powerful 5.1 audio system that was used with stereo audio to produce a more dynamic range of audio.

The sound emitted from the parametric speakers was a generative composition created in the visual programming language Max. The program assigned each of the 8 speakers a different bird song, and over time changed out different songs or silenced the speaker. The repetitions of the calls changed depending on the selected song.

Additional sound was played over the Deep Space 8k speaker system. The composition was built from accumulated sounds of unmanned aerial vehicles (UAVs or drones), in particular were consumer drones which bear the names of birds (Parrot, Global Hawk, Grey Eagle, etc.). As the performance progressed, the intensity and amplitude of the drones increased until overpowering all other sounds in the space until finally dissipating at the conclusion.

Recreating the effect of multiple shifting sound sources that added to the sonic effect of the outdoor installations for the indoor Deep Space 8k installation required a mechanical system to give programmatic control of the direction that each speaker faced at any given time. Each speaker was attached to a servo controlled pan-tilt mechanism, which in turn was attached to the ceiling of Deep Space 8k. This allowed the parametric speakers to point at specified locations within Deep Space or to move based on simulations of natural processes.

For the installation in Deep Space 8k, the speakers were assigned to individual boids in the small scale live simulation. The directional speakers would then move about the space with the simulated paper cranes and drones to recreate the constantly shifting audio that featured in the previous outdoor installations. Additionally, by being attached to a simulation that the audience could influence, additional aspects of the audio, location and movement, became part of the entire simulation of the event. A system for mimicking natural processes should consider the ways that all aspects of the system can support a particular behavior. While previous installations used only fixed position directional audio, it was the addition of a dynamic position system that connected the audio with the visual in a very direct way.

After the Deep Space 8k performance, it was clear that the installation and maintenance of the system were too intensive for the mobility of the piece, especially because of the variety of locations and spaces that the project must adapt. A review of the existing elements of the system and ways to improve upon it was conducted with the aim of easier installation, maintenance, and complexity.

One of the major problems from the Deep Space system was the control system wiring. The audio cables were cheap, light, and able to run long distances without concern for signal loss. This was not the case for the control cables. As distances became longer than 15 meters, the strength of the PWM signal diminished to the point of extreme jittering from the servo motors. Additionally, extra time was lost to cable installation and finding the optimal location for the control box. Other hurdles that were also related to the servo controller included limitations on the number of pan-tilt units due to using a centralized power supply. The existing 105W power supply would not be able to support additional pan-tilt units if needed in future iterations, further the Lynxmotion controller needed its own power supply and would require a second power supply for any extra servos. Finally, the Lynxmotion controller used a wired connection to the computer which added complications to the installation.

Based on these issues, it was decided to explore wireless solutions to sending control messages. Taking into account size and power limitations, the ESP8266 WiFi enabled microcontroller, as part of the ESP-01 module, was selected for testing. Each pan-tilt unit would now have its own microcontroller that would directly receive control messages over a wireless network. This eliminated the problems of cable length restrictions related to signal loss, weight of transporting cables, time installing extra cables, and the computer needing a wired connection to the servo control box. The problem of hardware limitations related to different numbers of pan-tilt units was now only limited by the router's limit on network traffic. Without the ethernet cables, each individual unit needed a separate power supply, which resolved problems with the centralized power supply, but also added the difficulty of needing to run power to each pan-tilt unit.

3.2.1 Hardware Design

3.2.1.1 Wired System (Deep Space 8k)

The directional speakers used for Deep Space 8k, and throughout the BSD project, were parametric speakers sold as kits by the TriState company (TriState 2015). These same speakers were used in prior iterations of BSD for creating the impression of many individual sound sources without having to use the large number of speakers required in other spatialized sound techniques, e.g. wave field synthesis (J. Berkhout et al. 1993). While the frequency response of the parametric speakers is limited to 0.4 – 5kHz, the speakers are able to reproduce bird song very effectively as the vocalizations used in the installation fell within this range.

Due to the small size and weight of the parametric speakers, the system could be built using inexpensive RC hobby servo motors. The parametric speaker, pan tilt mount, and electronics were mounted to a laser-cut acrylic plate. Power was transmitted, along with pulse width modulation (PWM) signals controlling the angle of each servo, through a single Category 7 (CAT-7) networking cable. Off-the-shelf cables and RJ45 connectors were chosen as the signal transmission medium as they could be easily found in a variety of lengths and easily replaced in case of failure. Signals were transmitted on three twisted pairs contained in the networking cable to reduce signal interference over the distance from the control box to the pan-tilt units. Testing with a variety of cable lengths revealed that up to 15m distances could be traversed with no effect on servo performance. The Lynxmotion SSC-32u USB servo Controller Board, built around the Atmel ATmega328, was used to send control signals to the 16 servos across the 8 pan tilt units. A separate power supply built from a Unitek 105W 21A 10-Port USB Charger Station provided separate 5V power to each pair of servos per pan-tilt unit. The PWM signal and power were combined in a custom designed RJ45 adapter board.

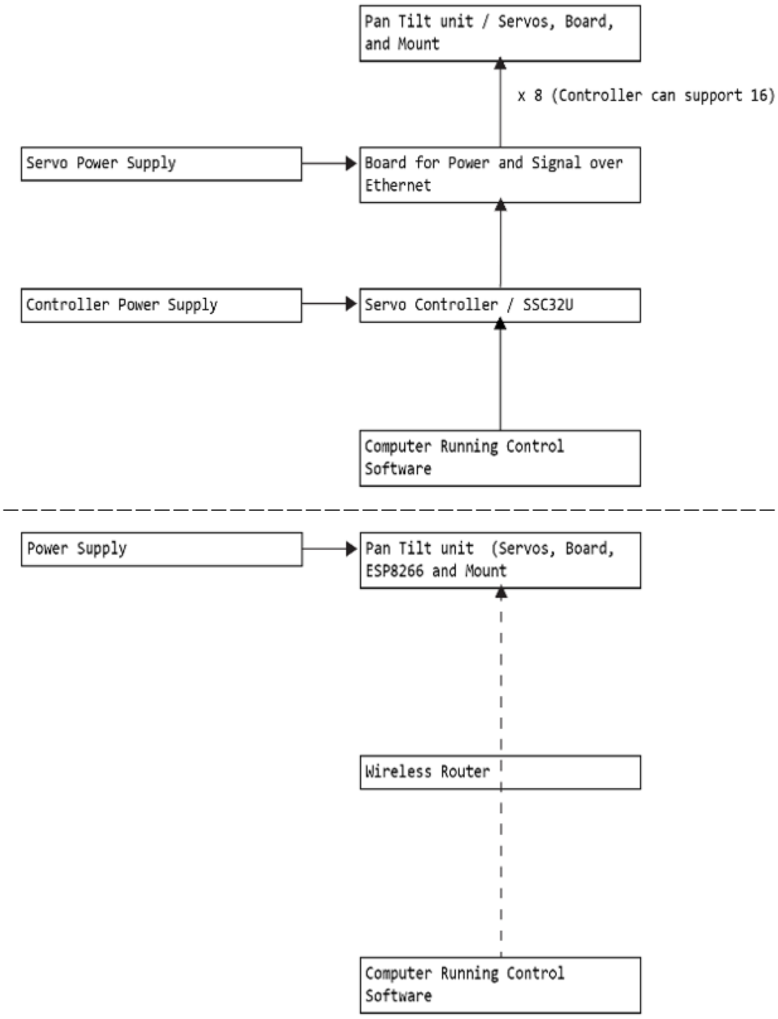

Pan Tilt unit / Servos, Board, and Mount
x 8 (Controller can support 16)
Servo Power Supply
Board for Power and Signal over Ethernet
Controller Power Supply
Servo Controller / SSC32U
Computer Running Control Software
Power Supply
Pan Tilt unit (Servos, Board, ESP8266 and Mount
Wireless Router
Computer Running Control Software

Figure 3.2: A comparison of the components and connections in the wired system (*left*) and the wireless system (*right*).

3.2.1.2 Wireless System

The shift to a wireless system meant that much of the custom hardware built for the wired system would need to be replaced or altogether discarded. The control box, which housed the Lynxmotion servo controller and power supply, the servo power supply, and the custom adaptor boards for combining the controller and power supply for transmission over ethernet were no longer needed. All of the steps between the computer sending control messages and the pan-tilt units could be replaced with a single wireless router. The rest of the system including computer, parametric speakers, pan-tilt servos, and mounting hardware remained the same as the wireless system.

On the pan-tilt units themselves, the custom PCB for receiving the ethernet cable and converting to separate pan and tilt cables was replaced by another custom board adapted for the ESP-01 module. This new board needed to receive 5V power for the servos while also stepping the power down to 3.3V for the ESP module. An 8-socket header was attached to receive the microcontroller module pins. In addition to the voltage regulator, two pulldown resistors were attached to GPIO0 and GPIO2 to keep the microcontroller from starting in flash mode. The two 3-pin headers allowed the pan and tilt servos to connect to the board. Figure 3.2 shows a comparison between the two systems.

3.2.2 Network Structure

For the wired system, the Lynxmotion servo controller handled the routing of messages to servos based on a protocol developed by the company (Lynxmotion 2018). The computer's control software sent values serially to the Lynxmotion, which passed them through the converter board to the respective ethernet cable. Though the setup of the wired system was an arduous process, the connectivity and response of the various components was very fast and reliable.

The ESP8266 microcontroller comes with support for WiFi (802.11 b/g/n) and can function as a station, access point, or both simultaneously (Espressif 2018). Depending on the module, the device can come with a variety of firmware that aid in software development and are often tailored for different use cases. Originally, the wireless system used the NodeMCU firmware (NodeMCU 2018) which allowed custom firmware builds to save memory by only installing needed modules. Another feature was the use of Lua scripting with an asynchronous, event-driven programming model (similar to Node.js) (LabLua 2018; Node.js 2018).

After powering on, each module would connect to a common access point and receive a static IP address identifying the pan-tilt unit. Once a module successfully connected to the access point, it would initialize a server to listen for and then parse UDP messages coming from the computer's control software. Each UDP message contained three values: the mode, a pan value, and a tilt value.

The mode was added to allow for calibration of the servos after it was found that, from servo to servo, the same value would result in slightly different positions. A user would manually positioning the pan-tilt unit to face straight ahead and parallel to the ground in "absolute positioning" mode, then send a "write" mode message to write the current pan and tilt values to memory as the "home" position for that pan-tilt. The default and normal operating mode is the "relative positioning" mode, in which the pan and tilt values are offset based on the "home" position. Pan and tilt values were simply angles between 0 and 180 degrees which would be converted by the server to the correct pulse values.

Testing with NodeMCU worked well initially, however it was soon determined that the underlying firmware was not able to receive UDP messages quickly enough for smooth, continuous control. When sending values from the computer, the servos would move in quick, discreet steps that would cause the speaker to bounce. An attempt at adding an easing function to reduce jittering proved fruitless, and an alternative firmware was pursued. Eventually, it was discovered that there was an Arduino compatible firmware for the ESP8266 (Forum 2019), and after testing out continuous control over a pan from 0-180 degrees and back it was clear that the movement was much smoother. The sane functionality as the NodeMCU firmware was reproduced for the Arduino firmware, however using native modules/libraries when available.

A comparison between firmware (NodeMCU and Arduino) of the visual movement and accelerometer data was done for the same unit after being instructed to move only the pan servo from 0 to 180 and back to 0 degrees. Using pixel tracking to follow the same point on each pan-tilt unit throughout the a video (1920x1080 FHD at 30FPS) of the servo movement, a plot of the x and y values over time displays a visual comparison between each unit's movement is shown in figure 3.3. While both position graphs appear similar, another plot was made to show the positional difference between each frame of video for both units. This graph (figure 3.4) though the movements are not perfectly aligned, shows a clear difference between the two movements as the NodeMCU displays much noisier and erratic motion than the Arduino firmware. Accelerometer data was also recorded at a sampling rate of 400 Hz. Figure 3.5 shows more intensity of vibration across the frequency spectrum for the NodeMCU firmware, confirming the an improvement in vibration after changing firmware.

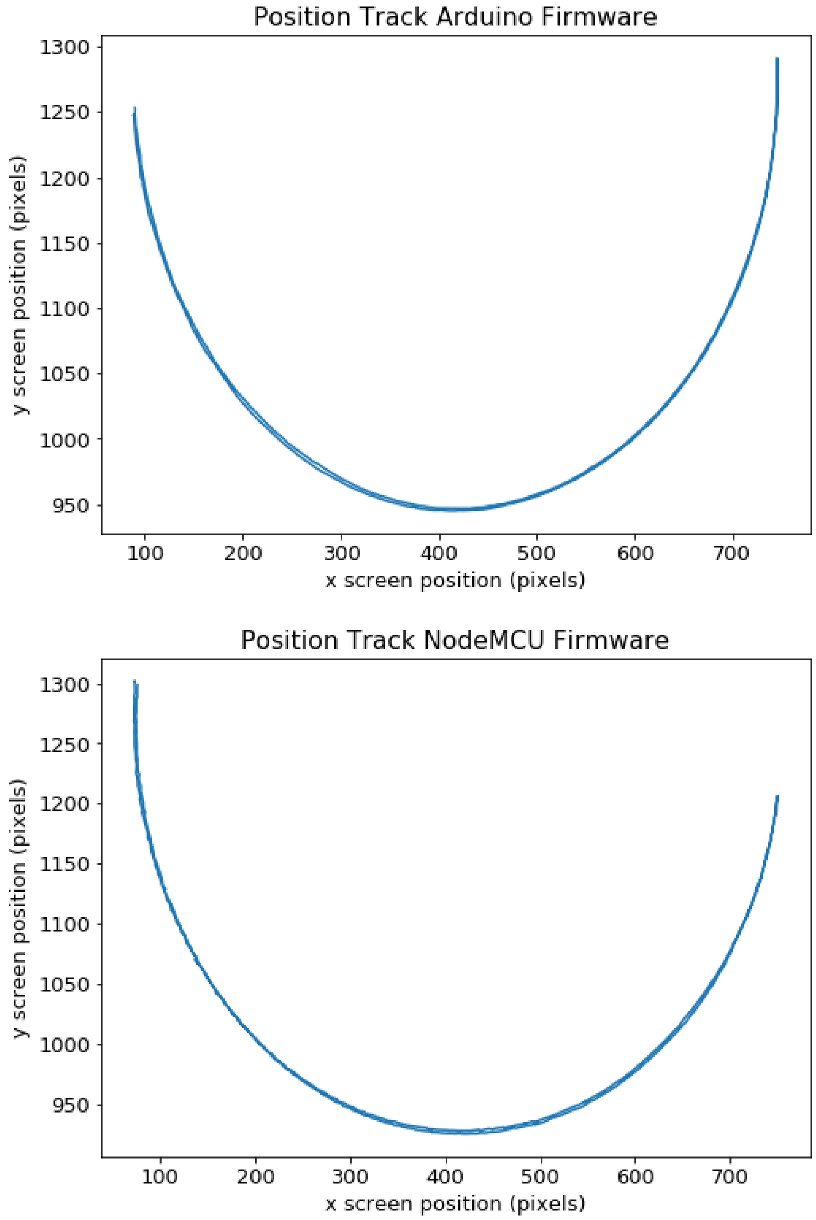


Figure 3.3: Graph showing the tracked x and y positions for the two firmware.

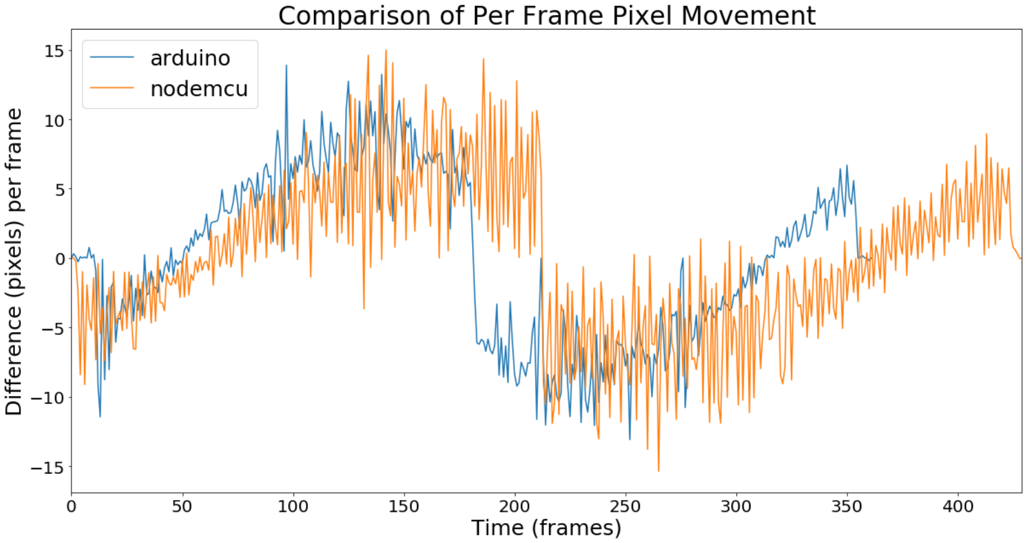


Figure 3.4: Graph comparing the frame by frame changes in movement, combined x and y, of the NodeMCU firmware (orange) and the Arduino firmware (blue).

3.2.3 Control System

The servo control board that facilitated servo movement, for both wired and wireless versions of the system, was in turn commanded by a computer running software that would facilitate the movement of the speakers and communication with other aspects of the BSD project.

When developing the software for the control interface it was important to consider a number of factors related to the BSD project and the development of sound movements. Reviewing the history of BSD shows a wide variety of installation spaces, both outdoor and indoor, small and large, so the ability to quickly mock up and visualizer an upcoming installation space was critical. Because additional components of BSD also occupied the same installation space, being able to display the physical movement of the speakers and the audio projection path within the installation space would give designers and artists clarity when trying to understand how the SIS system integrated with the overall installation. A further consideration came from the variation in collaborators for BSD, people might join or assist with the development of one exhibition and would have dramatically different backgrounds than people who had worked on other exhibitions. Therefore, it was necessary to make the control software intuitive and primarily a GUI where most or all of

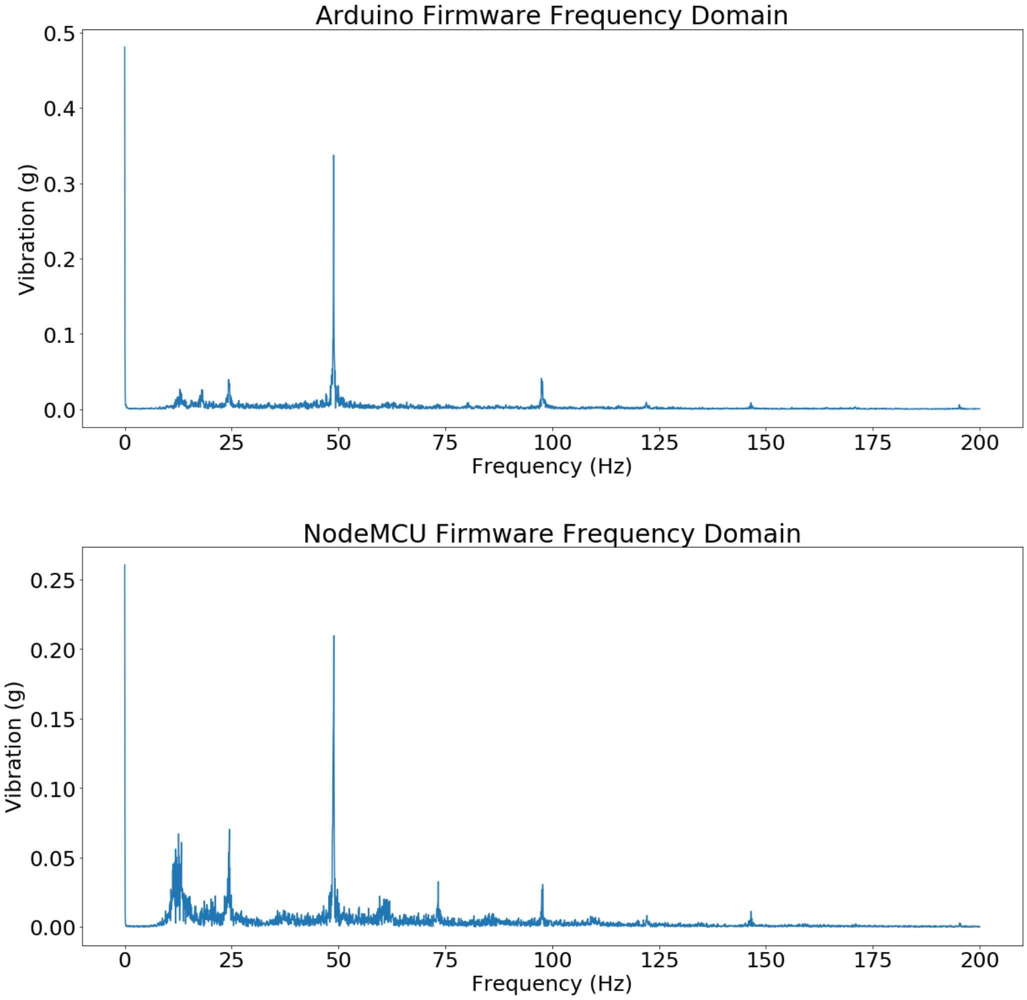


Figure 3.5: Comparison of the frequency of vibrations between the two firmware.

the movement design could happen without the need to code. Finally, due to the variety in installation spaces mentioned above and the desire to adapt BSD to work with the tools and features of each space, the software needed to easily interface with plugins, CAD and 3D modeling formats, communication protocols, and hardware systems for tracking or lighting.

With these requirements in mind, it was decided to build upon an existing, and moderately familiar, software rather than designing an entirely custom platform. The Unity (Technologies 2018) game engine and editor was selected due to its cross-platform support, existing prominence in the game development community, and also its growing popularity for use in visual and art installations. Creating a virtual representation for the physical installation would be eased by Unity's existing compatibility with a large number of formats for 3D models and in cases where a model of the space was not available, constructing low a low resolution mock-up of the space using the software's built-in primitives was also a possibility. In situations where working in the physical space was not possible, which was usually the case for BSD, having a virtual model of the space allowed for preliminary troubleshooting and testing which expedited the process of setup and installation after arriving at the location.

Unity offers a large store of assets and plugins, making the process of connecting with tracking systems or integrating simulation algorithms easier. When faced with the need to modify or write a custom object, users can create scripts in C#, which uses .NET 4.6 scripting runtime, that can attach to existing objects in a scene to incorporate new functionality (Wagner 2018). Because C# is a scripting language not limited to Unity, it is often possible to find existing C# libraries that will work with Unity, occasionally needing small modifications or recompilation using the proper frameworks. This allows users who are more familiar with Unity and scripting to develop more sophisticated solutions, while not preventing non-scripting from creating movements or positioning elements.

The control software is a collection of scripts, objects, and scenes that translate the angular rotation of virtual speakers to pan and tilt values for the corresponding physical unit. The plugin offers a number of scripts to communicate with different physical controllers, which have been added over time as the overall system evolved. Currently the plugin supports serial connections to the Arduino and Lynxmotion SSC-32U microcontrollers and can also send UDP packets to network addresses (here used with the ESP8266). The plugin uses Game Objects to create connections between the components of the system, these are: endpoint device objects (mention above), virtual speaker objects, and virtual target objects. Figure 3.6 shows a screen shot of a scene.

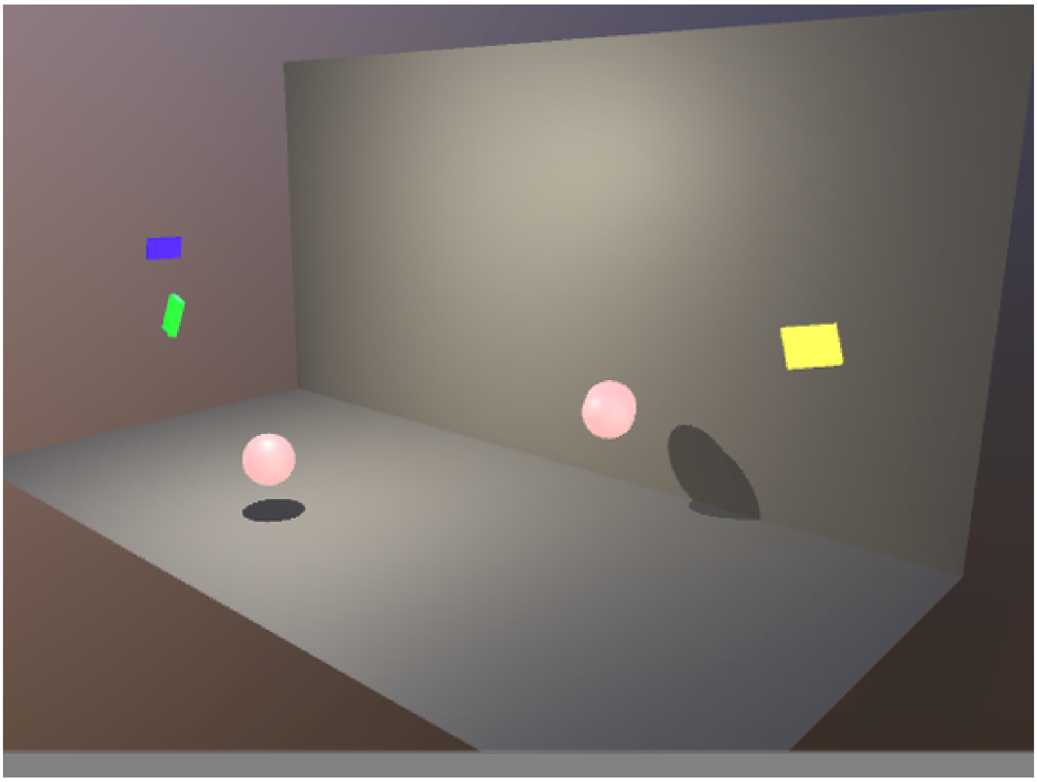


Figure 3.6: A screenshot of a Unity scene using the plugin. The scene is modeled after the Deep Space 8k space and features three virtual speakers and three virtual targets.

3.2.3.1 Unity Scene Configuration

For example, after creating a virtual version of the Deep Space 8k, the plugin let users place virtual speakers into the space, corresponding with the mounted locations of the physical speakers. The user could then “connect” the virtual speakers to the physical speakers by assigning the respective numerical value indicated on the physical network box. Once the virtual and physical speakers were connected, users could control the physical speakers in a number of different ways. While possible to manually change the rotation of the speakers, it is much easier to use the plugin’s targeting system to instruct the speaker where to aim. Any virtual object can be assigned as a speaker’s target and the speaker will always attempt to aim at that object. The physical speaker will only move as far as its range of motion allows, so limits of movement can be assigned in the software to prevent out of range movement. A target can be manually moved, animated, attached to physics simulations, or connected to another object’s motion which is often the case with a tracking system.

3.2.3.2 Integration of Tracking Systems

The support for tracking systems was one of the main reason's for choosing the Unity game engine as the basis for the SIS system software controller. To test this, two commonly used devices for tracking were integrated into the plugin and tested with a human participant set as the target of the pan-tilt speaker. The first device, a Kinect v2 (Microsoft 2016), was tested by placing the Kinect v2 in a room with two pan-tilt speakers. Wanting to make use of the Kinect 2's HD face detection and tracking feature, the existing Kinect plugin for Unity needed to be updated for the V2's SDK. After this was done, it was possible for Unity to receive up to 1000 facial points from the Kinect rather than only the 26 points of the skeletal pose (Microsoft 2014). The HD Face API uses semantically named constants to represent common positions along the tracked face, therefore the tip of the nose (`HighDetailFacePoints_NoseTip`), the inner corner of the left eye (`HighDetailFacePoints_LefteyeInnercorner`), or the edge of the jaw (`HighDetailFacePoints_LowerjawRightend`) can be selected to have the speaker aim at a specific location on a person's face by assigning the position of the face track vertex to the position of a speaker target object. For the test, the tip of the subject's nose was assigned to a target that was shared by both of the virtual speaker objects. Figure 3.7 shows the screen of the unity scene in which the two virtual speakers (blue and green boxes) are pointing in the direction of the tracked face of the subject (yellow points). See supplementary materials for a video demonstration of this system.

The second tracking system to be combined with the control software, is the OptiTrack system installed in the LargeSpace virtual reality system. The previous installation in the LargeSpace made use of the 20 unit OptiTrack Prime 41 system to have visual elements react to participant movement. OptiTrack makes a plugin that allows tracking data, live or recorded, to be streamed over a network from OptiTrack's Motive software (OptiTrack 2018). After adding the OptiTrack plugin to Unity, a test in which a speaker, mounted to the top of the LargeSpace's wall, would follow the position of

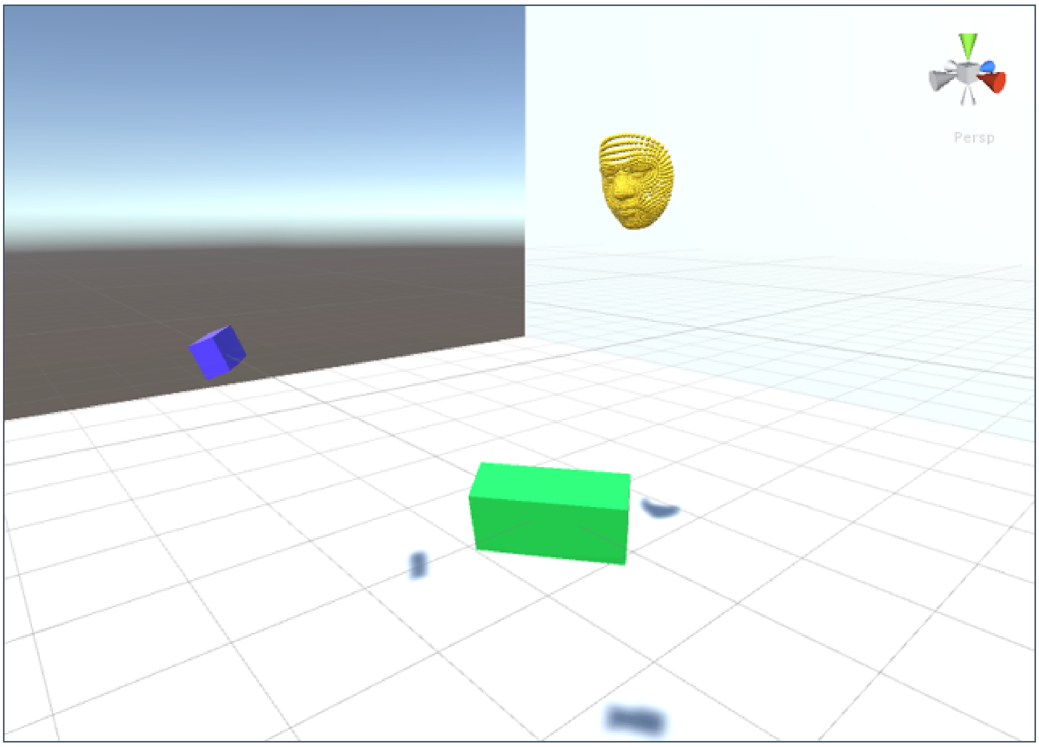


Figure 3.7: Demonstration of the Kinect v2 HD Face tracking system with the spatially immersive sound unity plugin. Both speakers are aimed at the nose of the tracked person.

a moving subject. The OptiTrack system was calibrated and set to track a pair of glasses worn by the subject, the position of which would be passed to Unity which then assigned this point to the target of a virtual speaker (and in turn the physical speaker). A video of this test can be found in the supplementary materials. It should be noted that the test was made using a wireless system with NodeMCU firmware, which lead to a lot of stuttering in the motion.

3.2.4 Calibration

Mentioned above in the description of the wireless system, was the addition of a calibration system to ensure that all of the pan-tilt units were not out of alignment. The software side of the calibration initially began as a python script with a CUI interface and keyboard commands to manually adjust the pan and tilt servos independently. When properly aligned, another keypress sent the current values to be stored in the microcontroller's memory.

The Unity plugin recreates the exact same functionality, though instead of keyboard movements, provides a GUI interface with buttons that allow manual adjustments.

After connecting the software to the physical pan-tilt units and calibrating them, a number of simple tests were done to demonstrate the functionality of the system. Videos of the tests can be found in the supplementary materials.

3.3 Implementation in LargeSpace

As mentioned previously, the LargeSpace offered BSD a much wider variety of possibilities due to the space's CAVE-like qualities, multi-person tracking system, and motion-base movement system. At the same time, this installation occurred before the SIS system which allowed for the parametric speakers to vary in direction. The solution to covering the area of the LargeSpace while using only fixed position speakers meant that a larger number of speakers had to be placed around the top edge of the LargeSpace.

3.3.1 Large Space Installation

In order to determine the number of parametric speakers to fully cover the floor surface, it was necessary to calculate the size of spread of the parametric speakers, given their mounted position along the top of the space. The floor of the LargeSpace is 20m x 15m with vertical walls at a height of 7.7m (Takatori et al. 2016). For the parametric speakers, the maximum angle of sound dissipation is 20 degrees. Based on this information, it is possible to calculate the diameter of spread s at distance d from the source, which has an angle of dissipation a using:

$$(1)\ \ s = 2d\ \tan(\tfrac{1}{2}a)$$

With a minimum distance of 7.7m, that is if a speaker was facing straight down, the diameter of spread would be 2.715m. A maximum distance of

30.154m, diagonally from the top corner to the far bottom corner, results in a diameter of 10.52m. More likely, a speaker will be aimed at the space immediately in front of it, so while useful to know the maximum spread for the LargeSpace, a distance of 30m is not a realistic value. Instead, a distance of 15m was used, angled toward the center, giving a spread of 5.289m. Based on spread sizes between 3-5m, it was determined that the floor of the large space could be divided into a 5 by 4 grid of 20 regions, each region comprised of a 3.75m x 5m rectangle.(figures 3.8 and 3.9)) This determined that 20 parametric speakers would be needed in order to cover a majority of the floor of the LargeSpace.

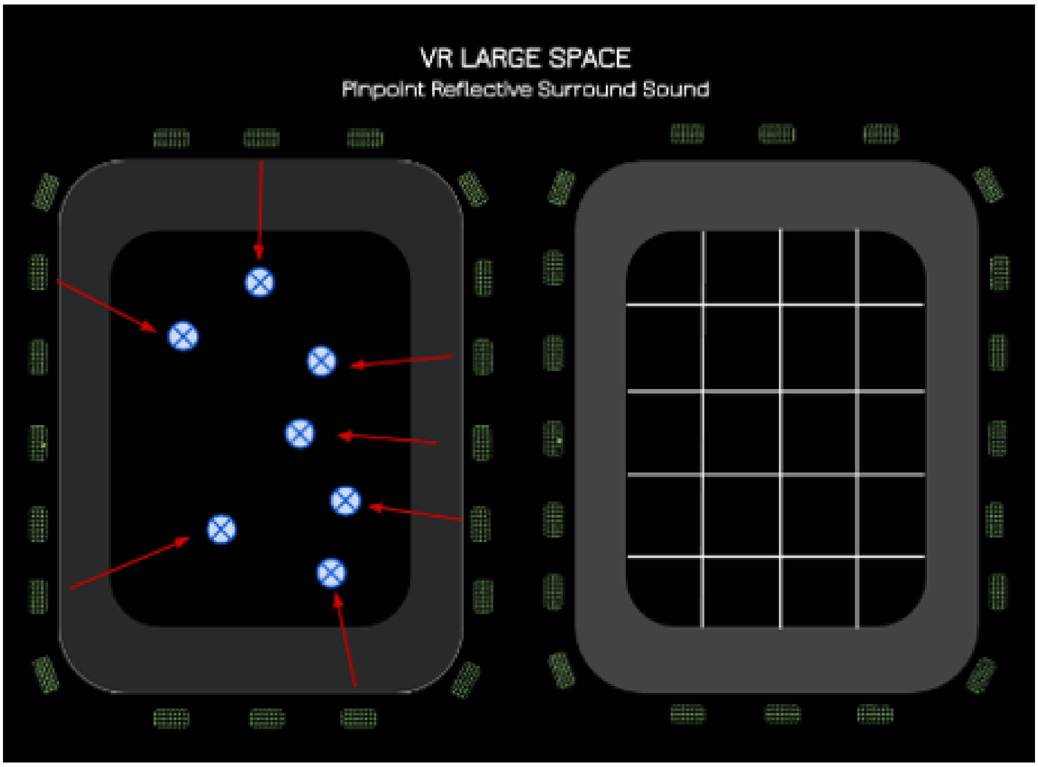


Figure 3.8: 20 square grid used in the LargeSpace performance.

During the installation, to give the impression of sound movement the audio amplitude of each region was assigned to a virtual grid running an ALife simulation of boid agents.(figure 3.10) As a boid moved between regions of the grid, the amplitude level at each region would be calculated by the boid's proximity to the center of each region. This process created a technique very similar to the that of multichannel panning described in **Chapter 2**.

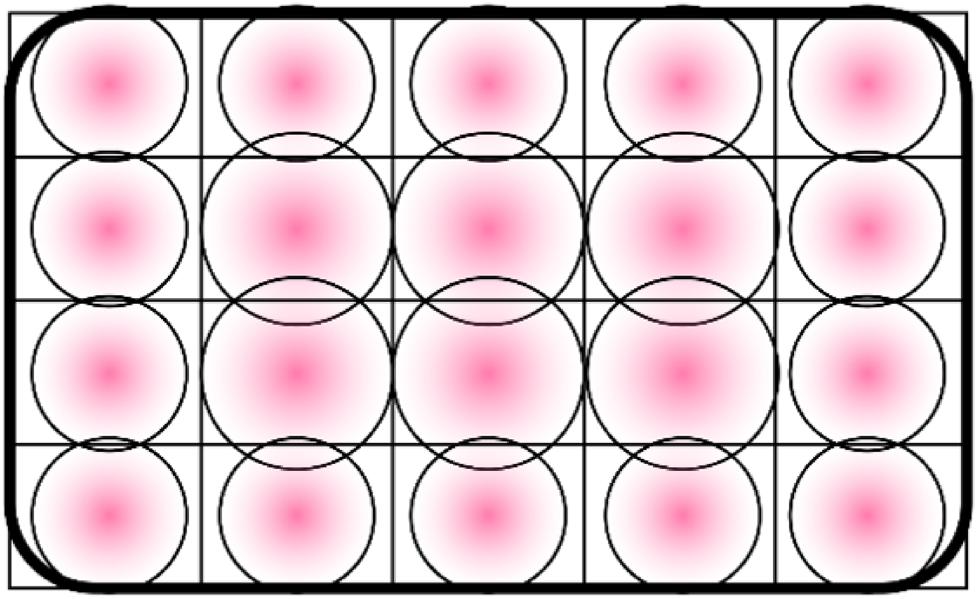
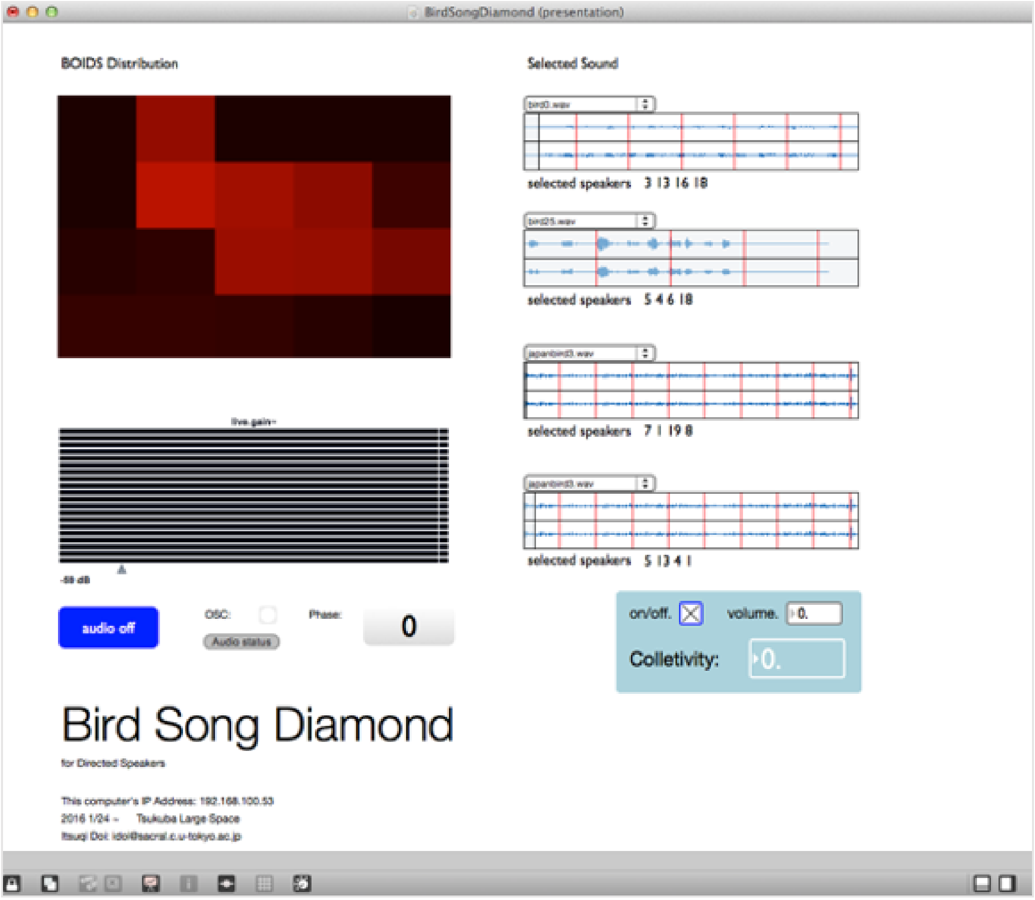


Figure 3.9: 5 by 4 grid overlaid with the calculated spread of 20 speakers mounted evenly around the top of the LargeSpace and aimed at the center of each grid region. Notice that spread increases when speakers are aimed closer to the center of the LargeSpace.

3.3.2 Single Unit Movement Data

Before implementing moving speakers in the LargeSpace, a comparison was made regarding the audio output of a fixed and moving speaker. Within the LargeSpace, two experiments were done, both involving a parametric speaker and a microphone. The first experiment involved a fixed position parametric speaker, aimed at 90 degrees pan and tilted downward at a 30 degree angle. A microphone was moved along an arc of radius 3m from the speaker, from 0 to 180 degrees in increments of 7.5 degrees. At each position, the amplitude of the microphone was recorded, resulting in 25 recordings as the microphone moved around the fixed speaker.

The second experiment fixed the microphone at a distance of 3m from the speaker at the angular position of 90 degrees. In this version the parametric speaker was able to pan from 0 to 180 degrees while maintaining a downward tilt of 30 degrees. Because the rotation of the speaker could be automated, whereas the movement of the microphone had to be manually measured each time, the audio amplitude at the microphone was recorded every 1 degree



selected speakers 3 13 16 18

selected speakers 5 4 6 18

selected speakers 7 1 19 8

selected speakers 5 13 4 1

volume. 0.

Colletivity: 0.

Bird Song Diamond

for Directed Speakers

This computer's IP Address: 192.168.100.53
2016 1/24 - Tsukuba Large Space
Itsuqi Doi: idoi@sacral.c.u-tokyo.ac.jp

Figure 3.10: The Max software used in the LargeSpace. The upper left corner shows the grid with amplitude values represented by color.

change in angular position of the speaker resulting in 181 recordings.

A side by side comparison of the experiments 1 and 2 is shown in figure 3.12. In both plots, the amplitude level has been normalized to more fully show the change of amplitude with change in the angular direction (here also normalized).

Because the moving parametric speaker allowed for automated recording at each individual angular position, a third experiment was conducted the chart the amplitude of a fixed microphone as the speaker panned from 0 to 180 degrees as well as tilted across it's entire range. The resulting graph in figure 3.13 shows the amplitude for each recording along the z-axis. There were 7,381 positions recorded in total, creating a kind of audio scan of the floor of the LargeSpace. The highest amplitude value corresponds to the

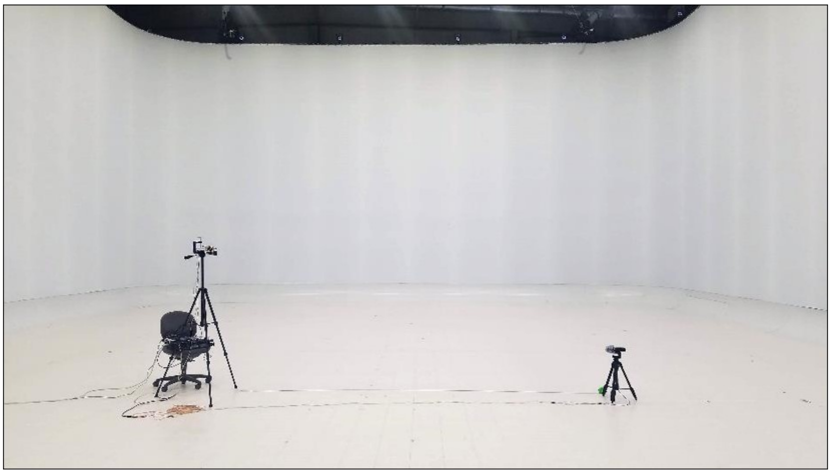


Figure 3.11: Experimental setup for comparing fixed and moving speaker.

location of the microphone with a similar panning falloff curve as the previous two experiments. The tilt falloff extends farther, potentially indicating the reflection of audio from the floor to the microphone. A similar artifact can be seen along the $y = 0.8$, in which the speaker was angled almost directly at the floor. A dual reflection, floor and ceiling, seems to be creating this artifact. This indicates that although the directionality of the parametric speaker is good at localizing an audio source, reflected audio can create lower volume artifacts at other regions in the space.

3.4 Implementation in CCD, Mexico City

For the SIS system, the installation happening in the Memorial space at the Centro de Cultura Digital (CCD) would be a large change from the first SIS test in the Deep Space. Aside from the physical differences in the space (**Chapter 1**), the largest difference was based in the presentation of BSD. The Deep Space performance occurred during a festival in which many different projects all shared the same theatre space, meaning that the BSD performance only lasted for about 30 minutes and the experience was mostly directed. For CCD Memorial, a long term installation in a space which BSD did not share with any other projects, visitors might come alone or in a group and could spend as much time as they like in the space without the aid of the artist or docent to describe the work.

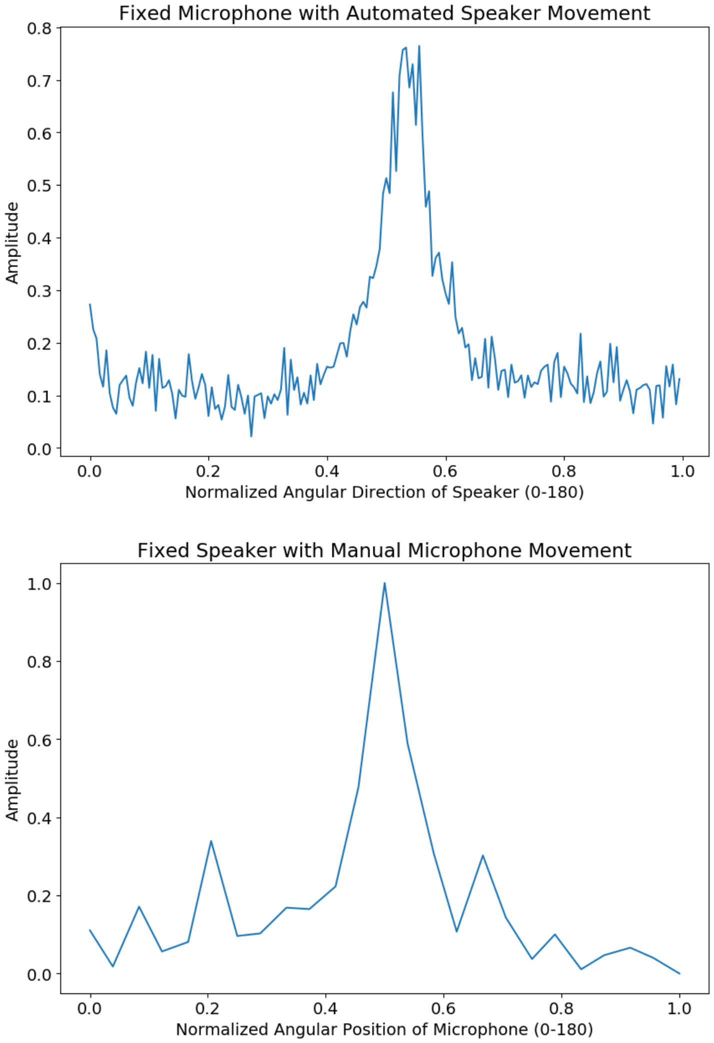


Figure 3.12: Comparison of fixed and moving speakers

3.4.1 CCD Memorial Installation

As mentioned in **Chapter 1**, the long duration of the installation meant that the SIS system could only stay for the first few days and then return to the LargeSpace for further tests. After receiving the layout of the space (figure 3.14) and receiving news that we would be unable to mount the speakers to any walls, a plan for mounting 8 parametric speakers, one for each column in the space, to the inner side of each column. Being that there were 8 standard loudspeakers, which would be producing sound, already mounted to the outer walls of Memorial, placing the parametric speakers on the inner column area would create a distinct change in sound between the outer and

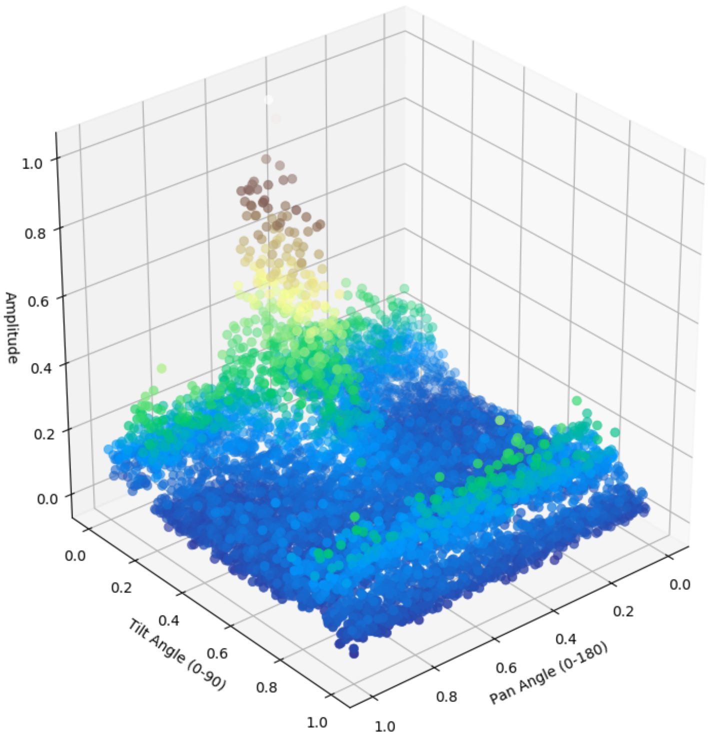


Figure 3.13: Using the pan, tilt, and amplitude values (x,y, and z respectively) to create a 3-dimensional map of the space.

inner sections of memorial. Upon arrival, it was clear that the speakers were not able to be mounted to the columns at all, however there were two existing wires running across the internal corridor of the two inner pairs of columns. After determining that the wires were sturdy enough so as not to wobble from the movement of the speakers, a new layout in which four pairs of speaker units hung across the center of the corridor was chosen (figure 3.15).

A further consideration was the 3m tall ceiling, which was quite low in comparison to the LargeSpace and Deep Space 8k. Based on the same method used for the LargeSpace, the spread for each speaker was estimated to be between 1m to 1.3m. Meaning that with 8 speakers, the sound could cover, at the least, 56% of the area of the corridor and, at most, 94% of the corridor area. A major difference from the LargeSpace is that each speaker was no longer fixed, and could therefore cover a much larger area of the corridor

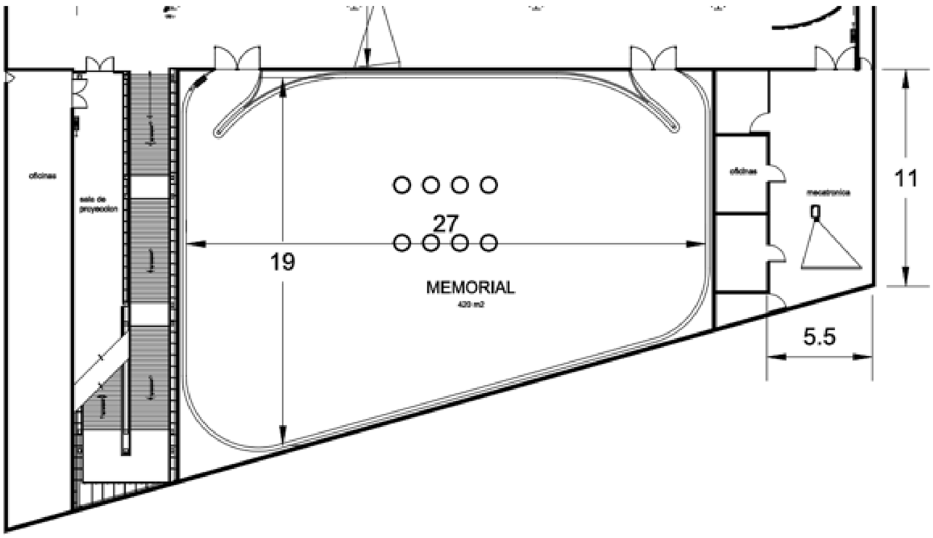


Figure 3.14: Floorplan of the Memorial space of CCD.

and areas of the outer section.

3.4.2 Speaker Movement Paths

Speaker positions in the initial layout allowed for each individual speaker almost full coverage of the interior corridor with small areas between some of the columns unreachable be the audio (figure 3.16). The change in speaker positions, hung from the interior wires rather than against the columns, significantly reduced the amount of coverage for each individual speaker. Only when considered as a pair (figure 3.17) is it possible to reach the majority of the inner corridor. An added problem to hanging the speakers above the center of the corridor, as opposed to the edge of the corridor, is that two bands of unreachable sections stretch across where each speaker is unable to turn any farther. This is alleviated by when considering the spread angle of each speaker which closes this audio gap after about 1m. A remaining problem is that when assigning separate sound sources for each of the 8 speakers, there will be areas behind each speaker which a sound will never reach. This is especially true for the speakers closer to the columns. Facing this situation, if full coverage of a space is necessary, assigning individual sounds to each pair of speakers for four total sound sources is the solution. The sound would

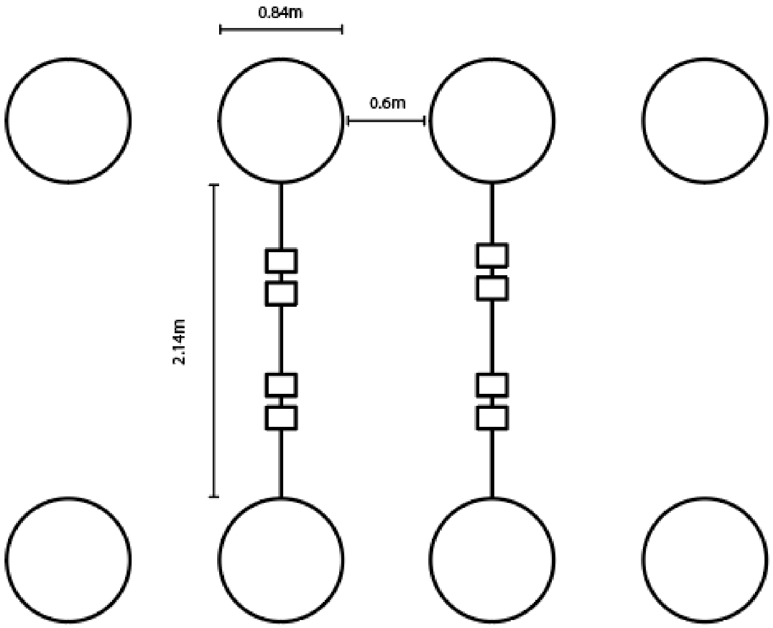


Figure 3.15: Close up view of central columns in Memorial, with the positions of the eight pan-tilt units suspended on wires between the inner two columns.

need to be passed between the two speakers as the target position moved from one speaker's coverage area and into the other speaker's side.

3.4.3 Integration with the Overall Exhibition

In the **Chapter 1** overview, it was mentioned that the movement of the speakers followed a boid simulation in which the target object for each speaker was an individual boid. This created a coordinated movement where some of the speakers moved to similar positions, while others might peel off onto a separate path. Even when the targets objects moved to the outer region beyond the columns of the Memorial space, the sound could reflect off of the column and prevent too much of the sound from leaking to the outer section. The sound in the outer region consisted of soft, synthetic sounds that slowly moved between the eight outer speakers. The synthetic sounds of the exterior were chosen to contrast with the natural sounds of birdsong produced by the moving parametric speakers. At the border between the outer and inner soundscapes, located at the two entry points to the corridor

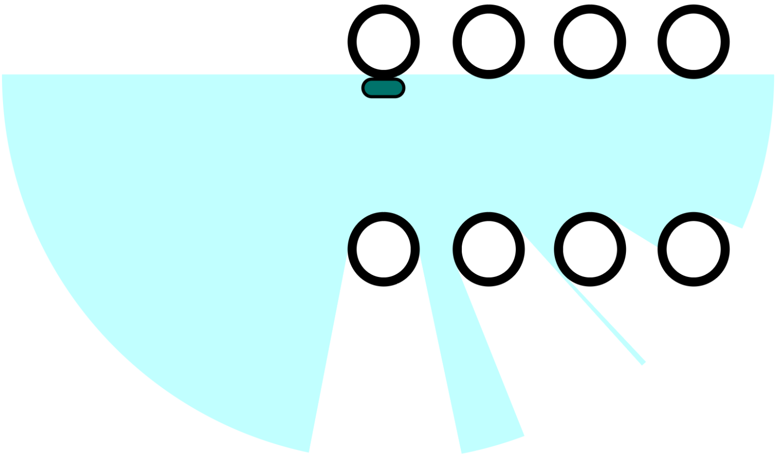


Figure 3.16: Coverage of a single speaker mounted to the column and facing inward. In this position, nearly the entire inner area can be reached by each speaker.

of columns, was the two stations of the BSD Mimic system. People participating in the Mimic system were positioned right at the boundary of the two sound worlds.

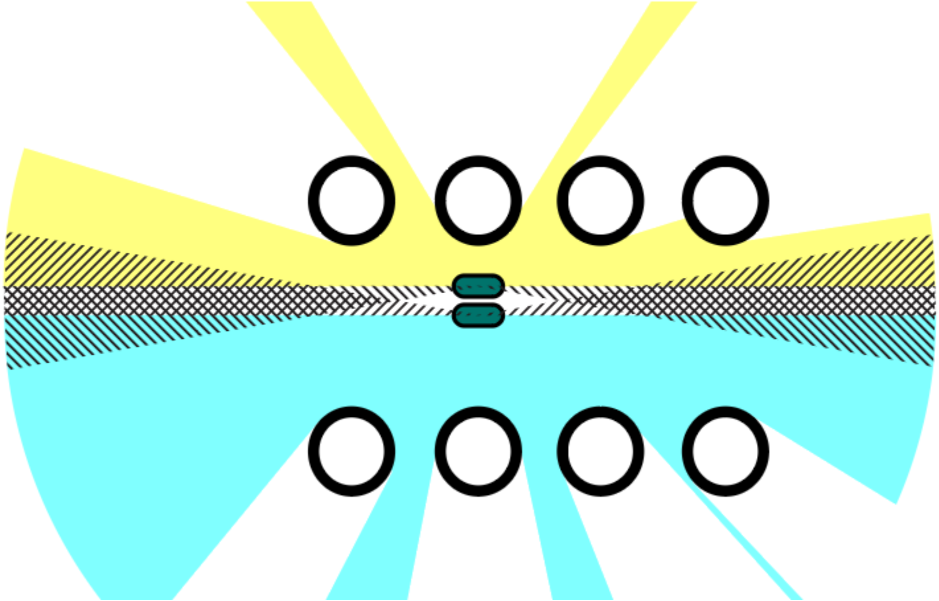


Figure 3.17: Coverage of a pair of speakers in the final positions used for the installation. Each color represents one speaker's line of sight. The black striped area represents extra coverage due to the 20 degree angular spread of the parametric speakers which reduces the amount of unreachable area.

Chapter 4

User Tests

Participant observations and anecdotal accounts of user experience had been verbally discussed between artists, collaborators, and participants over the course of the previous exhibitions. Generally, a successful installation was based on the level of interaction, time spent in the exhibition space, and attitudes of participants post-experience. From the perspective of an artist, the success of a work depends a lot on the imaginative, creative, or affective potential that a work instills upon those who engage with it. Additionally, the ability to attract users so that engagement with the work occurs in the first place, is another concern. The previous iterations all acted as experiments determining the effectiveness of BSD for both of these areas, so each time the work was installed, adjustments could be made to further the work. This iterative design process led to artistic insights regarding the beneficial nature of collaborative interaction for promoting audience engagement with the subject matter. Technical improvement of the BSD software, based on observations and user feedback, required researchers to consider alternative applications for their research technology.

Up until the exhibitions at the Deep Space 8k and CCD Memorial, the iterative design process had not taken a quantitative approach. The discussions surrounding improvements to the system were based on the observations of the artists involved. With prior expertise and knowledge regarding interactive installation design, the artists were able to make adjustments as needed.

With the recent exhibitions, attempts at incorporating quantitative methods from other fields by considering which aspects of BSD could provide fruitful information that could be used to improve part, or all, of the BSD system. By isolating different aspects of the system for analysis, particularly the SIS system, we hope to identify the ability of different facets of BSD for the empowerment of participants.

Finally, due to a lack of prior studies on movement of localized sound sources using parametric speakers, it was necessary to run experiments outside of the BSD installation. Removing the moving speaker from the context of the installation meant that structured tasks, isolated to the moving parametric speakers, could be performed by subjects without influence from other aspects of the BSD project influencing the subject's perception. Two experiments were carried out. The first involved subjects attempting to determine the direction of a single moving sound source, and the second was a constant stimuli psychometric test to consider the visual dominance of localized sound sources generated by a single parametric speaker.

4.1 Ars Electronica, Deep Space 2017

The series of performances at the Deep Space 8k was also the first attempt at collecting data on user interaction with BSD. The Deep Space 8k during the Ars Electronica Festival had to accommodate a wide variety of performances—from children's programs, to live performers, and a number of different physical elements including the BSD Mimic stands. Due to these constraints, BSD could only occupy the performance space for a total of 40 minutes per day, however with time required for setup and removal of equipment between 20-30 minutes remained for the performance. This meant for a limited number of participants who needed to be instructed on how to interact with the system, which is quite different from standard art installations in which visitors explore the work at their own pace. Because many audience members were in attendance and each performance was monitored by team members, there were no problems in finding people to participate and further instruction or guidance could be administered in the event of any confusion.

The BSD Mimic system was the focal point of the Deep Space 8k performance, and while the audio and visuals maintained their immersive quality, they mainly acted as a backdrop for the mimic performance. Improving from previous iterations of BSD Mimic, including the boids as part of the visual feedback added an additional aspect to connect the audience members with the immersive projection space. From a data collection standpoint, the Mimic system already calculated a score for each participant's attempt at mimicking a particular bird song. Therefore, the scores for all participants over the course of the performance were recorded in order to evaluate participant abilities to mimic various bird species' songs.

4.1.1 Mimic Data

Recording the scores of 211 mimic attempts over four days of performances (figure 4.1), the overall mean accuracy was 68.16%. Separating the attempts based on species of the bird yielded a wider range of results, with certain species such as the Song Thrush receiving consistently high accuracy percentages and others, Japanese Bush Warbler and Eastern Crowned Leaf Warbler, receiving much lower scores (figure 4.2). A variety of factors could account for the variation between species of birds including audience familiarity with only certain species of birds, acoustic distance of a bird song from natural human vocalization due to physiological differences, or simply differences in quality of bird song recordings.

4.2 CCD Mexico City 2018

The time between the Deep Space 8k performance and the CCD Memorial installation allowed for further reflection on the proper methods for evaluating user engagement and empowerment. A survey in hybrid research methodologies gave insights into potential ways to create a greater understanding of user engagement in an environment as uncontrolled as an art installation (Brown & Juhlin 2015; Lazar et al. 2017). As this would be the first real

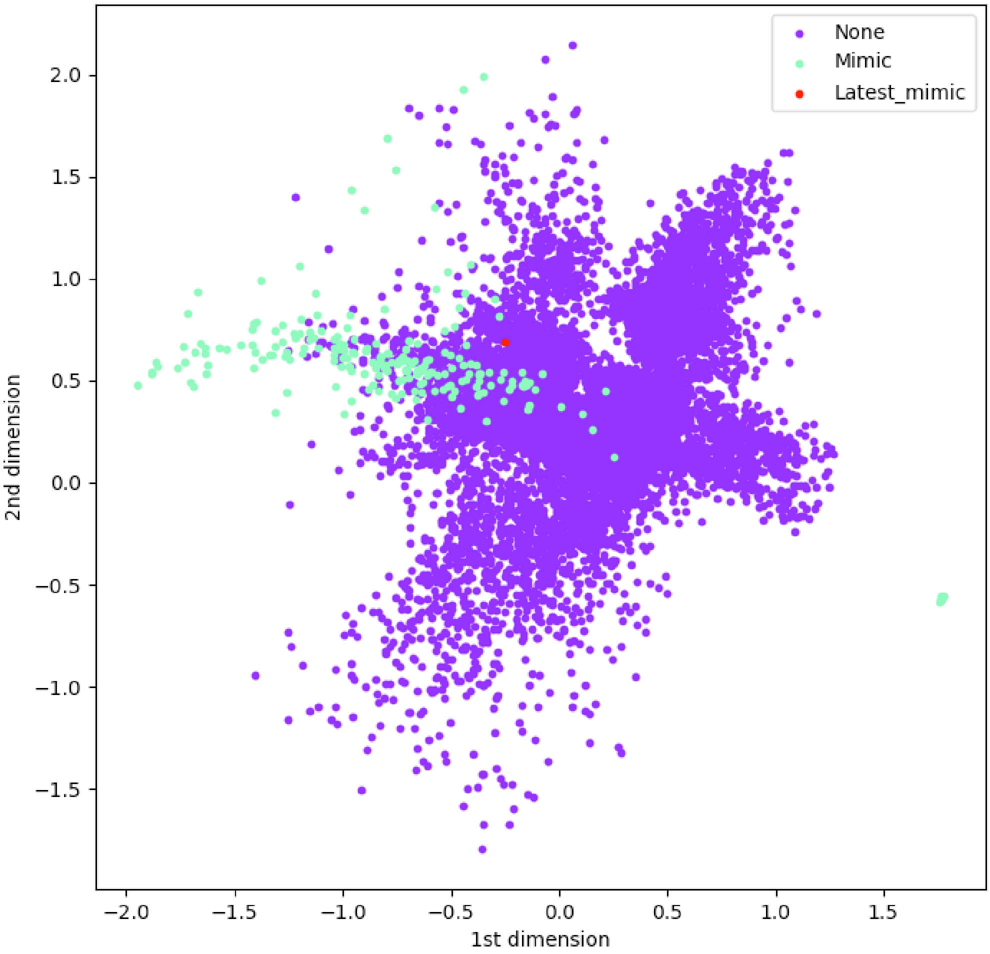


Figure 4.1: Scatterplot showing the two dimensions of the trained model (Dark Purple) superimposed with the Mimic attempts from the Deep Space 8k performance (Light Green/Blue)

opportunity for evaluating the dynamic SIS system in the context of an art installation, it was important to attempt to isolate the impact that the SIS system had on users. Therefore, a combination of surveys with both closed and open-ended questions, interviews with participants, recorded speaker movement values, and video recording of user interaction was prepared for the exhibition at the CCD. In recording video of user behavior, a comparison could be made with video recorded of the previous LargeSpace installation to compare the differences in behavior between an iteration with a fixed SIS system (LargeSpace) and one with a dynamic system (CCD).

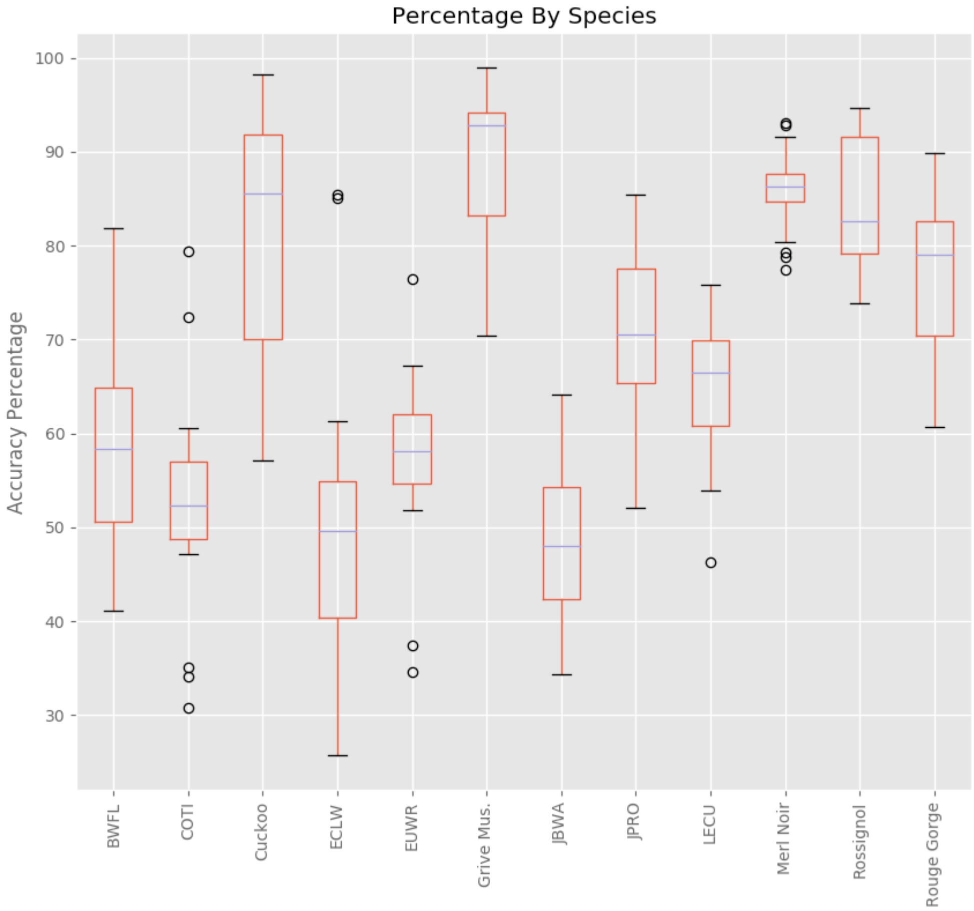


Figure 4.2: Box plot of the mimic percentages separated by species of bird. Unabbreviated species names can be found under additional information.

4.2.1 Paper Surveys

One of the most important methods for considering the effectiveness of the BSD installation in CCD was to receive feedback directly from those who had experienced the installation. Not limited to only those who directly interacted with the work or received explanation of the function and artistic intent, the surveys aimed to gather a broad cross-section of participants to more fully understand affective potential of the installation. Each survey contained nine scaled questions that asked respondents to what extent, from strongly disagree (1) to strongly agree (5), they agreed with a series of statements. The statements themselves focused on three aspects of the installation: recognition of the spatialized sounds, connection between the audience and the installation, and the connection between the thematic aspects of the installation and spatialized sound. The final question was an optional written response in which participants could write any further impressions or comments about the installation. In all, there were eighteen respondents who filled out the nine scaled questions, eleven of which answered the final question. While interviews with participants had been planned, due to differences in language and difficulty in finding interpreters to assist with communication it was not possible to conduct this portion of the installation study (See **Appendix 1** for the survey form).

4.3 Classification of Behaviors

Due to the inherently different user experiences across the LargeSpace and CCD installations, it was important to consider changes in user interaction in each installation. A common technique for analyzing user behavior is to record video of the full installation space over a set time period of interaction. This enables detailed analysis of multiple users which can be performed outside the context of the installation space in which most collaborators on the project were occupied with maintenance or operation of different aspects of the work. From the video, both quantitative and qualitative data can be collected and assessed. Because the ultimate goal of the project is related to the engagement and empowerment of users in relation to the subject of bird song, it is necessary to gauge the levels of interaction within the installation space.

For a quantitative aspect of analysis, the movement of users was considered as in relation to interactive artworks, a higher level of user activity often correlates with higher engagement with the work. Therefore, to determine the success of the various aspects of the work in which participants must relate to bird song through interaction, especially in relation to the movement of the speakers in the SIS system, consideration of user movement is paramount. It follows that from prior research on bird behavior regarding responses to modified bird song (Taylor et al. 2017), the process of recording a subject's movement in relation to interest has a close connection with the underlying research at the heart of BSD.

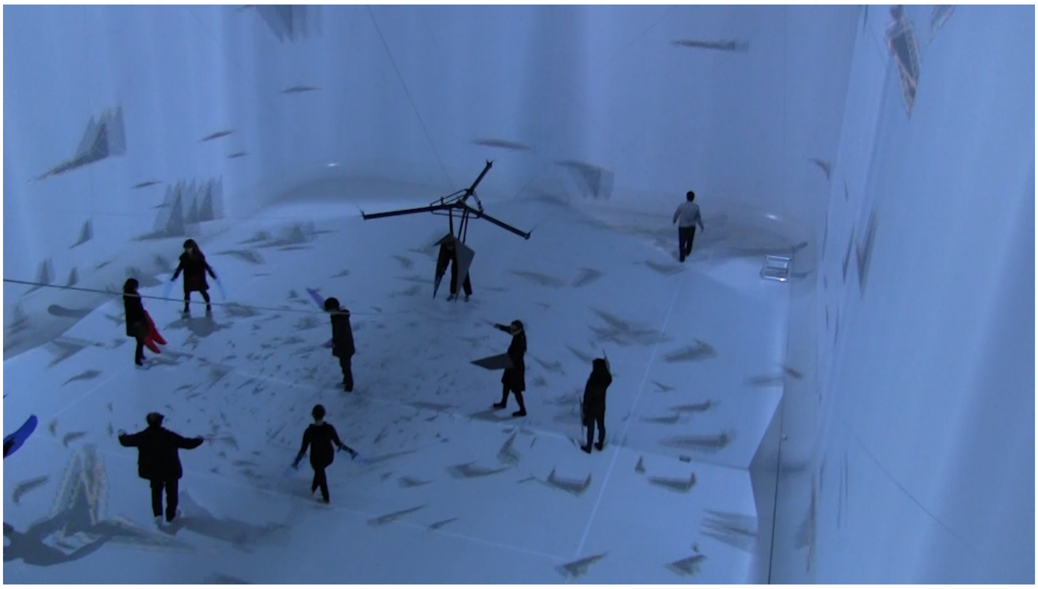
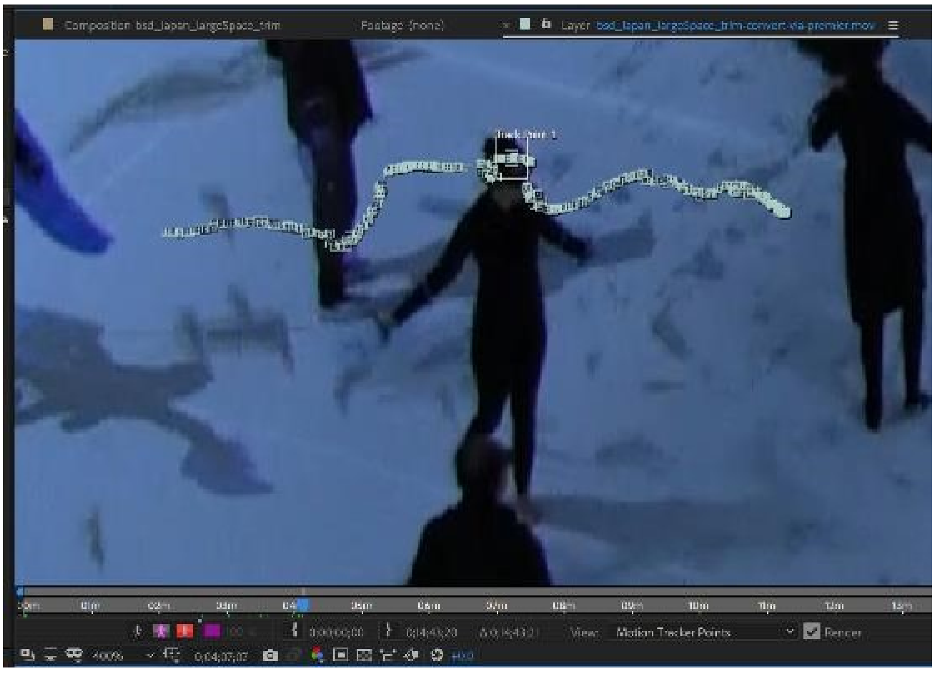


Figure 4.3: Angle of view for the LargeSpace performance recording.

In Taylor et al's *Sensitivity...*, the behavior of an identified California Thrasher, *Toxostoma Redivivum* (CATH), subject as it reacted to differently constructed bird songs was intended to show that songs constructed with attention to grammatical detail were more affective at inducing a behavioral response than randomly constructed songs. In the field, a camera was set to record a subject's territory in which a camouflaged speaker reproduced the experiment's songs. A combination of experimenter notes and manual video analysis, automated tracking was not feasible as the bird's size in relation to the frame was miniscule, lead to the estimation of each subject's distance to the speaker. Similar to the process of determining a California Thrasher's interest in a sound source, video observation of an installation can be used to gain some insight of interest and engagement.

Fortunately, for the sake of comparison, the LargeSpace performances were live streamed and a recording of a full fifteen minute performance was recovered. The camera recording the performance was mounted in the structural support at the top corner of the LargeSpace and, due to the height of the large space walls, could view a majority of the floor of the LargeSpace with only the area directly under the camera out of view (figure 4.3). For this performance, five participants were randomly selected to have their movements tracked. The process was far less intensive than the manual tracking done in the CATH research, with larger users and more predictable movements, but especially as the video length was only 15 minutes as opposed to the 40 to 60 minute long CATH videos. The primary difficulty in following the movements of participants was that the projections of the LargeSpace shifted over the course of the performance; changing from bright to dark to minimal to complex. While the larger participants allowed for an automated tracking process using Adobe's After Effects software (Adobe 2019), the variation in backgrounds, changing light, as well as participant occlusion and exiting frame, usually meant that the process could only be partially automated (figure 4.4) with manual interventions to correct a track or having to advance one frame at a time to maintain correct tracking. The resulting data, comprised of 2-dimensional (x,y) screen space position for each tracked user, meant much more detailed analysis could be done than on the CATH data in which only observable distance from the speaker was recorded.



Composition bsd_Japan_largeSpace_trim

Footage (none)

Layer bsd_Japan_largeSpace_trim-convert-via-premier.mov

View: Motion Tracker Points

☑ Render

Figure 4.4: Screenshot of the After Effects motion tracking process. The white square represents the area of interest which aids in stabilizing the track. The green squares show the locations of the track for each frame, and in case the automated tracking begins to drift, the user can realign the track.

Moving to the CCD installation, the difference in user experience, and especially the difference in the architectural space, called for a slight variation in the methods used for video recording. Where the LargeSpace was a 15 minute performance in which all users remained in the space for the full duration, the CCD installation allowed users to come and go whenever they felt the urge. Over the course of recording user behaviors different groups and total number of users varied, so this allowed an additional metric of duration spent in the installation to be measured. Therefore, the selection of users was based on groups that entered and exited the installation with the aim of following users of different groups who entered and exited within course of the video recording. Five users were selected based on this goal, precedence was given to users of different groups, however this resulted in the tracking of some users who were already in the installation at the start of the video recording. Further difficulty was due to the low ceiling of the space, only 3m as opposed to the nearly 8m of LargeSpace, and CCD had additional mounting restrictions meaning that only a handheld camera could be used in the video recording process. The placement was done at the far end of the room looking toward the columns and at a slight angle so that both mimic stations were visible, however the columns in the center of the room and the low angle of the recording meant that a lot of visual occlusion took place. Such problems made tracking users a more involved process and ultimately lead to issues in evaluation (see **Chapter 5**). Before actual tracking could take place, the shaky camera movement inherited from handheld recording with a phone camera (though higher resolution than the LargeSpace camera) had to be stabilized for both position and rotation to make sure that the tracked user movements were not shifted with the camera 4.5.

To compliment and give more meaning to the movement behavior it was necessary to add an additional observational element to the tracking data. Participants in the space moved throughout the space while doing a variety of activities and the type of activity can provide valuable understanding for analysis of engagement. Further, large amounts of movement might not always indicate engagement, similarly a lack of movement does not point to a lack of engagement. In the BSD Mimic module of the work, a participant stands and sings into a microphone while also waiting and listening for new songs to mimic. This is a situation where the amount of movement alone would hide that a participant is directly engaging with an aspect of the work, and another layer of information would provide valuable data regarding the

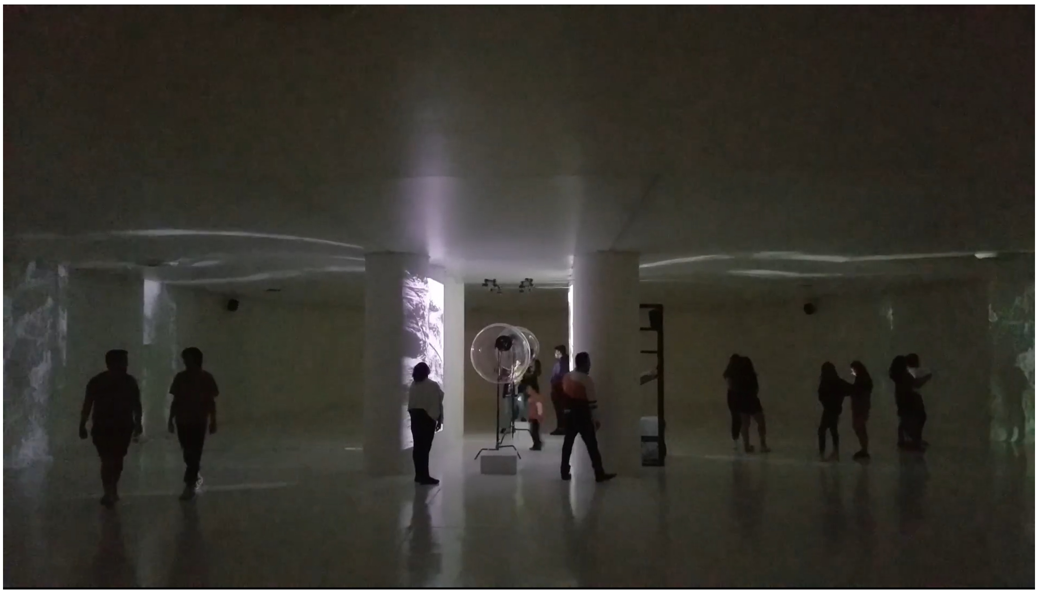


Figure 4.5: Angle of view for the recording of the CCD Memorial Installation. Note how it was nearly impossible to take an overhead video of participants given the space.

quality of movement. Based on observation in the installation space and further observation of users in the video recordings, four generalized classes of general user behavior were determined: Interacting, Watching, Listening, and Documenting. Related to the movement data, each of these qualities could occur during two movement states: standing and moving.

Primary to these qualities of behavior is the distinction between direct interaction and indirect interaction with the work. Direct interaction, classified here as “Interacting”, meaning time in which participants are actively participating, playing, and connecting with the work by intended or unintended means. Indirect interaction corresponds with the remaining classes (watching, listening, documenting) in which participants are observing or enjoying the work without necessarily activating the interactive qualities of the work. The documenting behavior was added after following behaviors in the CCD installation in which the prevalence of photography of the work was to such an extent that it merited its own category. It could be argued that this type of interaction could be an even more indirect form that those of watching and listening based on the additional layers of mediation created by portable devices and the associated presence of social media. Where watching and listening point attention toward the work itself in the form of cognition and understanding related to presence in the physical space, the documentation behavior could be interpreted as user attention pointed outward from the work itself and contextualizing it through documentation on a platform unrelated to birds or bird experience. Nevertheless, this form of interaction has become prevalent in contemporary art viewing, and will be included in this study based on its presence in the video documentation.

4.4 Parametric Speaker Experiments

Outside of user behavior in the context of the BSD installation, it was also necessary to focus on parametric speakers, their functionality, and the perception of audio movement using moving ultrasonic speakers in the context of the SIS system developed for BSD. While a few of the related research projects listed in **Chapter 3** conducted user studies with parametric speakers (Ochiai et al. 2017), and prior studies on moving audio sources have been conducted using a non-parametric speaker apparatus (Carlile & Best 2002; Mills 1958), no available studies on human perception of moving parametric speakers could be found. Because parametric speakers exhibit spatialized sound properties that are very different from other spatialization techniques, and also the techniques generally used for moving sound studies and psychoacoustic experiments involving sound movement, it was important to conduct an experiment to explore how subjects perceived of moving ultrasonic sound.

4.4.1 Demonstration of Directionality

It was necessary to establish the effective directionality of the parametric speakers before proceeding to the psychometric evaluation. As mentioned in **Chapter 3**, an experiment confirming the equivalence in angular spread between fixed and moving speakers was established. Following this experiment, an additional demonstration was performed in order to validate the reflective properties of parametric speakers. When a parametric speaker is aimed at a surface, the audio appears to come from that surface rather than the speaker itself, and depending on the smoothness of the surface, may reflect based on the angle of incidence of the audio source. This test involved a single moveable parametric speaker and a stereo recorder placed in a room with a single reflective panel (figure 4.6). The stereo recorder, comprised of two microphones angled away from each other at 45 degrees, was modified to include a small board separating the left and right microphones. The moveable speaker produced a constant amplitude white noise signal and was aimed in the direction of the microphone at three different angles in 15 degree increments. At each angle, the amplitude level received at each of the reorder's two microphones was recorded. The first angle was aimed directly at the recorder, the second was 15 degrees to the left of the recorder and aimed at the space between the microphone and the reflector, and the third angle was another 15 degrees further aimed directly at the reflecting panel. As can be seen in figure 4.7, when the speaker is aimed directly at the microphone as well as in the space between the microphone and reflector, the median amplitudes are roughly similar with channel 1 slightly below 0.05dB and channel 2 slightly above 0.05dB. When the speaker moves to the reflecting panel the amplitude at channel 2 increases above 0.10dB while the amplitude at channel 1 remains the same. This can confirm the reflecting capabilities of the parametric speakers as the channel 2 microphone is positioned on the side closest to the reflecting device, while the channel 1 microphone is on the other side of the microphone separators. If there was no audio reflection, then neither of the microphones would receive heightened signals when the speaker is aimed at the panel.

4.4.2 Movement Experiment

While demonstrating the directionality of the parametric speakers through the use of sensors enables a quantitative confirmation of the speaker's function, the ultimate use of the device is to create a unique spatialized experience for humans. Therefore, a user study involving human participants was necessary to confirm that humans are able to perceive this same directionality. Even further, due to a lack of research in the human perception of audio source movement of when produced by parametric speakers, an experiment in which participants determined the movement of audio sources

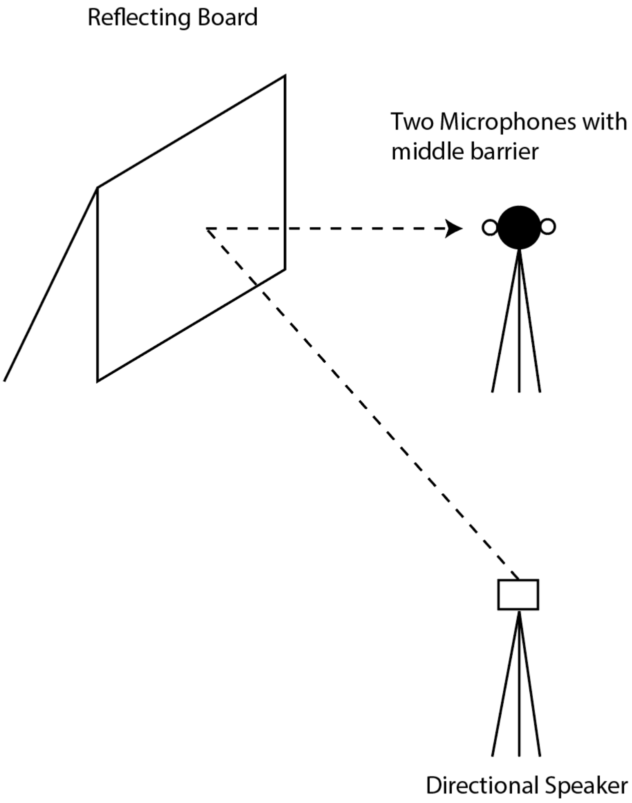


Figure 4.6: Diagram of the equipment and layout used in the reflective panel experiment.

was of great importance. As such, two experiments in which participants estimated the directional movement of a single audio source were performed with the second experiment consisting of a slightly more complex variation of the first.

The experiment was conducted in the Large Space and tested a moving directional speaker and its ability to convey motion (figure 4.8). One moving parametric speaker was mounted to the top of the structure of the LargeSpace at 7.7m above the floor. The speaker, aimed toward the wall opposite itself, reflected sound off the LargeSpace wall and back towards a subject. Through the reflective properties of parametric speakers demonstrated above, the subject would perceive that the sound is coming from the wall. Given that the angular position of the speaker could be remotely adjusted and animated, it was possible to use this apparatus to conduct the aforementioned movement experiments.

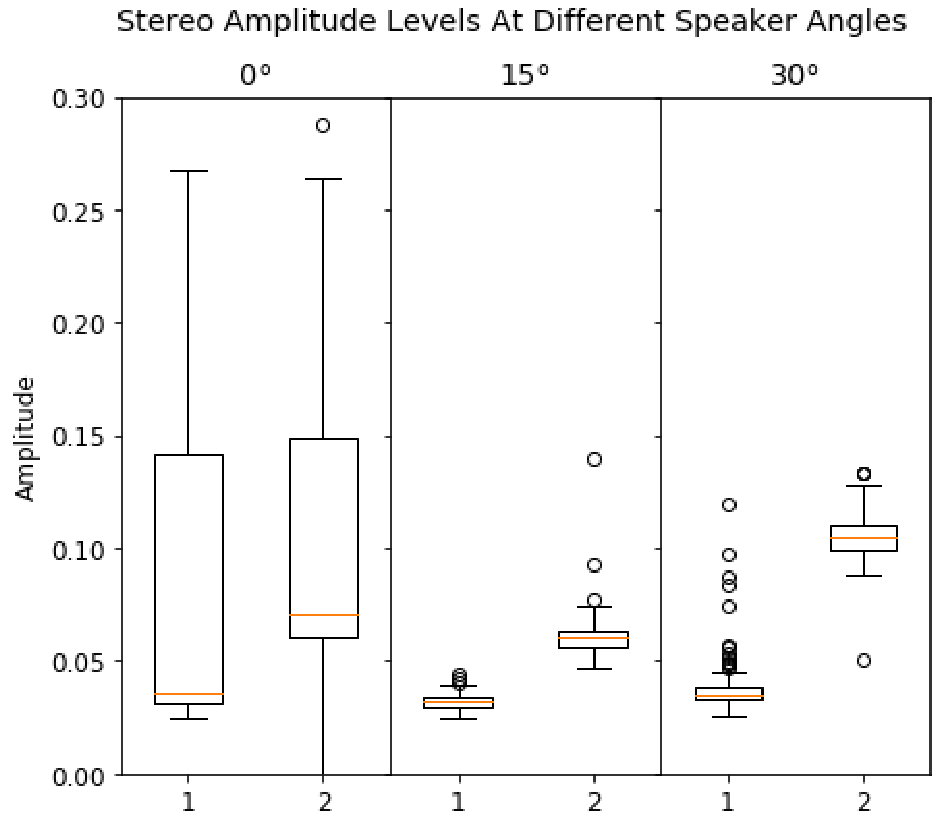


Figure 4.7: Comparison of stereo amplitude for different speaker angles using Box and Whisker plots.

In the following experiments each subject stood on the midline of the LargeSpace's floor at 7.5m from the reflecting wall. Subjects faced away from the speaker and towards the wall from which the reflections emanated. The speaker was positioned 3.55m to the left of the subject, and positioned downward and to the right in such a way that the sound produced from the speaker would hit the far wall and reflect directly at the position of the subject. This speaker position, 8 degrees to the right of center and 16 degrees below horizontal (pan: 82 and tilt: 106 as servo angles), was considered the *center* position and all experimental movements were based relative to this position. In this central position the virtual location of the sound on the wall was 1.18m to the left of the subject and 3.23m above the floor. While this *center* position results in an audio source that is offset from the directly in front of the subject, it enables the virtual source to move diagonally or vertically along the wall without causing additional interference reflections that

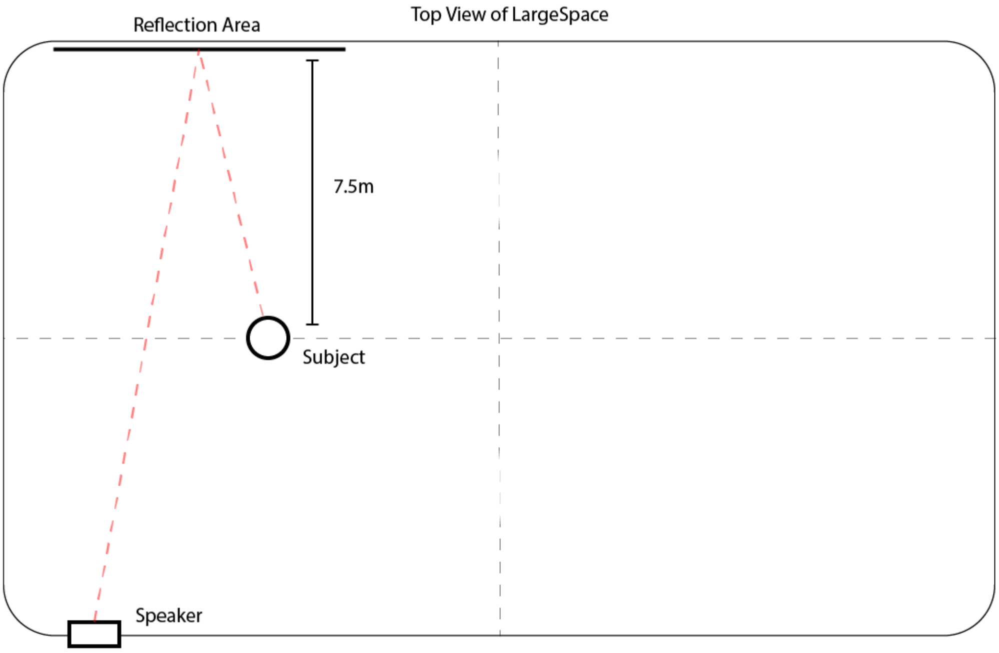


Figure 4.8: Diagram showing overhead view of the LargeSpace setup used to conduct both movement and constant stimuli tests. Subjects are placed at the midpoint between the speaker and the reflection area. The dashed red line demonstrates the pathway sound takes to reach the subject when the speaker is set to the central position.

might confuse the subject. If the speaker were placed directly behind the subject, any downward vertical movement would result in sound reflecting on the floor in front of the subject or even on the body of the subject causing unintended directions. As will be seen in the next chapter, errant and misleading reflections are very difficult to control and, even with precautions, may be an inevitable problem when using ultrasonic directional audio.

Using this experimental apparatus, an experimenter can vary the angle of the speaker to give the impression of the sound source moving along the wall. Based around the central position on the wall, if a constant sound source moved from the left side to the right side of the wall, the subject would be able to perceive the movement based on the differences, over time, in sound between the left and right ears; known as interaural time difference (ITD). Subjects can be asked to identify whether the virtual audio source is moving from left to right, right to left, top to bottom, etc. to validate if existing figures of sound motion perception, primarily the minimum audible angle (MAA) and minimum audible movement angle (MAMA), apply for parametric speakers in a SIS system (Hermann et al. 2011).

A number of tests were considered for measuring the directional accuracy of virtual sound sources for participants. Because non-moving directionality of parametric speakers has been demonstrated and the primary contribution that this SIS system makes to the spatialized sound is through the dynamic movement of directional speakers, the tests considered were all intended to evaluate the movement of sound using the system. Early possibilities included identifying linear movement of a sound source, identification of simple shapes (square, circle, triangle, etc.) as the path was traced by the virtual source, and attempting to draw a path freehand after listening to a nonlinear path. Prior to the experiment, a number of feasibility tests examined each of these trials. One test, for example, consisted of participants trying to discern the difference between a rectangular path and a circular path. It was quickly found that freehand drawing and shape identification were far too difficult for an untrained listener, but tests involving linear motion were consistent and could also be modified to increase difficulty.

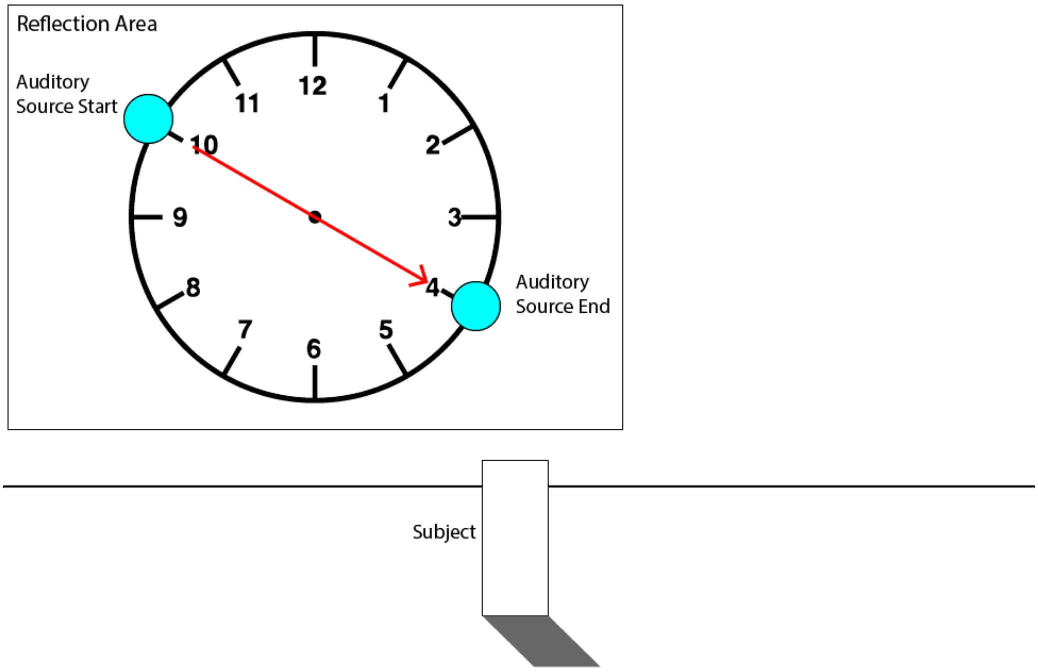


Figure 4.9: Diagram showing a side view of the LargeSpace wall and demonstrating the way in which sounds move across the reflection area during a trial. During the experiment there are no visual indications of position projected onto the LargeSpace wall.

The first experiment measured a subject's ability to identify the linear direction of a moving sound source. The sound will move on the reflecting wall from one side of the central position to the other in a straight line and the reflected sound will pass across the subject's field of audition (meaning the area) as they remain in a stationary position (figure 4.9). After presenting one of these sound movements, the subject will mark on a paper the starting and ending point of the sound. Beginning with the four cardinal directions (up, down, left right), the number of possible paths was increased in preliminary tests until paths became difficult to clearly distinguish one from another. This resulted in subdividing the four cardinal directions with eight additional directions for 12 possible starting points each 30 degrees apart (corresponding to the 12 marker positions denoting the hours on a clock's face). These values were well above the MAA determined in prior psychoacoustic studies which varies depending on a number of factors, though can be as low as 1-2 degrees (Mills 1958).

Each subject would receive 48 randomized positions, while ensuring that each position was tested equally. Additionally, there were four possible speeds in which the sound source could move along the path, and every twelve trials a different speed value was set. The inclusion of differing speeds was based on prior trials for the MAMA which showed that the value was directly related to the speed of the audio (Carlile & Best 2002). Worksheets were given to subjects which contained 48 circles with twelve labeled marks on each circle. Every trial began with no sound and the pan-tilt unit moving to the starting position of the path. At the start of the path movement, the speaker would emit broadband noise and continue until the end of the path when both the pan-tilt movement and the sound would cease. The movements themselves used a short ramp at the head and tail of the move and a full movement was based on the speed of that set of trials (1,2,3, and 4 seconds). The total movement distance covered 8m along the wall, with 4m from the starting point to center and another 4m from the center to the ending point (see **Appendix 2** for testing sheet).

An additional experiment was added to increase the complexity of sound movements, while also considering how accurately subjects could place the beginning and ending positions of a movement path. A continuation of the first experiment, all aspects regarding hardware, sound, speed, movement distance, number of positions, number of trials, and worksheet remained the same. The modification which added difficulty was that the ending position of the path was also randomly placed. The path movement would now start at a position, move to the center, and then change direction to end at a different point along the circle. It was also possible to have a movement identical to that of the first experiment. A subject would have to identify the direction of the sound source's movement as it entered into the subject's of audition, and then identify another direction of movement for a sound source exiting the subject's field of audition. As the protocol was identical to the first experiment, each participant evaluated 48 randomized movement paths at 4 different speeds.

4.4.3 Constant Stimuli Experiment

The use case for SIS is often in relation to visual elements, and often within a visually immersive environment like the LargeSpace. Because of the inevitable pairing between both sound and visual elements, consideration regarding the current SIS system's ability to accurately place a virtual audio source so that it appears to emit from a virtual visual visual source. While a perfect, identical placement between audio and visual sources would be optimal, the required calibration and mechanical precision, especially in a space covering many square meters, is not feasible. Further, human perception of spatialized sound sources is much less precise than that of visual elements, meaning that there are levels of inaccuracy which are acceptable for developing virtual spatialized audiovisual sources. The acceptable displacement between visual and auditory stimuli is important information, when considering how accurate the SIS device needs to be for creating successful experiences in which the audio and the visual are intertwined.

It is a known perceptual phenomena that the vision and audition are not treated equally by a human receiver. It has been shown through a number of studies that in the activity of auditory localization, the presence of a visual stimuli greatly biases, and supersedes, the auditory source; often "snapping" it to the position of the visual stimuli. This notion of visual dominance, mentioned briefly in **Chapter 1**, is commonly referred to as the "Ventriloquist Effect" or visual capture (Pick et al. 1969), and is also related to the Colavita visual dominance effect, named after the researcher who experimentally demonstrated this effect in the early 1970s (Colavita 1974). More recent studies have considered this effect in which the audio and visual stimuli are temporally and spatially separated, finding that visual dominance effects are significantly pronounced in situations of audio and visual coincidence (Koppen & Spence 2008; Koppen & Spence 2007). Because the goal for certain SIS installations is to achieve true coincidence of an audio and visual element, where the auditory and visual stimuli occur at the same time and in the same position, it can be assumed that SIS is often working in situations where the visual is most dominant over the auditory.

Though spatial factors have been tested to confirm that this effect occurs at close range distances (less than 1m) using panning techniques for sound localization, and that stimulus localization is often a factor of the sum of weighted sensory inputs (Battaglia et al. 2003; Alais & Burr 2004), the perceptual threshold in which a human can identify a localized sound source as spatially distinct from a visual source has not been thoroughly explored using a parametric speaker setup in a larger scale environment. Using the same experimental system as the previous movement-based tests, a third experiment was conducted to evaluate the ability for subjects to identify the location of the sound source in relative to a visual marker on a wall. Based on the results from this experiment, a better understanding of the point at which sound and visual stimulus are perceived as coincident or separate should emerge.

The LargeSpace SIS system described in the previous section, a single speaker reflecting audio on different positions of a wall, was used with the addition of a short throw projector to display visual markers on the reflecting wall with a projection width of 10m. The participant was again positioned at 7.5m facing the wall, and again offset to right side of the projection area so as not to block the speaker's audio during vertical movements. A constant stimuli method was used to measure participant responses over 90 trials, separated into 6 blocks of 15 trials with half of the blocks dedicated to horizontal localization tests and the other half to vertical.

A single trial consisted of a bimodal source, auditory and visual, activating for a period of 2 seconds, after which the participant was given a two-alternative forced choice (2AFC) (Fechner & Wundt 1889) whether the audio source was left-right (horizontal) or down-up (vertical) in relation to the visual source. Broadband noise was used for the sound source, and the visual source was a 1m diameter circle projected on the wall. The audio source would be randomly assigned one of 5 distance separation values: -2.5m, -1.25m, 0m, 1.25m, 2.5m, corresponding to the distance between the sound and visual stimuli, where negative values indicate left/down distances and positive right/up distances relative to the center of the visual source. Each of the five intensity levels received at least 15 participant responses. The visual source could appear at three possible positions: left, center, and right, with the center position corresponding to the audio center (see previous section) and the left and right positions offset by 2.5m in the respective direction (figure 4.10).

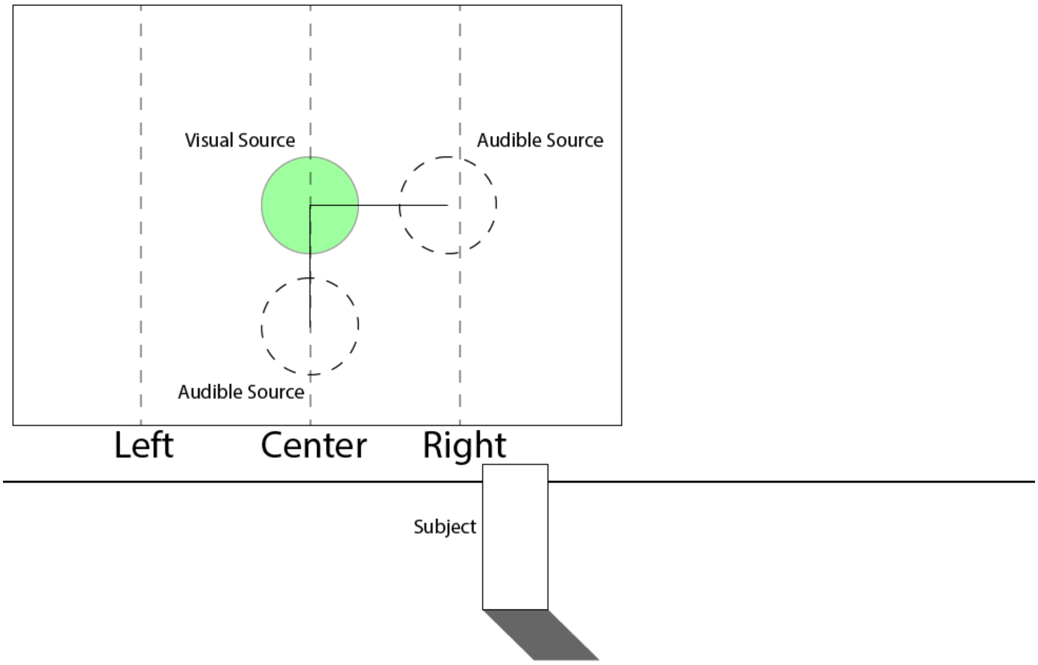


Figure 4.10: Diagram showing a side view of the LargeSpace wall and illustrating the way in which sounds move across the reflection area. During the experiment there are no visual indications of position projected onto the LargeSpace wall.

The 5 intensity values were chosen based on the the maximum width allowed by the projection area (10m wide and 6m tall). These values also correspond to the diameter of spread of the sound source when arriving at the wall, and in preliminary testing the -2.5m and 2.5m values provided consistent responses for selecting as the low and high intensity limits. Though many potential conditions are possible to isolate for evaluation (horizontal, vertical, left, center, right, and even the combinations, e.g. vertical-left or horizontal-center), the lack of trial numbers for every condition restricts analysis to only conditions with greater than 20 trials. In total, ten subjects participated in all three experiments.

4.5 Conclusion

The collection of a variety of data, from qualitative visitor feedback to video analysis to quantitative perceptual threshold estimation, covering performance of parametric speakers for localizing sound and SIS implementation in artistic contexts was intended as a means of further exploring the impact of SIS and the effectiveness of SIS for the empowerment of users. While independent studies may not directly engage with the topic of empowerment, if the system itself is well understood and can be optimized to maximize the affective capacity of SIS or a larger multi-modal immersive environment, then SIS as a tool wielded by artists and experience designers will inherently lead to the empowerment of both the creators and participants.

Chapter 5

Discussion

The importance of a broad range of tests, experiments, and surveys for collecting evidence on the viability and effectiveness of a system are necessary to form a convincing argument. In the case of SIS, the previous chapter provided a survey of the variegated methods used to collect information about the system and its potential as a tool for increasing empowerment of participants. Basic system validation was a necessary starting point for demonstrating that such hardware can accomplish spatialized sound as well as dynamically moving spatialized sound. The next step involved experiments to gauge the ability of listeners to recognize a moving spatialized sound source so that the limitations of the system could inform future movement and SIS design. An additional experiment to estimate the influence of visual capture on localized sound with parametric speakers was conducted with similar goals of understanding the effects of SIS systems on participants. Finally, behavioral data and user reporting was collected in the context of a real world installation involving SIS, which, due to the iterative nature of the work, allowed for comparison between prior versions that did not implement the same SIS system. The following sections will present the findings for each of these studies regarding the effects of SIS on users and its potential for empowerment.

5.1 Parametric Speaker Experiment Results

Tests with the parametric speakers were initially done to show the extent of the spatial properties offered by such a system. This can be seen in the previously discussed experiments involving the ability of a speaker to record the output of a parametric speaker (**3.3.2**, **4.4.1**). The first of these tests, **3.3.2**, demonstrated that a moving parametric speaker could fully reproduce the same angular spread as that of a fixed speaker (figure 3.12) and further validated that the horizontal angular spread of the parametric speakers used in the SIS system was 20 degrees. Further implications of this test can be seen in figure 3.13, in which, without manually moving the microphone and programming the pan-tilt unit to incrementally cover its entire positional range, a scan of the reflective properties of a space can be conducted resulting in a sonic cartography for a region based on amplitude levels. The second, **4.4.1**, demonstrated that a reflection method can be used to create the impression of a physical object emitting sound. This technique is necessary to establish as all further experiments and implementations with the SIS system involve the reflective technique to position virtual sound sources (as opposed to a direct tracking technique or the dual-speaker interference technique). While the experimental results (figure figure 4.6) may not show a dramatic change in amplitude levels for each channel between the direct (0 degrees), indirect (15 degrees), and reflected (30 degrees) positions, due to difficulty in fully isolating the left and right microphones of the recording device, there is still a 0.05 decibel increase in the median amplitude for the channel closer to the reflected source (channel 2). These tests illustrate that the speakers are spatially functional as expected and further tests can be done to estimate the perceptual capabilities of parametric speakers and SIS.

5.1.1 Results of Movement Experiment

The movement experiment provided the first controlled test of a range of participants for recognizing spatialized sound. The same ten participants were involved across both this experiment and the constant stimuli experiment described in the following section. The participants were all affiliated with the University of Tsukuba and participant ages fell between 20 to 40 years old, with a median age of 26. None reported any problems with hearing, and some reported experience as musicians or had some familiarization with activities involving spatial recognition of sound (for example, playing games which convey directional information through sound). The methods described in **4.4.2** was carried out for each user, and, prior to the start of the trials, a detailed explanation of the worksheet, possible sound movements, and step by step procedure was given which included a demonstration of the sound to be played as well as one of the sound movements. Because the system itself, as well as the activity of following the movement of a sound source (with no visual source), is quite esoteric, it was deemed necessary to provide such an explanation so that all participants would have an equal understanding of the experiment.

After completing the experiment, a number of casual interview questions were asked to gain immediate feedback about the subject's own feelings regarding their performance. Questions included statements such as, "At which speed did you find it easier to discern the direction of the sounds?" or, "Did you develop any strategies for better sound recognition?" The aim of these questions was to try and understand how participants approached this activity and how they felt afterwards about SIS. Across the full range of subjects was a lack of confidence, with most participants expressing some form of "I hope that my answers were OK" or "I hope you are able to use my data". A number of subjects claimed that certain speeds felt much easier or difficult, with some feeling that faster speeds did not allow enough time to predict the positions or that slower speeds created more artifacts which made it difficult to guess. Most participants felt that keeping their eyes focused on the reflection wall aided in prediction, though a few participants closed their eyes during the experiments to reduce any visual distractions. Ultimately, the variety of responses received between participants only reflected an opinion of the overall experiment being quite difficult, while no unified technique or speed level received a majority of support.

(2)

$$error = \begin{cases} \pm error_{max} - (a_2 - a_1), & \text{if } |a_2 - a_1| > \frac{1}{2} error_{max} \\ a_2 - a_1, & \text{otherwise} \end{cases}$$

After adding the participant responses to a table containing the actual positional values as well as the speeds for each trial, each test received an *error* score to indicate the number of positions a user's response was from the actual position of the sound. Two types of errors were calculated, one indicated the direction in which the user tended to err, positive values being clockwise and negative values counterclockwise to the actual position, ranging from 0 to ±6 with 0 representing a response that matched the actual position. The second error indicated the absolute value of the error, ranging from 0 to 6, and indicating the distance away from the actual position without a directional component. The data was formatted according to the tidy data standard (Wickham 2014), and an example of the first five rows of the data can be seen in Table 5.1.

Table 5.1: Example of the participant data set for the single movement experiment. Because participants were only instructed to record the starting value, in this test the ending value, *MoveEnd*, is always a fixed amount from the starting position, *MoveStart* (6 positions away), the responses of the subject, *SubjectStart* and *SubjectEnd*, only contain values for the starting point. The *MoveTime* column indicates the speed of the movement. Each row represents a single trial.

ID	Test	MoveStart	MoveEnd	MoveTime	SubStart	err	abs_err
1	1-pt	7	1	2	6.0	1.0	1.0
1	1-pt	8	2	2	6.0	2.0	2.0
1	1-pt	9	3	2	0.0	3.0	3.0
1	1-pt	2	8	2	0.0	2.0	2.0
1	1-pt	4	10	2	9.0	-5.0	5.0

Because the second movement experiment closely imitated the first in both form and structure, a similar process was followed for inputting the user responses for starting and ending positions. When calculating the error for the second experiment, it was now possible to determine three different measures of error for each trial. The first, identical to that of the first experiment, was the error between the actual and reported starting positions. Second, as there was now an additional position for the participant to guess, was the error between actual and reported ending positions. And finally, in the interest of determining a combined error for each trial, the third error was simply the combination of the absolute starting and ending errors and could range from 0 to 12.

Table 5.2: Each column displays the error counts for each level of error indicated in the leftmost column. To the right of the *Error Amount* column, the *1-Direction* column shows counts for the single direction experiment. The remaining columns split the two direction experiment into starting and ending error counts.

Error Amount	1-Direction	2-Direction Start	2-Direction End
0	97	76	86
1	172	126	130
2	84	109	104
3	53	48	54
4	26	58	44
5	31	38	36
6	15	25	26

With the difference errors calculated between user and actual positions, it is possible to use these values to compare the ability of participants for determining the direction of movement with this specific SIS system. Beginning with the overall counts in magnitudes of error, some initial observations can be made. The histograms shown in figure 5.1 reflect the number of each error value, with directional error to the left and absolute value error to the right, recorded for all users during the single direction tests ($n = 478$). A cursory glance at these shows much higher counts as the error approaches zero, with the directional error showing the highest count at zero with 97 successful attempts. Taking the absolute value of the error shows that much more often, a participant was off by a one positional value with 89 in the counterclockwise direction and 83 clockwise to the actual value, making a total of 172 attempts which were off by a single position. With successive increases in error starting from 2, the number of recordings in absolute error decreases with counts of 84, 53, 26, 31, and 15 for 2,3,4,5, and 6 respectively (Table 5.2).

The mean of directional error is expected to be zero in for a normal distribution, and the single direction experiment results in a mean directional error of 0.10 which is quite close given the low number of participants. In a one-sample T-Test comparing directional error with an assumed mean of zero, with a statistical significance threshold of 0.05, the resulting p-value of 0.620 shows that, while the mean is not actually zero, the data is not significantly different from a mean of zero. The absolute error for all subjects in the single direction experiment came to 1.77 with a median error of 1, indicating that recognition of position regarding movement using this system is feasible. When compared with errors calculated from randomly generated responses using a Wilcoxon Signed-Rank test, again with threshold of 0.05, it is shown to be statistically different, $p = 0.00691$, which confirms the visible trend shown in 5.1 that the participant responses are not chance guesses.

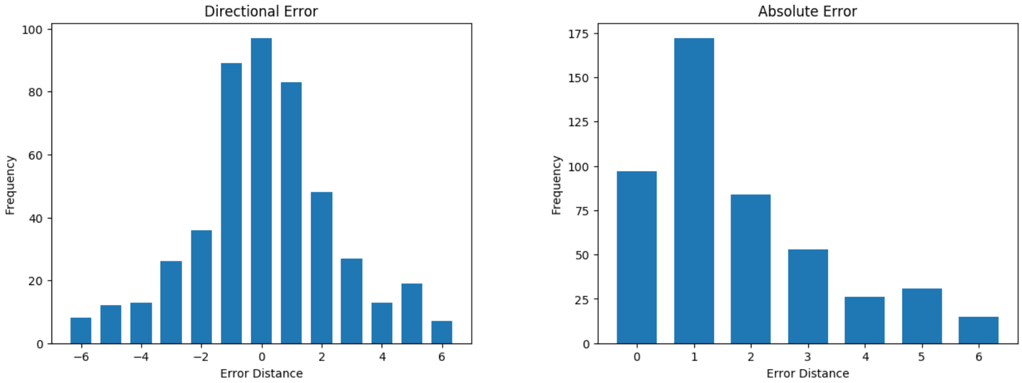


Figure 5.1: Histograms showing the frequencies for directional (left) and absolute error (right) separated by error distance for participants in the single direction movement experiment.

Examining the data collected for the two-direction experiment, performed by the same ten users with 48 trials each, the data can be split depending on the first, or starting position, and the second, ending position. Overall, both starting and ending histograms, figures 5.2 and 5.3, show higher occurrences at larger error intervals, specifically interval 2 and 4, with lower occurrences of exact, 0, or 1-position errors. A comparison of absolute error counts with the single direction test in Table 5.2 shows that the absolute error interval of 2 surpasses the 0 interval for both starting and ending positions going

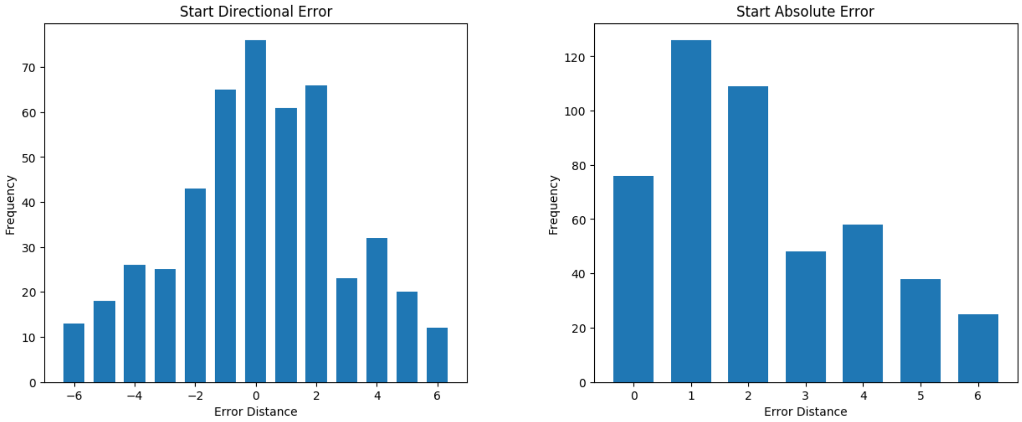


Figure 5.2: Two direction experiment histograms for the starting position. Directional (left) and absolute error (right)

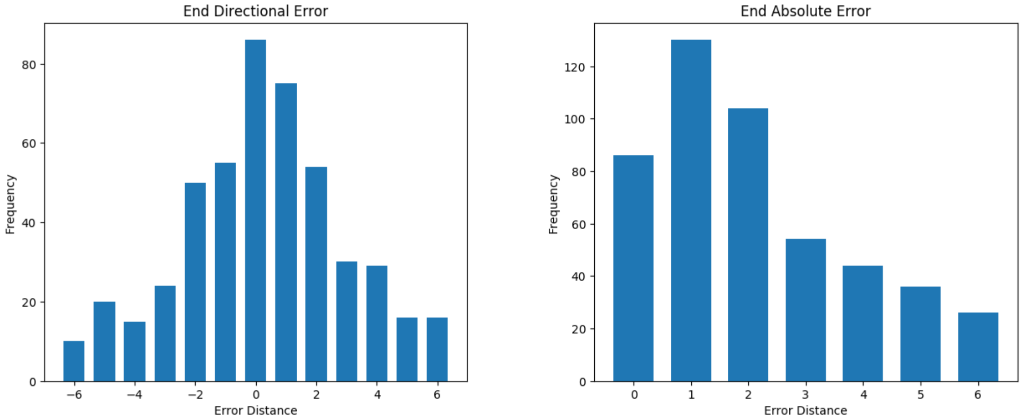


Figure 5.3: Two direction experiment histograms for the end position. Directional (left) and absolute error (right)

from 84 occurrences in the single direction test to 109 start position and 104 end position occurrences in the two direction test. The addition of complexity in the two direction experiment seems to effect both guessing attempts resulting in higher levels of error. The mean absolute error for the starting and ending positions have increased to 2.20 and 2.10, respectively, and the median absolute error for both positions has also increased to 2.

Interestingly, participant performance on the starting position does not seem to affect performance at the ending position. For comparing the mean absolute error for each user in the single direction experiment with the two mean absolute errors of the two direction experiment, a Wilcoxon Signed-Rank test was used due to the paired nature of the experiment along with the low number of participants ($n = 10$) (IDRE 2013). Using a predetermined threshold of 0.05, the test resulted in p-values of 0.021 and 0.059 for the start and end absolute errors, respectively. Here the starting error distribution shows a statistical difference from the single direction experiment, while the ending error distribution, albeit just on the border, does not result in a difference. Given this, it may indicate that the incoming or outgoing movement of sound may have a greater effect on a participant's ability to discern the direction of movement. As described in the previous chapter, the first position will always move inward towards the center, or sweet spot, of the auditory field of the participant, while the second position always moves outward from the central location. To absolutely verify this hypothesis, a stronger statistical value to indicate difference needs to be shown. As such, perhaps conducting a study more focused on this particular aspect of sound movement, or simply a larger study involving many more participants will provide a more robust dataset which can accommodate a stronger analysis.

The error rate for participants at different positions is another concern, as prior tests in perception regarding sound localization and movement have reported that horizontally placed sound, positioned along the azimuth of the user's auditory field, outperforms sound which is placed vertically, along the elevation relative to the user (Hofman & Van Opstal 2003). Based on this research it is expected that higher error rates should occur when positions, single, starting, or ending, approach upward and downward values (considering the clock face upward values correspond to 11,12,1 and downward values to 5,6,7; note that 12 is converted to 0). To explore the effects of direction on error magnitude, a confusion matrix was developed which pairs the actual sound positions with user reported positions. Figure 5.4 shows this relationship, in which higher counts are displayed as darker colors to indicate groupings of responses. An ideal graph would have all data line up on the diagonal, where every participant's response was exactly that of the actual value. The single direction graph shows user responses to be primarily grouped along this diagonal, however there are two spots in which the density of the diagonal line is reduced around the positions of 3 and 9. From another perspective, it appears that much of the answers are grouped around the 0 and 6 positions. Figure 5.5 shows confusion matrixes for the starting and ending positions where this tendency is even more pronounced. These results seem to contradict the claim that subjects perceive sound more accurately when it appears along a horizon, as here the figures show that higher error intervals occur with horizontal positions while lower error intervals are primarily at the vertical positions. Creating a box-whisker plot showing the absolute error for each position in the single direction test, figure 5.6 highlights that positions 3, 4, 8, and 9 all have a median error of 2, while the rest of the positions all display a median error of 1. The plots for the two directional experiment are shown in figure 5.7. It is unclear whether this result is an artifact of the experimental procedure, or if this type of system actually induces a different perceptual response than the types of systems used to determine such responses in the past. One possibility is that participants, faced with a ambiguity in possibilities (e.g. position 2, 3, or 4), might conflate responses to the more cardinal directions of 0, 3, 6, and 9. This might explain why in the end position plot of figure 5.7 that position 4's median is elevated even further than that of position 3. However, if this were entirely the case, there would be clusters around the four cardinal directions instead of two. As these are the first experiments of this kind which use parametric speakers as the directional sound source, it may require further testing in order to dredge up the cause.

Because the speed of the sound movements was varied over the course of the experiment, it is also of interest to look into whether or not each speed category had any effect on participant error. Figures 5.8 and 5.9 show a full comparison of error in both single direction and two direction experiments. User responses were grouped by each movement speed, where the movement speed denotes the duration, in seconds, of each trial. The grouped responses are displayed using a Box and Whisker plotting technique to aid in comparisons between different times of multiple statistical dimensions (mean, median, quartiles, range, and outliers). Looking at the collected data, there aren't many notable differences in error distribution between the various speed levels. In the single direction experiment, there is a slight difference in the range and upper quartile of directional error for speeds of 1 second, which could account for a bit of increased bias towards erring in a clockwise direction for higher speeds. For absolute error in single direction trials,

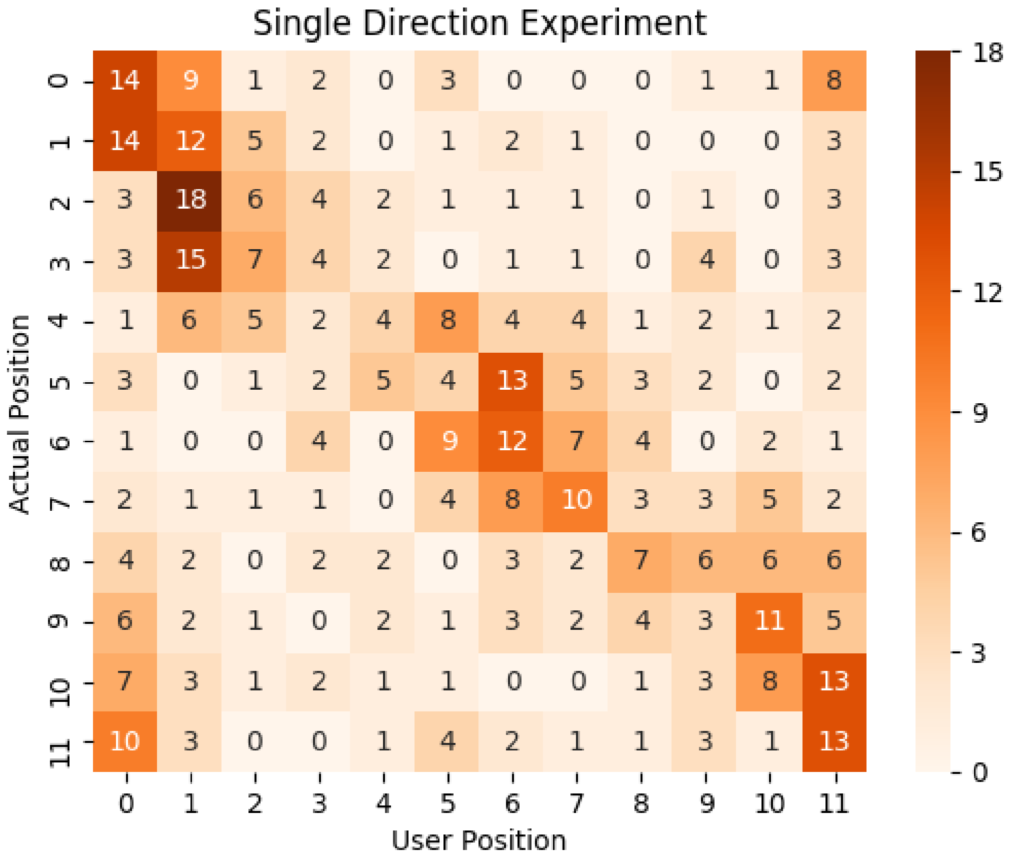


Figure 5.4: Confusion Matrix representing the overall accuracy of participants across each positional value. Darker colors indicate higher counts.

the 1 second group no longer stands out, while the 4 second duration now indicates a reduced range and upper quartile. In conjunction with fewer outliers in the directional error plot, this would indicate that slower speeds could reduce error which contradicts a few participant interviews that mentioned slower speeds to be more confusing. When considering the data for the two direction experiments, there are fewer standout differences between the movement speeds. The 1 second groups still show a bit worse responses, with the 1 second group for directional end position error reflecting another bias towards clockwise error. Absolute errors for the starting and ending positions show an interesting tradeoff in which starting times that show higher error, 1 and 3 seconds, show lower error for ending positions and vice versa. It is uncertain as to what is causing this behavior, however it may simply be an indication of the low sample size for the experiments.

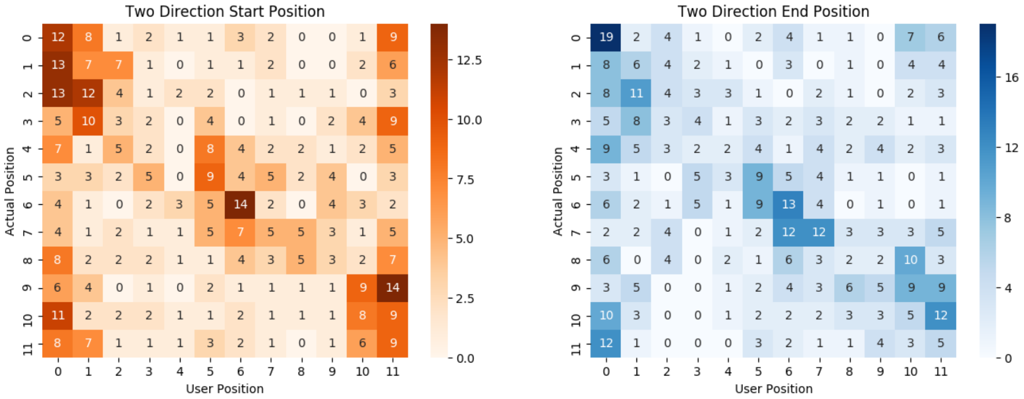


Figure 5.5: Confusion Matrix representing the overall accuracy of users for the starting (left-/orange) and ending (right/blue) positional values.

5.1.2 Results of Constant Stimuli Experiment

The constant stimuli experiment was meant to understand the perception of SIS sound while in the presence of a visual stimulus, especially in when such stimulus is the intended source of the localized sound. Collecting and interpreting the participant data for the constant stimuli experiment, the same 10 subjects from the movement experiments participated in the constant stimuli trials, involved creating plots of the mean responses for each subject and then using these points to estimate each subject's psychometric curve. Based on the collected subject's curves, it is possible to consider how each subject perceived the relationship between the auditory and visual stimuli.

The following figures will display the mean values, generated curves, or both, based on each of the 5 intensities: -2.5, -1.25, 0, 1.25, 2.50, which correspond to the distance of the auditory source to the visual source as described in the previous chapter. In each figure, the y-values represent the proportion of participants who thought that the sound was to the right or above the visual source (through a 2AFC process) with values ranging from 0.0, in which none responded that the sound was to the right/above, to 1.0 where all identified the auditory source as to the right/above the visual source. For example, in the horizontal tests, when a sound was 2.5m to the left of the visual source (-2.5), ideally zero participants would respond that it was to the right resulting in 0.0 as the mean response for -2.5. The vertical lines

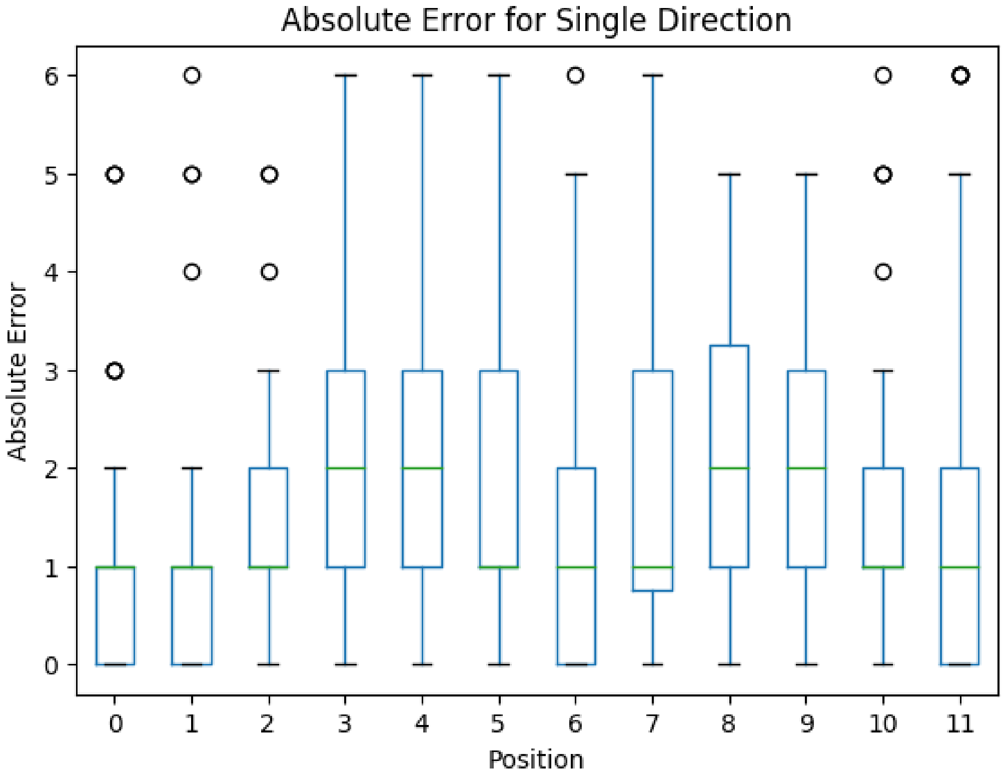


Figure 5.6: Box and Whisker plot showing absolute error distributions for each position of the single direction experiment.

mark the theoretical point along each curve that a participant would guess that a sound was to the right 50% of the time, $y = 0.5$, which defines a value, X_{50}, known as the point of subjective equality (PSE) (Meese 1995).

According to the graphs in which participant data could considered for generating a psychometric curve, more on this below, the PSEs for the majority of participants occurred when the auditory source was to the left of the visual source. Because of this perceptual bias, each successive figure attempts to isolate, as much as was possible given the number of participant responses, a different parameter in order to evaluate what effect each had on the overall outcome. With 90 responses for each subject, very specific parameters (e.g. direction: horizontal, position: left) would only have 15 trials per subject. As the specificity of the parameters increases, the validity of the results are diminished due to low trial numbers. An activation function used in fitting a psychometric curve, and for the generated data a logistic sigmoid function was used with L as the maximum of the curve, k as the growth

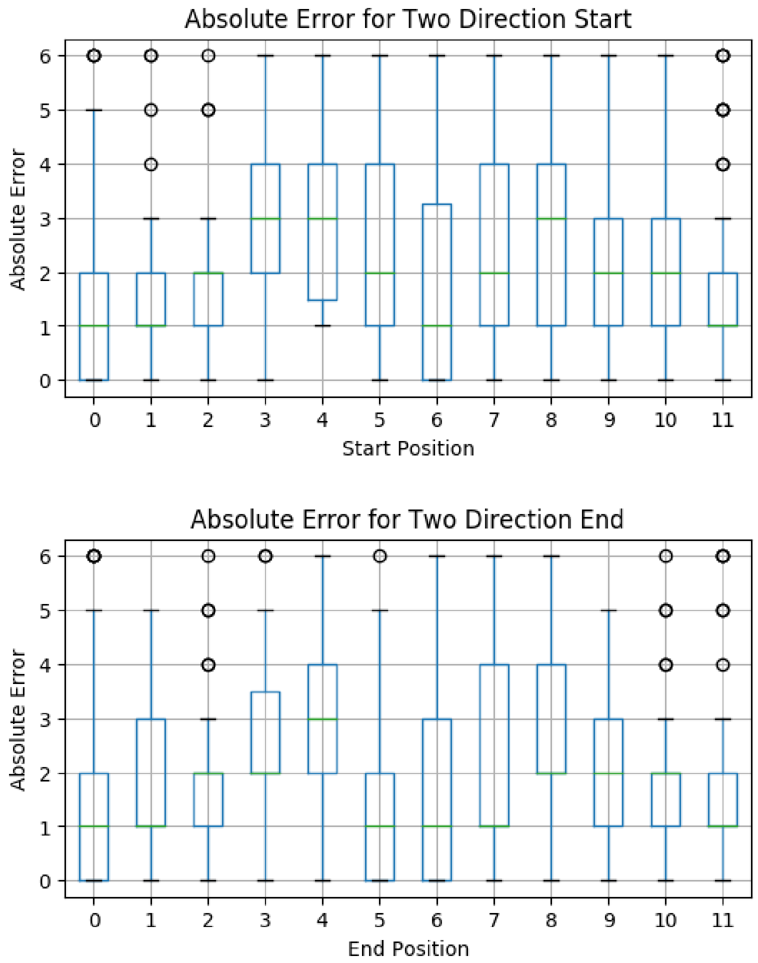


Figure 5.7: Box and Whisker plot showing absolute error distributions for the starting (top) and ending (bottom) positions of the two direction experiment.

rate, and x_0 as the midpoint:

(3)

$$f(x) = \frac{L}{1 + \mathrm{e}^{-k(x-x_0)}}$$

An ideal psychometric curve would correspond to a centered PSE, where $X_{50} = 0.0$ and the extreme values would correspond to $X_0 = -2.5$ and $X_{100} = 2.5$, with the curve moving upward as the intensity values become more positive as higher percentages of participants identified the sound source to the right of the visual source. As mentioned in the previous chapter, the -2.5m and 2.5m values were rough estimates of values that would result

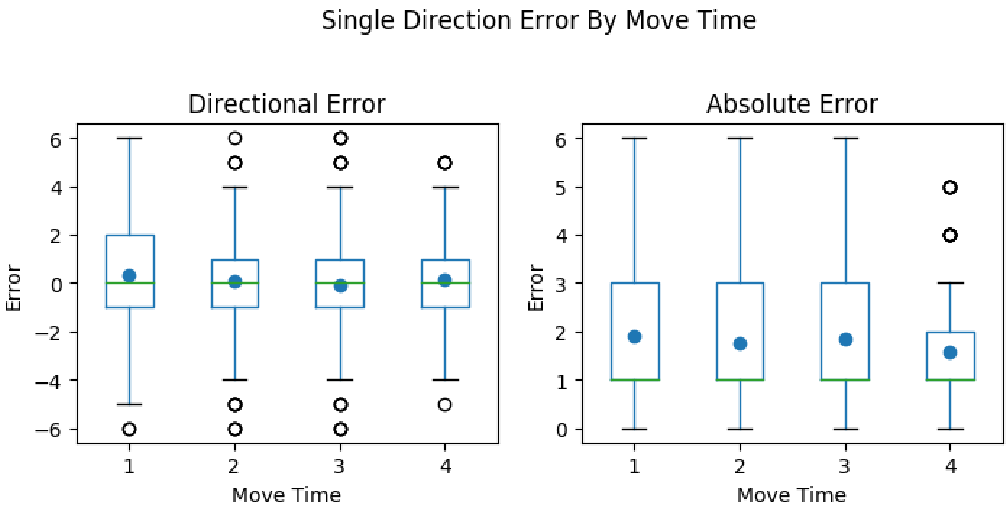


Figure 5.8: Box and Whisker plot of directional and absolute error grouped by movement time for the single direction experiment.

in 100% correct guesses. Figures 5.11 and 5.10 show the generated curves for the combined user means and means separated by user, respectively. The mean values for each intensity level did not filter for any directional or positional tests, so the first plot displays a combination of parameters that were later found to negatively influence the results and may have compromised some of the experiment. The second figure, 5.10, shows the psychometric functions for 7 of the 10 participants due to the exclusion of participants who had data that could not be fitted to a curve. The PSE of the 5.11 comes to -0.88m, which is actually closer to the second intensity level of -1.25m, while figure 5.10, when calculating the mean of the 7 available participants, results in a PSE of -0.97m.

Because of the positioning of the participants, with the reflection area was to the left of the of the participant, the “center” visual source was also to the left side of the participant. It could be due to the leftward direction of the visual source that most participant’s PSEs were also biased to the left of the actual identical position. In figure 5.12, comparing the horizontal and vertical responses, it seems that more participants were able to properly identify when the sound was “up” rather than “right”, due to the lower degree of perceptual bias for the PSEs. An additional factor might be due to the diameter of the visual stimulus, 1m or 0.5m offset from the central point,

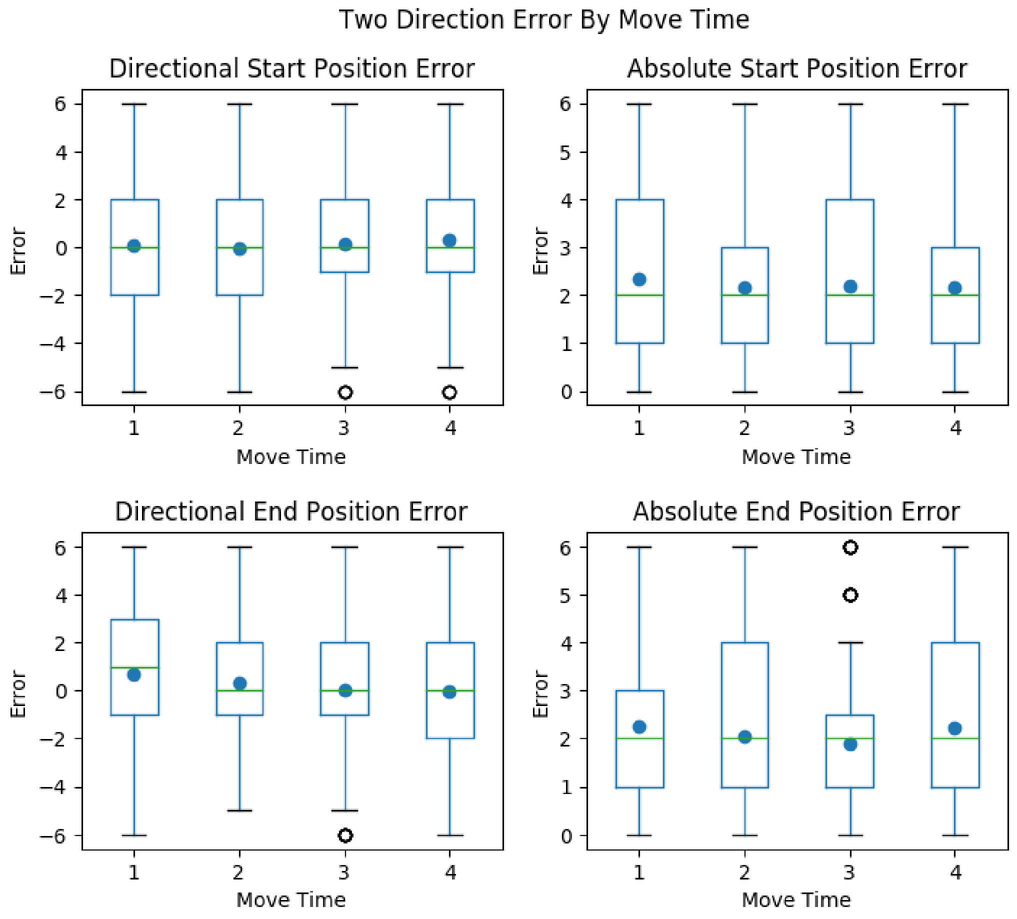


Figure 5.9: Box and Whisker plot of directional and absolute error grouped by movement time for the two direction experiment. Start error is shown on top and End error is shown on the bottom.

aiding in this shift because of the visual capture phenomenon. However, if this is truly the case, then the perceptual bias would also affect positions to the right and above the visual source still resulting in a symmetric curve.

Isolating the tests based on specific parameters, it became clear that certain combinations of auditory direction and visual position resulted in very erratic and often only a few users could even generate a curve fit. In order to represent all participants in the data and to try and parse out the outlying values, the combined mean values (without curves) at each direction and position can be seen in figure 5.13.

When looking across different participant's results, it became evident that some were very disoriented by the directional sound. Though the same users

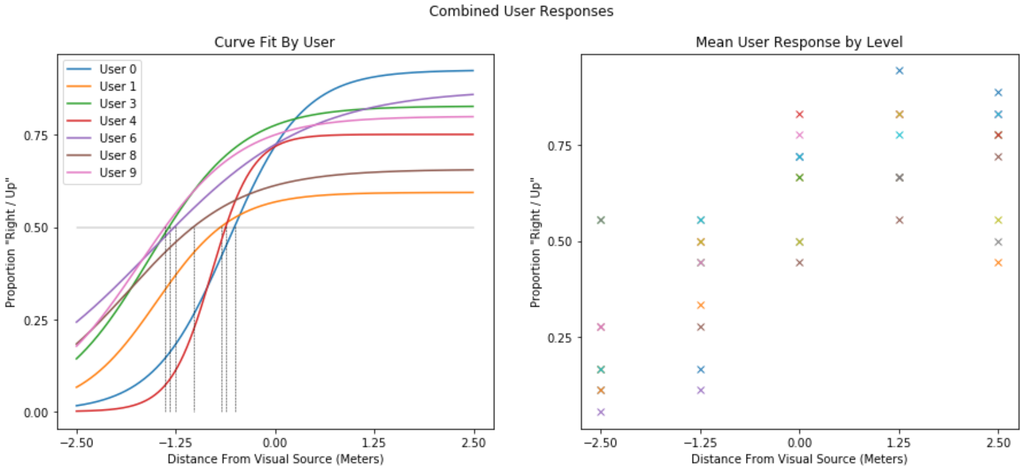


Figure 5.10: Generated curves for the majority of participants (left). Plot of all participant's mean responses for each intensity level (right).

participated in the directional tests prior to the constant stimuli experiment and all proved, more or less, that they could distinguish direction, some users were unable to accurately identify the location of the auditory source with one participant even responding with entirely opposite results: up was down and left was right. At the end of the experiment, most participants were not confident about their performance, however nobody reported any hearing problems or issues with being able to hear the auditory source. When a participant's responses were extreme, inverted or close to random guessing, it was not possible to generate a psychometric curve based on the recorded values, therefore different participants had to be excluded from the plots.

Overall, and in the face of many experimental problems, it is apparent that certain positions resulted in fundamentally flawed responses. In particular, nearly all responses given for the "right" position, in both horizontal and vertical directions, were erratic and often unusable for calculating a psychometric function. After returning to the experimental setup and running further tests with sounds in the "right" position, the problem causing these strange participant results revealed itself. When the speaker was aimed toward the rightmost side, especially when the sound source was positioned below the visual source, the subject's head would catch some of the sound before it reflected against the wall. After filtering the data to remove these aberrant values, there was still the problem of the leftward bias which as

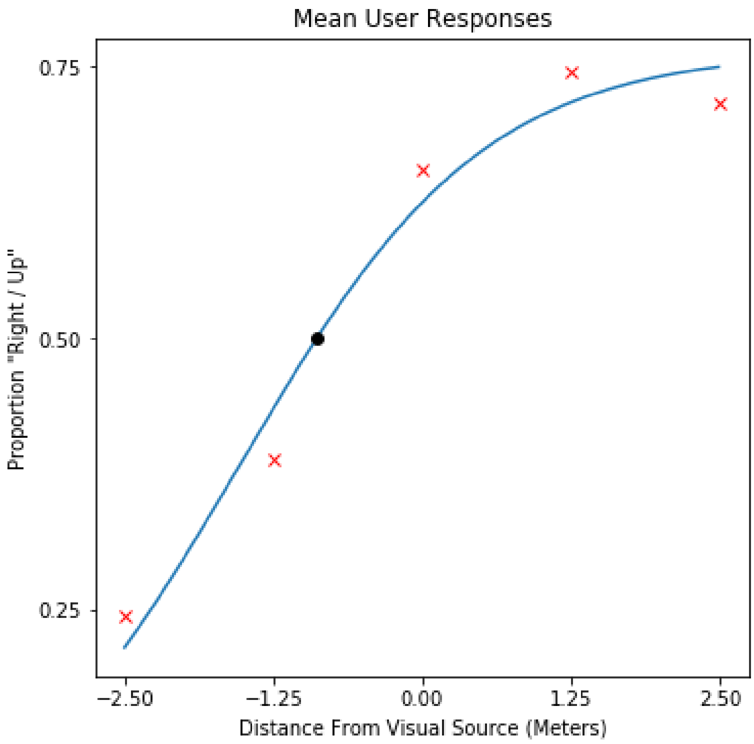


Figure 5.11: Curve calculated by averaging across users the intensity response scores.

mentioned before was likely due to the overall leftward positioning of the audio and visual sources to the participant. In the future, further experiments can be conducted to evaluate the effects of this positioning on the biased results of the experiment, particularly conducting duplicate experiments but with centered or rightward positioning of visual and auditory sources.

5.2 Empowerment of Users based on Prior Installation Data, Surveys, and Interviews

Collecting user tracking data across the earlier LargeSpace exhibition and the more recent CCD Memorial exhibition provided very detailed information about the positions of users who visited and participated in the BSD installation. The video recordings allowed each user's position on the screen, the x and y pixel values, to be recorded on a frame by frame basis, resulting in extremely detailed positional recordings. In fact, the high sampling rate

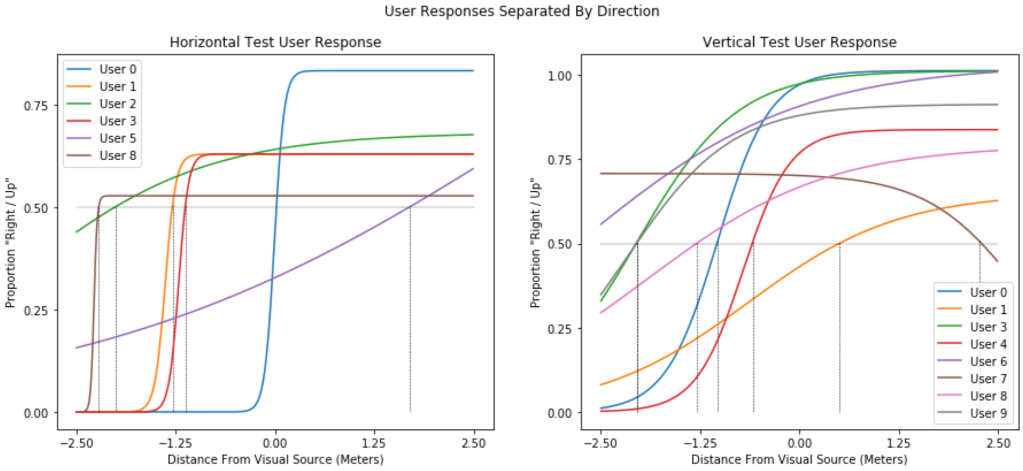


Figure 5.12: Separation of user responses by the direction of the audio source relative to the visual source.

of the recordings created very noisy tracking data which made it difficult to follow the movements of each user. Smoothing of the data was achieved using a rolling mean technique with the window of 50 frames (Pandas 2019). After preprocessing the data, plots were created showing the movements for the five selected users in each of the two installations (figures 5.14 and 5.15). Unfortunately, the differences in aspect ratio, perspective, and time scale, make it very difficult to compare the raw tracking data between the two installations. This is especially evident when considering the CCD Memorial installation user positions, in which nearly all vertical movement is focused in a small region of the plot. Because the documentation angle was so low to the ground, it compressed one of the dimensions of the user movement (the y-axis) so that it is incredibly difficult to evaluate each user's movement based solely on this data. While the recording and tracking of participant movement in the exhibition spaces was a useful exercise, the lack of consistency between the recordings made the data unusable and has taught the author an important lesson in regards to setting up video documentation for use in comparative studies.

Regarding the behavioral classification for user participation, which was also determined from the video documentation, the nature of identifying participant behavior was able to avoid many of the pitfalls that were problematic for the aforementioned video tracking data. As was described in the previous

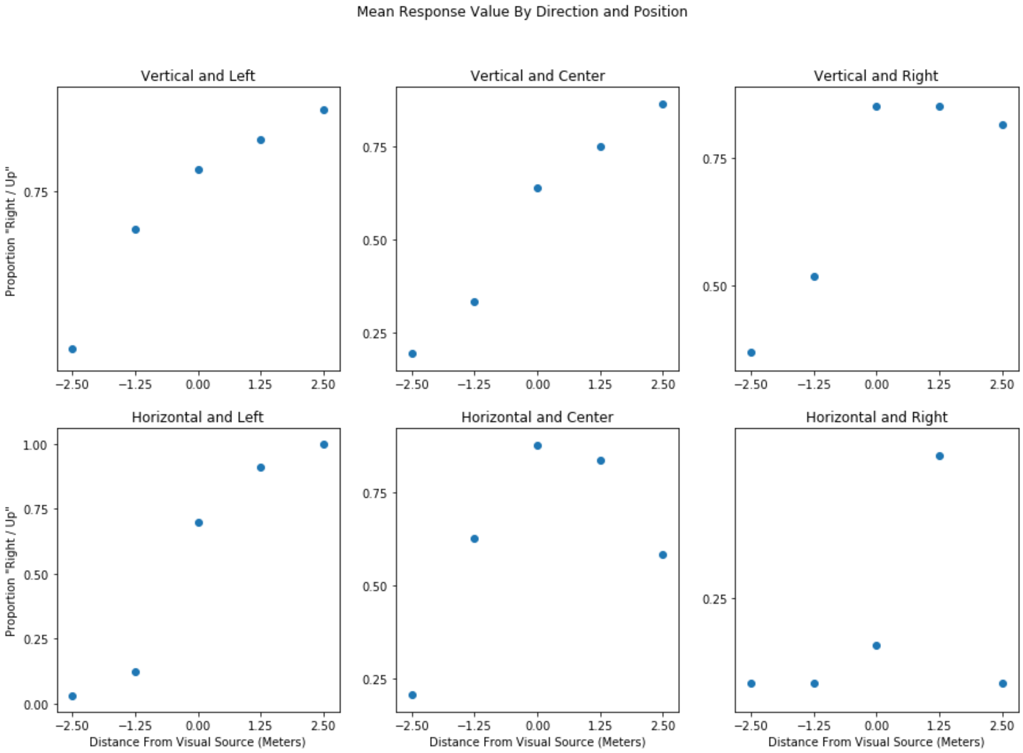


Figure 5.13: Plots showing the means for the full 6 conditions.

chapter, there were four behavioral classes: Interacting, Watching, Listening, Documenting, however when faced with annotating the video data with these categories, it was quickly apparent that the "Listening" class was nearly impossible identify, given the lighting conditions and relative size of users in the space, and was exchanged for a different class called "Socializing" in which participants were observed speaking to other participants. After aggregating the data for each class of behavior, and considering participant movement for each behavior action, it was possible to compare the duration of time spent for each class of behavior between the two installations. Shown in figure 5.16 it is clear that the LargeSpace installation receives much higher durational values than for the CCD installation. This is expected, as the LargeSpace installation required participants to remain in the space for the duration of the performance (roughly 15 minutes), while the CCD installation had no time constraints and participants spent, on average, between 4 and 6 minutes in the space. Interestingly, is that the CCD installation far outpaces the LargeSpace in the category of Documentation. When returning to the video documentation, there is a marked difference in the

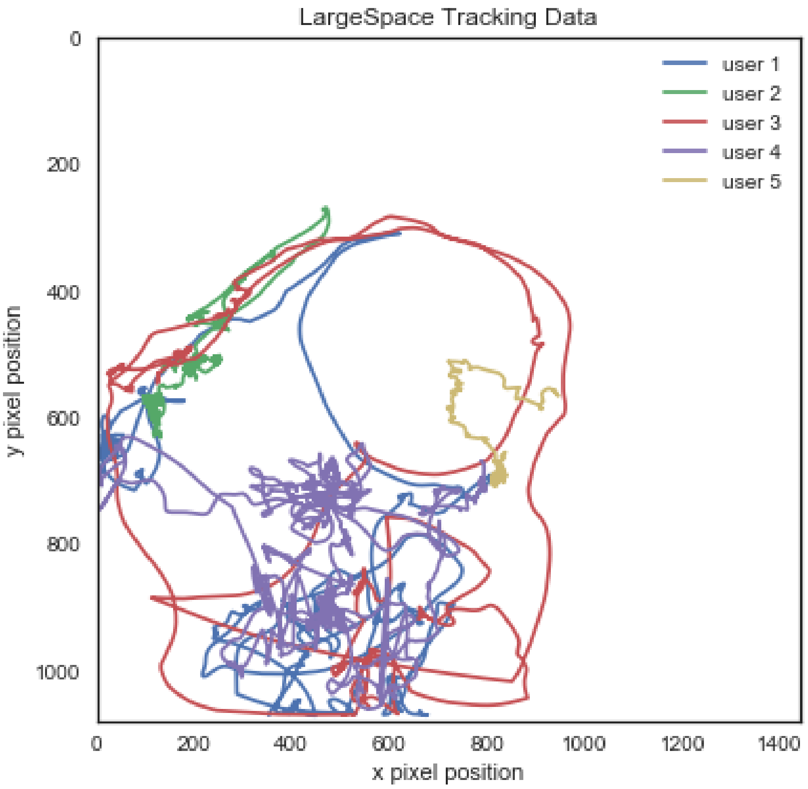


Figure 5.14: Smoothed tracking data for all LargeSpace users (1-5).

readiness for CCD users to take photographs and interact with smartphones than the users in the LargeSpace (in which only a single user contributed to this category). Relative to other categories, the CCD users are also more willing to socialize rather than interact with the work, whereas the ratio of interaction to socializing for LargeSpace users is inverted. While these behavioral differences are interesting to consider, they ultimately reflect the radical differences between the two installations and highlight the difficulty in comparing one installation to another using these factors alone. It begs the question as to whether there are more flexible or appropriate surveying methods for gauging participant interaction with artistic installations.

Finally, the questionnaires which participants filled out during the CCD installation can hopefully provide some insight into audience member perception of the SIS system used in the installation. Many of the survey questions focused on various levels of recognizing spatialized sound, from simply noticing that sound was spatialized (questions 1,2,3), to whether or not the user

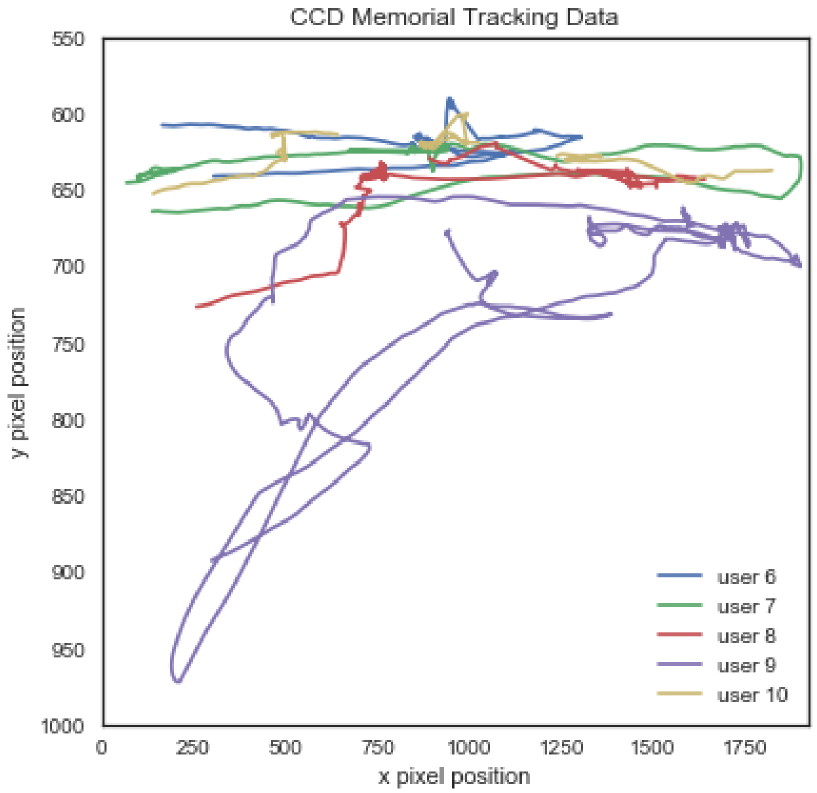


Figure 5.15: Smoothed tracking data for all CCD Memorial Users (6-10).

could control the positions of the sounds (question 4). Other questions were related to the thematic nature of the work (questions 6,8) and also directly asked participants if they felt more engaged with the work due to the sound spatialization (question 9). Across the majority of respondents all questions, except one, were marked with a score of 4 or 5 meaning that most participants strongly agreed with the majority of the questions. The only question which did not receive a "strongly agree" mark was question 4, which stated "I could control the movements of the sounds". In fact, because there was no direct user control over the locations of the sounds in the CCD installation, this is an accurate response. For the open response portion of the questionnaire, most answers were positive regarding the sound: "It makes the sounds interact perfectly with the voice and the movement", "The sound is very well appreciated", "Great work, brought back the connection of bird songs and human language, I heard about years ago". Other responses that were not as positive were related to difficulties in translation as the installation was only running with English instructions at the time: "There are no clear in-

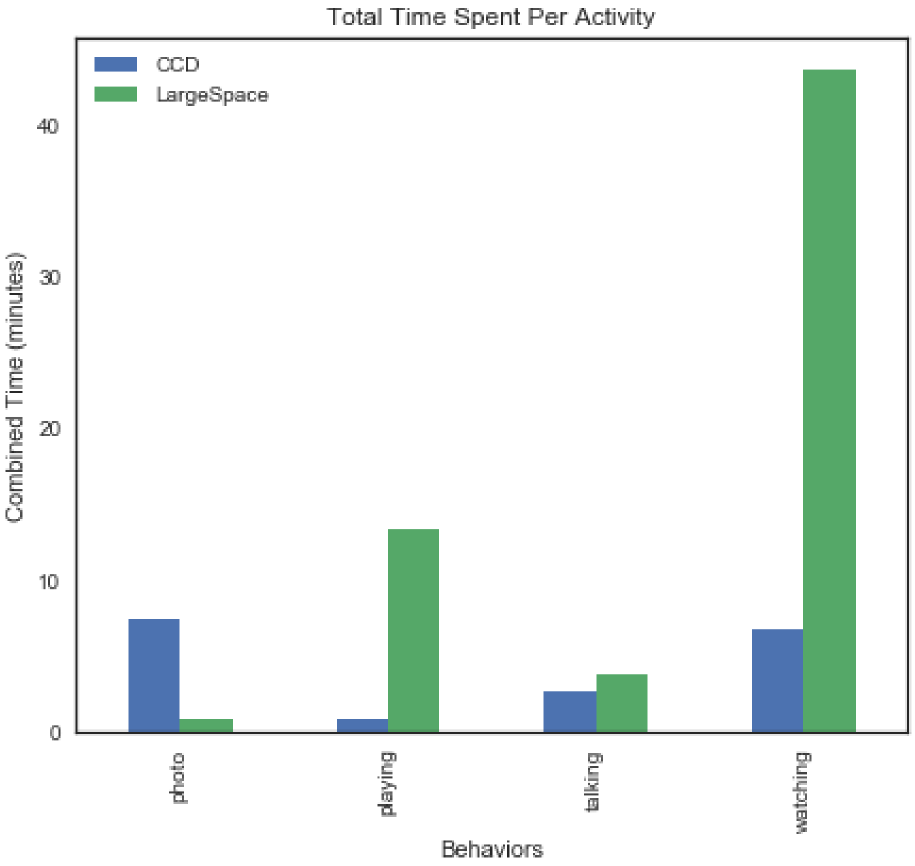


Figure 5.16: A comparison of time spent on each behavior between the CCD Memorial and LargeSpace.

dications of how the devices work in Spanish because of this detail, I did not understand anything about the piece", "I could not get involved with the exposure enough to control the sound with the microphones". Based on these survey responses, there is at least a modicum of evidence supporting the positive benefits for incorporating SIS in artistic installations.

5.3 Effectiveness of Spatially Immersive Sound System for Bird Song Diamond

As BSD was born as a work directly concerned with the acoustic relationships between bird, human, and the overall ecosystem, it naturally fits very well with SIS and has actually grown to take advantage of the capabilities of SIS in the more recent iterations of the project. While nearly every version of BSD has made use of parametric speakers for projecting spatialized sound into an exhibition space, the difficulty in manufacturing an effective experience using fixed position parametric speakers has always been a major challenge faced by those involved with the project. As described in **Chapter 3** many of the goals for developing the SIS system were directly related to problems from BSD, and aspects of the system were designed as a response to the difficulties in realizing certain aspects of the work. From the standpoint of the artist and designer of the audio portion of the work, the system provides an interface which is easy to develop for, and the Unity plugin provides further simplicity when designing the movements of the speakers. For audience members, it is evident from the survey responses in the previous section that there is an appreciation for spatialized sound and that it strengthens the thematic aspects of the work. The added dynamic nature of the sounds creates the perception of movement, whereas the fixed parametric speakers could only approximate movement. The SIS system has offered a large expansion in the sonic possibilities for future BSD exhibitions.

5.4 Contributions to Human Informatics and Potential Applications

The previously described experiments in movement and perception using the SIS system have sought to create a basis for which the potential of SIS within the field of Human Informatics can be realized. Though both of the experiments are very much initial developments into the effectiveness of moving parametric speakers, both of them are able to show that SIS is capable of reproducing sound movement and localization at a drastically reduced cost and number of components when compared to other methods of open-air sound localization (see **Chapter 2**). This cost reduction creates a nice middle ground between headphone related solutions, which can block out environmental sounds, and multi-speaker setups, which can be bulky and difficult to achieve high levels of localization without a large number of units. The above movement experiment indicated that most users are able to accurately discern the general direction of sound movement as it passes across their field of hearing, and even without extreme levels of directional precision, the added dimensions of directionality and movement can support many existing applications for sound in Human Informatics. And while the constant stimuli experiment was very much the beginning of a series of experiments regarding visual capture and SIS, it nevertheless supported a rough approximation for bimodal localization. This combination of sound and visual elements is a natural fit for fields of augmented and mixed reality, allowing additional layers of audio information to enhance human experiences in the world. Even further, the previous experiments have only dealt with a single moving parametric speaker, and the SIS system is equipped to handle many speakers. The creative potential for immersive audio environments using multiple SIS units has only been tested in the BSD project, but there are still many unrealized projects and applications which can implement a larger number of SIS speakers.

Chapter 6

Future Works

While the above discussion on spatially immersive sound remained in the context of the Bird Song Diamond project, further applications for SIS are possible within a variety of contexts and applications. Some other projects have already been presented in **Chapter 2** which make use of moving ultrasonic speakers for applications in performance and media art installations, increasing realism of augmented reality agents, and sound art installations. Applications for large scale performance, media art installations, and sound art naturally come from the above discussion in which SIS is able to enhance the immersion of audience members into a work by encouraging exploration and participation. The locative aspects of sound sources and the ability to direct and create sonic behavior within a space expands the possibilities for artists to express ideas. The ability to direct the movement of sound sources in space combines choreographic aspects to sound design that require addressing the acoustic and physical context of a work.

While the possibilities for artistic expression are one aspect, it is worth returning back to the functional and practical applications of localized audio. Spatially immersive sound is inherently a technique for creating immersion in virtual and mixed reality environments, especially larger spaces where SIS extols the exploratory potential of a space, however the flexibility of the system to expand or contract with a given environment means that SIS adapts well to a variety of implementations. The field of sonification and the development of auditory displays can profit from the small form factor of the parametric speaker for generating spatialized audio. Driving in a car, for instance, requires visual attention on the road ahead, so auditory cues are often used as non-visual indications. Localizing an auditory icon to a very specific location encodes additional directional information that might be useful to a driver, and achieving such localization without headphones, means that sound localization can occur without dampening other important environmental sounds (Nelson & Nilsson 1990). In Computer Supported Collaboratory Workspaces (CSCW), the development of teleconferencing tools for communicating with multiple remote participants faces issues when individual voices of the participants all come from the same position in the room. Spatially separating participant's voices can enhance recognition by listeners in a shared space, as well as reducing the overlap of similar sounding voices (Phua & Gan 1999). While, existing spatialization methods (**Chapter 2**) can be used in both of these applications, it is worth considering the unique features that parametric speakers offer based on the desired number of sources, the localization precision, and the spatial attributes that may restrict the size of a system.

6.1 Alternative Methods for Spatially Immersive Sound

In considering an expanded version of SIS, a number of alternative methods have been attempted by the author with the aim of creating the spatial engagement that separates *spatialized sound* from *spatially immersive sound*. The combination of location technology (GPS, Distance Relationships, Tracking Systems) and spatialized sound techniques outside of moveable parametric speaker systems can exhibit such possibilities. The following two projects have been developed in conjunction with the BSD SIS system, and try to engage with location and collaboration across space through the use of sound.

6.2 Head in the Clouds: GPS Audio Augmented Reality for Infrastructural Exploration

Network protocols, which enable computer-to-computer communication and thus mediate everything that travels over a digital network, perform silently and are often highly abstracted and simplified when a user receives a notification. Unless an application is specifically for network traffic analysis, it would be inappropriate for a developer or designer to include such traffic as part of its graphical interface. Yet to achieve a solid awareness of digital space and to develop abilities for traversing and modifying hybrid spaces, one cannot remain oblivious to the underlying structures and semaphore of digital communication. There exist tools for uncovering the particulars of networking traffic, yet are, for the most part, for system administrators and require some preexisting understanding of network jargon to make much sense of a stream of data (e.g. Wireshark, Carnivore Client, Little Snitch). These types of software are indispensable for monitoring and packet analysis. They can provide indirect reminders that routers and data centers are functioning properly, and can even describe the pathways taken by individual packets via commands like traceroute. However, these tools only function when a user explicitly wants to discover network traffic and consequently fail to represent the continuous existence of the infrastructural elements that facilitate such data transfer whether the user considers them or not. The creation of tools which bridge the gap between the immaterial and the material are becoming more and more necessary as companies seeking to simplify and streamline their devices continuously push the function and execution of algorithms, protocols, logic gates, and electromagnetism into the realm of mythology.

In the domain of auditory displays, most of the previously mentioned tools only provide a minimal amount of sonified information. These types of output are used to assist in monitoring a network and highlight important events such as a network aberration or a remote access request. As such, the sounds are discrete and meant to alert the user of an action that needs to be taken (Nees & Walker 2011). While successful at conveying that a network event has occurred, the audible components in isolation fail to provide any added understanding or insight regarding the overall networked system or its relationship to outside systems.

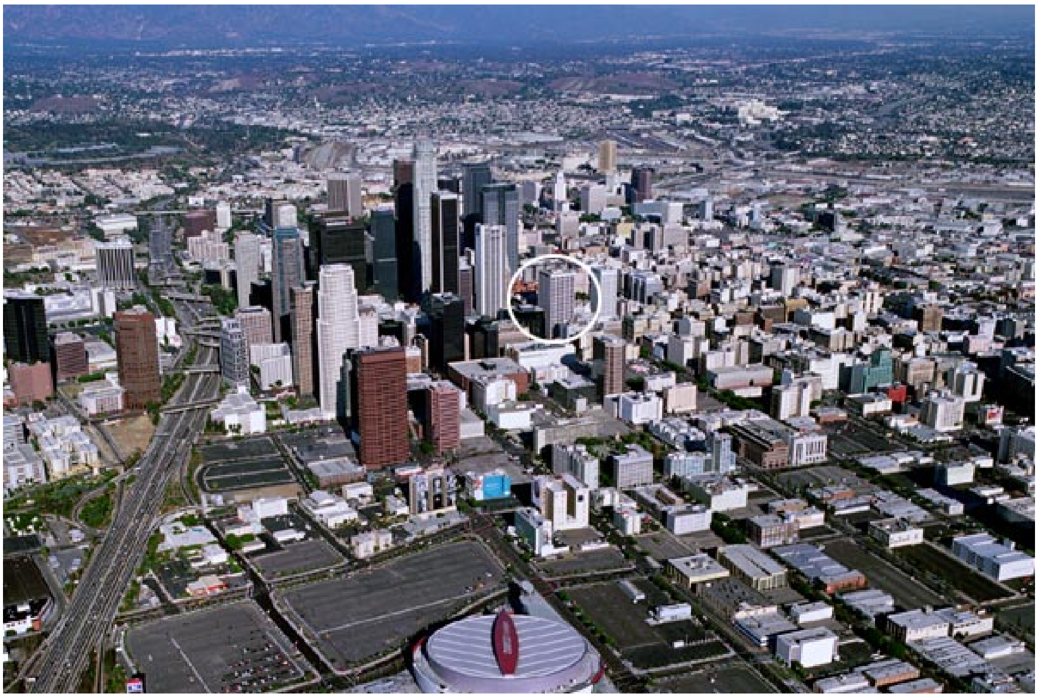


Figure 6.1: One Wilshire in Los Angeles houses one of the largest groupings of datacenters in the world

Even when successful at depicting the constant flow of data, monitoring software remains primarily within the physical boundaries of screen space (the bezel or frame of the screen). Ignoring the physical nodes and edges supporting our networks presents a detached and incomplete perspective for the user. It is not sufficient to simply provide the positions, addresses, or satellite views of the buildings (data centers) housing this equipment, as it fails to elicit an embodied relationship with infrastructure based on the physical constellation of the collected human and nonhuman actors. In the same way that the stepping points of a network are inert until activated (performed), the experience of the surrounding data centers (figure 6.1) should also be revealed through the active participation in space with those sites (Thrift 1997).

By using a portable device, such as a smart phone, to add perceivable characteristics to a data center *entity*, one can sense the presence of data centers and gain a more concrete understanding of their infrastructural surroundings. Further, rather than building a standalone application that can drive the entire experience from within the smart phone, an overview of existing apps can be found below, a web application requires pinging a nearby datacenter or repeater and is itself housed across hosting servers, performing handshakes, requests, and callbacks to activate the infrastructure that it uncloaks. Venturing into physical space to encounter data centers lets users achieve an experiential and embodied understanding of their relationship to infrastructure. Because sound can so readily fade into and out of our focus, yet remain ever present, it provides the most effective characteristic for a data center to virtually emit.

6.2.1 Sound Geographies

Using sound in conjunction with geography can be traced back quite far. While initial experiments using GPS, and what are now quite standardized audio annotation techniques, started in the late 1990s with the Hear&There project (Rozier & Karahalios 1999), there have been more analog approaches to exploring the sonic non-spaces within urban geography. Cage's *Imaginary Landscape no.4*, scored for 12 radios is a piece tied to the RF saturated information space that would soon be overshadowed by the more visually oriented cathode ray tube as the decade progressed (Cage 1960). Later, Cage would expand and complicate his sonified network system in his *Variations IV* by incorporating live, recorded, and broadcast sound all influencing and being influenced by the venue and moment of performance (Cage 1963).

More directly related to the city, sound and exploration, Max Neuhaus's sound walk performances, in which he would lead audiences through a city or town to discover the sounds of the everyday, actively engaging with the surrounding geography, but in a decidedly focused and particular way. Neuhaus sought to continue in the steps of the Italian Futurist, Luigi Russolo (Russolo et al. 1967), heightening and embracing the everyday clatter that is so often tuned out as noise. After spending an afternoon visiting power plants and other scenes, once primed to receive everyday sound, a listener could leave with a new, empowered set of ears. Pauline Oliveros would further this notion into a fully developed practice and aesthetic known as Deep Listening

(Oliveros 2005).

While not directly inserting or overlaying a media object or sensation between the receiver and source, as is traditionally the case with augmented reality, these composers championed methods for focusing one's consciousness to discover sounds that would otherwise not exist in the conscious mind of the listener. Beyond sound, one could also point to many other artists that sought similar goals with different mediums (Berger et al. 1973; Plant 2006).

Turning back to more conventional methods of augmented reality, recent years have seen a number of mobile phone applications aimed at providing the public with ways to incorporate sound into their environments. For the most part these are standalone apps which enable users to place sounds at particular locations, either to set up a personalized tour, guided walks, or artistic experiences. Designed to be trivial to pick up by non-technical users, this creates many limitations in the experiences that can be designed. Nearly every application requires manual placement of audio, either virtually or by recording it while at a location, therefore scope is extremely limited by a lack of API or open codebase.

More interesting examples emerge in the forms of alternative reality games which often use customized hardware and software to create location based augmented reality experiences. In particular, the group Blast Theory has created a variety of audio-locative experiences in the form of personal narratives, complex puzzles and mysteries, and high intensity chases. (BlastTheory 2007).

The aim of this project is to simultaneously develop a more open and available toolkit for mobile web developers to construct more customized, larger scale, and complex intersections of audio and location. Additionally, the development of a self-referential work that embraces the historical explorative motivation of midcentury sound artists is important for demonstrating the capabilities of the system.

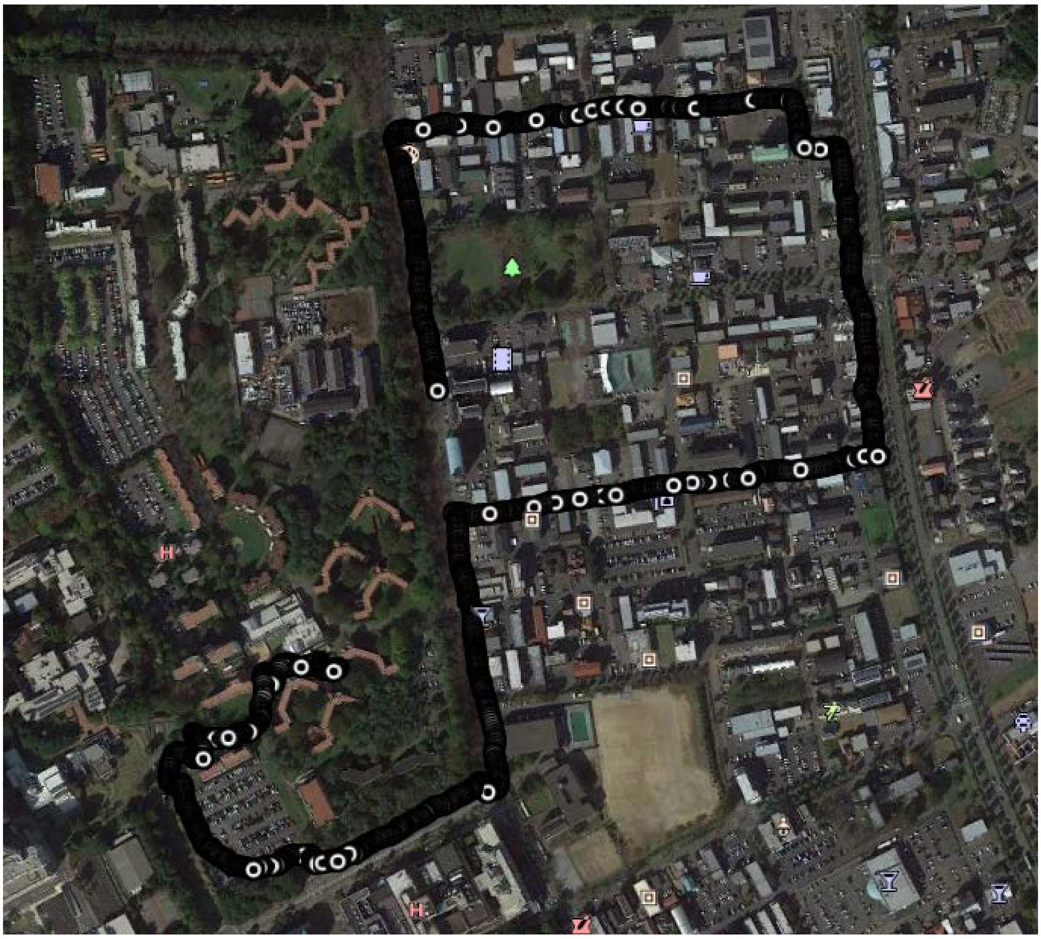


Figure 6.2: Early test recording the path of a user moving while using the application.

6.2.2 Application

Adding audio characteristics to data centers is achieved through the use of stereo headphones attached to a smartphone. The phone must be capable of running a major mobile browser (Chrome, Safari Mobile). The web application is the primary method for connecting a user's position (location and heading) to an existing database of actual datacenters. Other than the initial over-the-air download, and occasional check when a user has moved a large distance, the user can properly use the app within the web browser with limited-to-no browser connectivity. Figure 6.2 shows an early test at recording user movement.

In developing the application, we tested a number of systems for constructing the virtual space as currently there are no web-based augmented reality frameworks that support spatialized audio. After attempting to build the entire web application from scratch using only browser supported JavaScript APIs (WebAudio, Geolocation, and DeviceOrientation), and then using a combination of an audio library and JavaScript APIs, it was found that, depending on the device, the browser, and even the version of the browser, the application could behave erratically or provide incorrect location, orientation, or audio data. Eventually, we found a JavaScript game engine called Babylon.js which provided enough stability across devices to enable further system testing (Catuhe et al. 2018). This engine allowed for faster prototyping due to convenience functions in vector math, audio loading, and spatialized audio. The lack of additional plugins continued with the conceptual thread of the entire application being delivered through the browser immediately, without having to download plugins or register with a service. Possible alternatives that would violate this requirement, would be the Unity Web Player, however location and orientation services may still require additional libraries.

While Babylon.js offered phone orientation functionality, it was found that using a separate library for orientation, Full-Tilt.js (adtile 2018) would offer more reliable orientation as well as multiple checks for the fragmented device and browser support that plagues mobile development. Unfortunately, even this library was not fully up to date, and it was necessary to add support for a different orientation API used in newer versions of the Google Chrome browser. This was to take advantage of“absolute orientation” which allows for user orientation in relation to magnetic north. Many browsers now default to “relative orientation” which gives user orientation based on the direction that he or she was facing when starting the application. The rest of the web application uses standard HTML, CSS, and JavaScript with additional JavaScript libraries for formatting and interactivity.

6.2.3 Location

When a user connects to the application, the server queries an up-to-date repository of publicly listed data centers throughout the world, then checks against a database stored on the web app’s server for any changes in data center locations. To reduce latency for the end user, the server returns a JSON object with a subset of centers filtered based on the user's location. In order to provide a user's position, the user must grant the application permission to access her device's GPS and orientation sensors. Once the locations of both the user and the data centers have been retrieved, the client side of the application can initialize the audio.

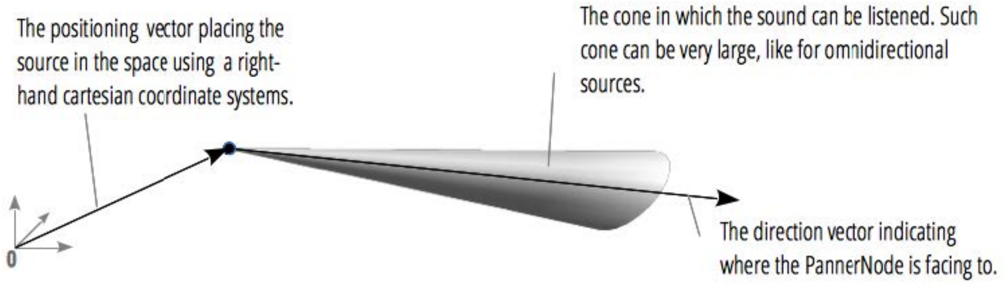

The positioning vector placing the source in the space using a right-hand cartesian cocrdinate systems.
The cone in which the sound can be listened. Such cone can be very large, like for omnidirectional sources.
The direction vector indicating where the PannerNode is facing to.
0

Figure 6.3: The *PannerNode* object takes multiple parameters. This project specifies a position while leaving the sound cone omnidirectional. Image by Mozilla Contributors is licensed under CCBY- SA 2.5.

6.2.4 Audio

All audio is generated and controlled using the JavaScript Web Audio API which allows web-based applications to take advantage of much more sophisticated control over audio without the use of plugins. Following the location data retrieval, an audio context is initialized and the position of the user, a vector containing latitude and longitude coordinates, is assigned to the main *AudioListener* object. Additionally, data collected from device orientation sensors is processed to produce a compass heading ranging from 0 to 360 degrees and assigned to the *AudioListener* orientation. Both of these values are updated whenever new sensor data is made available.

For each data center, a *PannerNode* is created based on its respective global coordinates. The directionality of sound moving away from the node is controlled by the shape of a sound cone shown in figure 6.3, however in this project all nodes are set to be omnidirectional. The node on its own does not emit any audio, therefore once a node is created it will immediately be assigned a unique sound.

With both an *AudioListener* and at least one Panner- Node created, any audio produced by a node will be filtered and attenuated based on the orientation of the listener and the distance model of the respective node. The *PannerNode*s will remain fixed, while all changes in audio are determined by the listener's change of position and orientation. As the user moves through a region, whether by train, car, or on foot, the levels and positions of the audio will shift and change, creating unique mixes of the various centers within that region.

The head-related transfer function (HRTF) used by the Web Audio API to provide accurate spatial discrimination for the user, using impulse responses from human subjects, did not create a significant enough difference between sounds located directly in front of a user and sounds located directly behind a user. In systems where small changes in position provide clear indications regarding the distance of a user to a sound object, the problem of inverted direction can be detected and corrected quickly by the user (Hermann et al. 2011). However, when the amplitude falloff of a sound may only become noticeable after hundreds of meters of movement, as in this system, clarity of a sound object's spatial position will be the primary source of information used in navigation. Because of this, further processing of the sound based on the orientation is necessary to prevent misunderstandings and frustration for the system's users.

The decision to have such parameter mapping, to represent extremely slow and gradual auditory feedback for users of the system, is reflective of the geographic size of the region exposed by the auditory display (figure 6.4). Given such a slow feedback loop, and provided that the collective data center soundscape maintains an amount of interest, the full sonic capacity of an individual data center can slowly emerge. A visual analog to this experience might be driving toward a benign looking mountain, far along the horizon, and gradually realizing its imposing stature as you move ever closer. Rather than downplay the importance and presence which data centers hold in our lives by using a more immediate audio perception action-loop, their importance is magnified through the energy we must expend in order to significantly modify the sonification model. This highlights the affective nature of spatialized sound by combining the auditory qualities of the sound sources (see below) and the performative nature of such a labor-intensive perception-action loop. However, it is also possible that because the interaction design of this system goes against the transparency between action and effect which characterize most successful sonic interaction designs, that participants may become frustrated or disinterested by lack of immediate feedback.

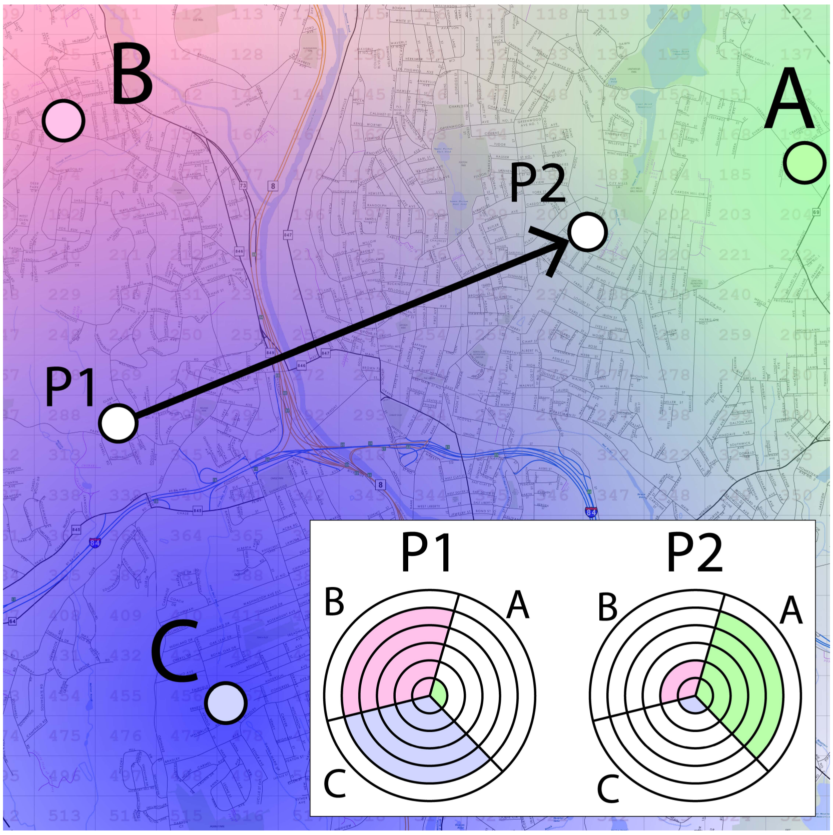
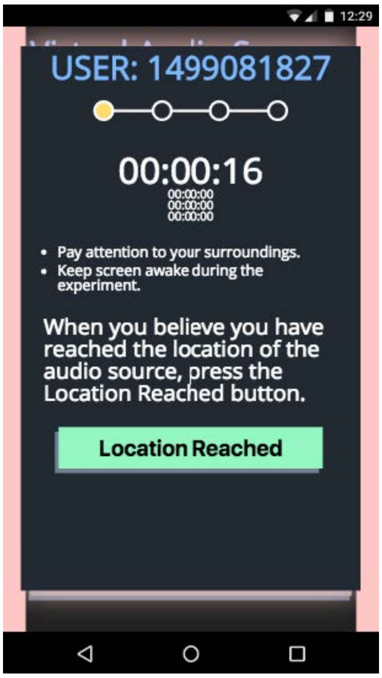


Figure 6.4: A representation of the change in sound influence of data centers A, B, and C as a listener moves from points P1 to P2. Each center has a color gradient which shows the size of its sound influence and falloff based on a user's distance from the center. The diagrams show a rough approximation of the loudness and spatialized position of the centers for the two marked points.

6.2.5 Testing

After developing the application, it was necessary to determine the accuracy of audio wayfinding in comparison to visual and tactile modalities. For an initial test, a modified application (figure 6.5) was created, which made used all three types of feedback and allowed users to guess which direction a particular source was coming from. Each participant guesses ten different orientations for each feedback type for thirty guesses in total.



USER: 1499081827

00:00:16

00:00:00
00:00:00
00:00:00

- Pay attention to your surroundings.
- Keep screen awake during the experiment.

When you believe you have reached the location of the audio source, press the Location Reached button.

Figure 6.5: The user interface for a future experiment for testing the effectiveness of both orientation and position in the field.

To isolate orientation from position, the application maintains a fixe distance from the source to the listener. Each time a participant guesses the orientation of a source, a new source is semi-randomly generated at a location surrounding the listener, with randomization biased toward selecting new source positions with larger angular distance away from the participant's current heading. The order of the different feedback types is chosen at random and each type uses all ten guesses before moving on to the next type.

To maintain consistency between tests, the same device and headphones are used for each participant (Samsung Galaxy S7 Smart Phone and Sennheiser HD 280 Over-Ear Headphones). Further, the testing facility was an empty, quiet room with no indications of heading, where users could stand and freely rotate in order to determine the correct orientation. The web application displayed instructions, further clarification could be given by an assistant, and the user can begin the experiment whenever she wants. The technique for audio spatialization is the same as explained above in which an HRTF, *AudioListener*, and *PannerNode* is used. During non-visual portions of the test, the screen is black except for a button for a user to select decide when he believes that he is facing the target source.

For the tactile feedback portion of the experiment, the screen is also blank with a single button. Tactile feedback is given via the use of the JavaScript Vibration API, however the API only offers the ability to control whether or not vibration is on or off. Additionally, the latency involved in this mobile application makes high speed PWM an unsatisfactory solution. Slower PWM is possible, therefore the angular distance between the participant's orientation and the target source's position is translated to the frequency between vibration pulses. These pulses vary from 60-2000 milliseconds between pulses when a user is between 0-180 degrees from the target, respectively.

The visual portions of the test turn on a virtual, directional, light in the virtual space that allows users to rotate the screen of the phone in order to find a sphere at the correct location. The only other element on the screen is the button for indicating an orientation guess (figure 6.6).

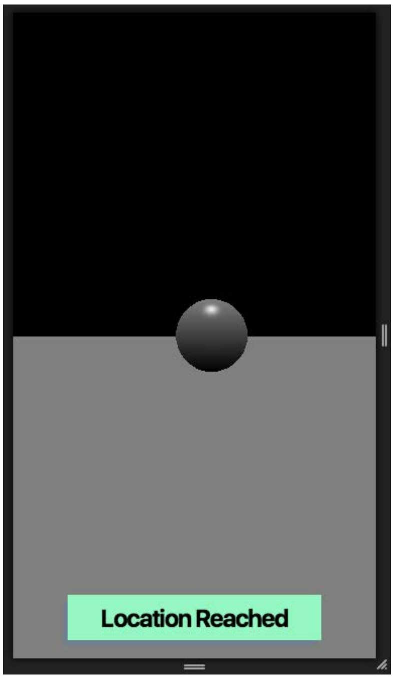


Figure 6.6: The visual portion of the orientation experiment, where a user guesses when they are facing a randomized target.

6.2.6 Results and Discussion

Preliminary results from a number of individual tests can be seen in the figures 6.7 , 6.8, 6.9 show results from the visual, audible, and tactile feedback orientation tests. Divided by feedback type, the graphs represent the time and angular distance to the targets (upper graph) followed by the time in seconds to complete each stage of the test for each feedback type (lower graph).

These tests show a distinct difference in accuracy and behavior from the sound-based feedback versus the visual and tactile feedback. While the times to complete the stages remained roughly equivalent between types, it is clear from the plotted points of the angular distances over time that the accuracy of estimation is diminished for audio feedback. It is also possible to compare the shapes of the plots for each stage, where audio-based feedback presents a strikingly different and irregular path as the user moves toward the goal. The visual and tactile graphs have much more patterned curves.

A greater difficulty for accurately locating audio elements is not surprising as the ear more often aides in a more general aiming process which the human eye then completes. The success shown by the tactile feedback is more unexpected in both time and accuracy. Because there have only been a limited number of participant in the tests there cannot be a more generalized result, so further tests will need to be completed.

Given a greater difficulty for audio-only positioning, it helps to show that an audio-based feedback system is the most appropriate for this project. In order for the work to highlight the ambiguity of the datacenter, it should not give a precise and easy to ascertain location; a GPS unit, map, or turn-by-turn navigation system fulfills this need quite well. Instead it forces a listener to engage further with her surroundings, to consider them, to determine her own route by improvising over, under, and around obstacles. Audio allows for the use of other senses, and in the case when a user is walking or driving through an active environment, it is important to afford the user his awareness of such an environment rather than pull him out of it.

6.2.7 Sound

The sounds attached to each *PannerNode* are created from modified audio samples with adjustments to the buffer speeds of the samples based on the relationship of the listener to the data center and other listeners. Historically, there have been many examples of communication devices being used as tools for extracting sounds, either incidental to the device or indicative of the network's medium. From Thomas Watson listening to natural radio through a telephone wire to works by Paul DeMarinis, e.g. *Rome to Tripoli* (Ouzounian & DeMarinis 2010)), these examples both channel and examine the natural energy of electromagnetism that saturates our atmosphere (Kahn 2013). With the current dominance of digital communication protocols adding a layer of separation from the analog, it becomes necessary to also explore the underlying components that drive digital communication. Therefore, the methods for generating sounds which intend to explore the energetic activity of contemporary networks, in the same vein as the above examples, should similarly embrace the systems, protocols, and physical components which underpin them. Because the focus of this project is to consider the spatial relationships of physical bodies and structures, the sounds draw from the hardware and mechanical qualities of the network rather than exploring the higher-level messaging protocols or software.

Upon visiting a data center, the visual stillness of the racks of servers, the neatly strung cables, and the uniform fluorescent lighting was immediately overshadowed by the filtered noise of the arrays of fans simultaneously pumping air through the machines. It brought to mind the dichotomy of the static and solid external appearance of the computer against the inner chaos of the CPU, GPU, and hard drive produce billions of operations per second. This relationship of an unremarkable exterior belying chaotic internal activity informed the process of creation for the sounds in this project.

Each sound begins as a simple audible waveform, with a relatively low frequency, recorded into a buffer. The speed of the sample is then multiplied by a factor to bring the frequency of the waveform far above the range of human hearing (20-20,000Hz) and into the same frequency spectrum of computer hardware operation speeds (MHz, GHz). These goal frequencies are determined by various milestones within the history of computing systems, for example the original IBM PC had a clock speed of 4.77MHz. The radical shift in frequency, coupled with the degradation of the sample inherent in the process of modifying the sound, results in a drone of subharmonics both chaotic and stable.

Rather than sonifying the user's position and orientation information solely from the perspective of communication efficiency, using beacon chimes or spoken descriptions of the occurrences at each location, the sonic nature of the overall soundscape depicts the continuous activity of a large network ecosystem. This presents problems regarding audio stream segregation when many different data centers are emitting audio in close range to one another. Though spatialization is usually aids in the process of separating sounds (Neuhoff, 2013), the complex and continuous nature of the sounds require further assistance with differentiation of each audio stream. A solution is found in the bioacoustics of rainforests, where the large numbers of vocalizing animals create a highly-crowded frequency spectrum. Rainforest species have adapted to occupy "niche" frequency bands within the overall audio spectrum, which allows them to communicate with others of the same species within their own audio territories (Krause, 2011). Just as different species of animals within a rainforest differentiate their sounds using their own species-specific frequency bands, the different datacenters of a metropolitan region can each inhabit their own bands within the frequency spectrum. Combining virtual sound spatialization with bioacoustics inspired frequency differentiation will aid users in pinpointing each unique sound source while not detracting from the overall auditory scene.

By creating a system that enables data centers to emit virtual sounds across large distances, it is hoped that users will consider the constant presence that these complex entities maintain on all aspects of global networked communication. Additionally, this system provides a very specific implementation of a more generalized system for creating audio based augmented realities. More possibilities exist for local, regional, continental, and worldwide installations using this system. Further exploration of these variations in scale as well as new contexts that, with the help of SIS techniques, can gain additional meaning.

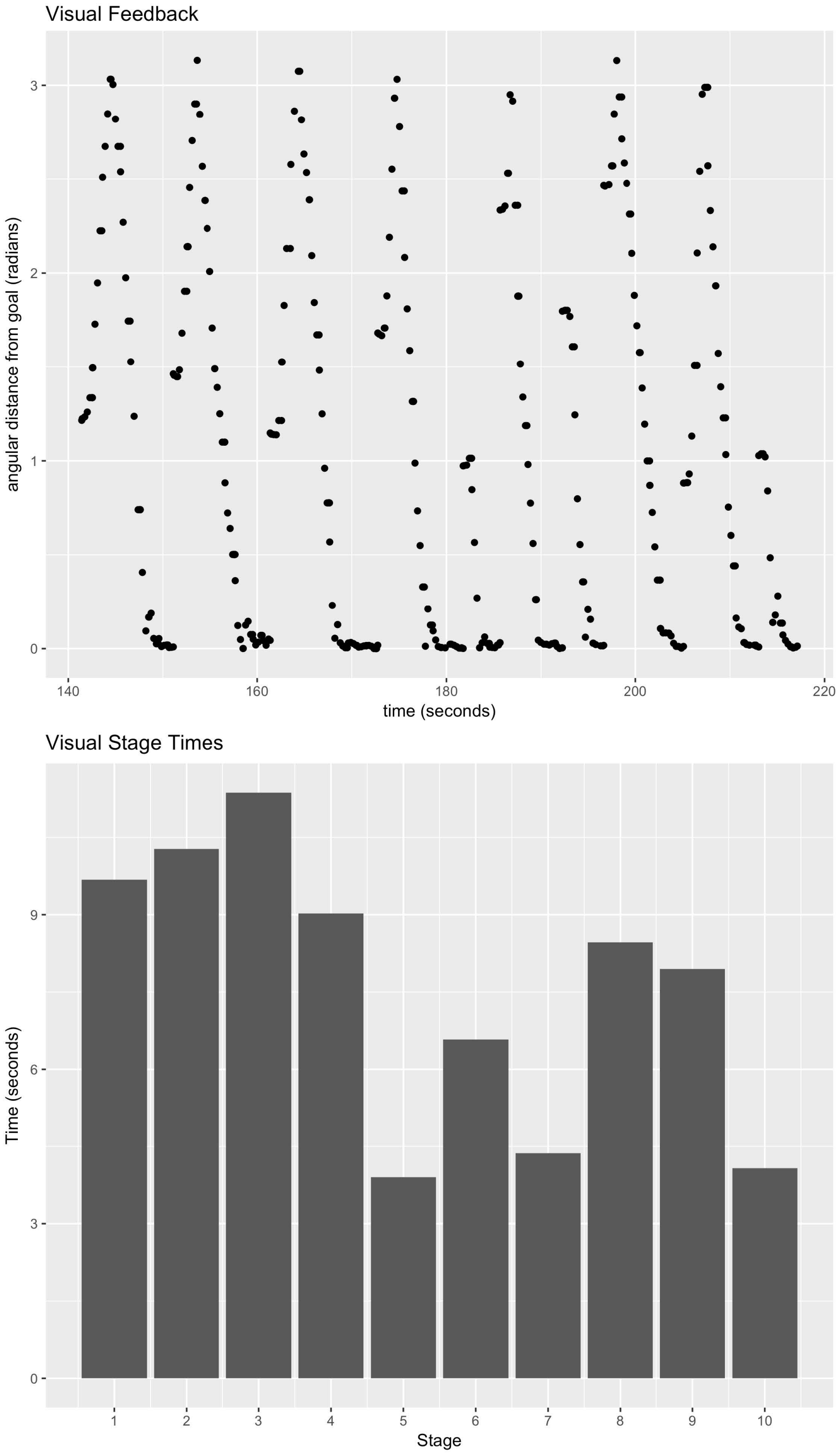


Figure 6.7: The graph at the top represents the angular distance of a user over time for the *Visual* test. The graph on the bottom corresponds to the time it took the user to guess each orientation for the same test as the graph on the top.

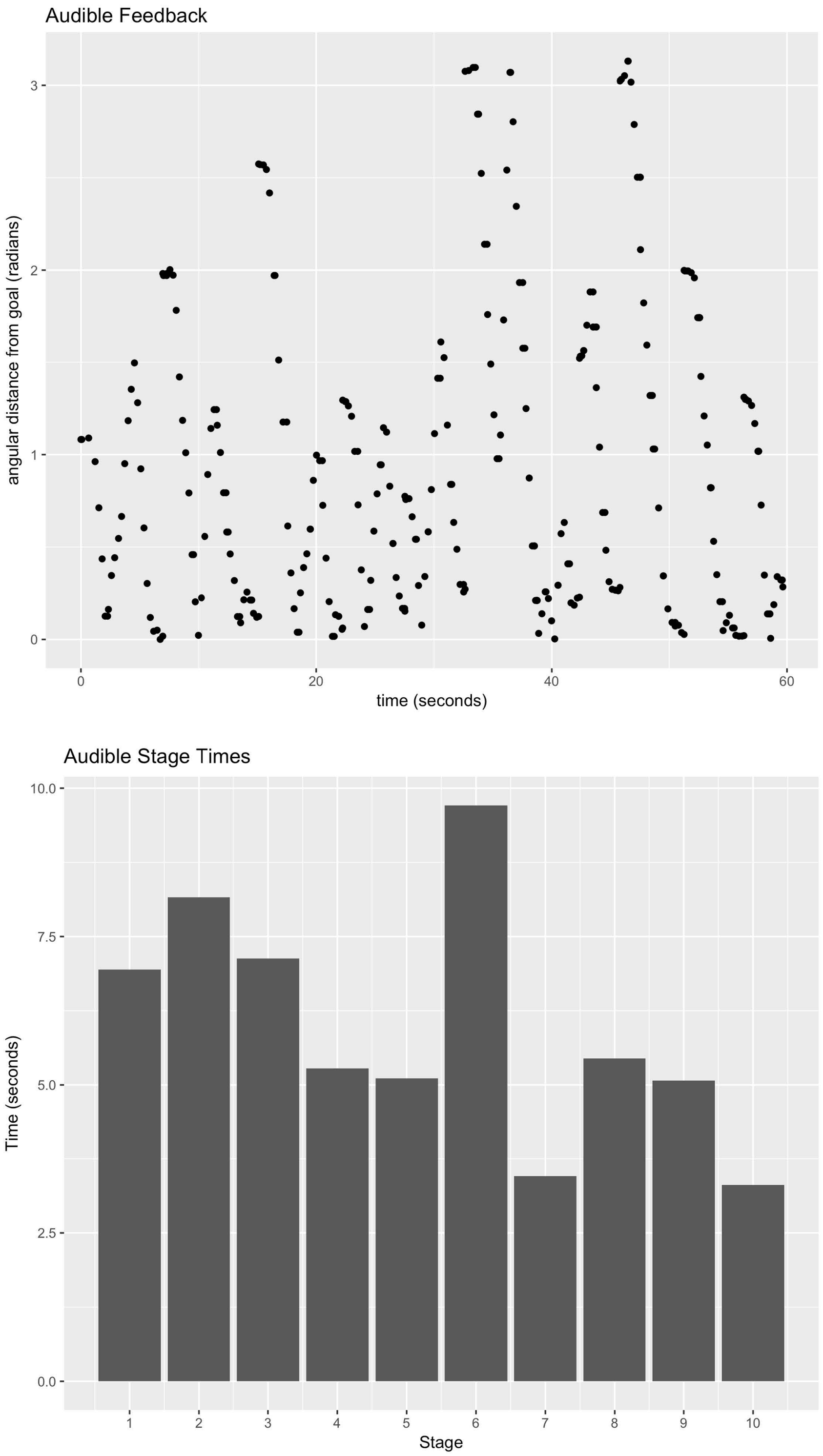
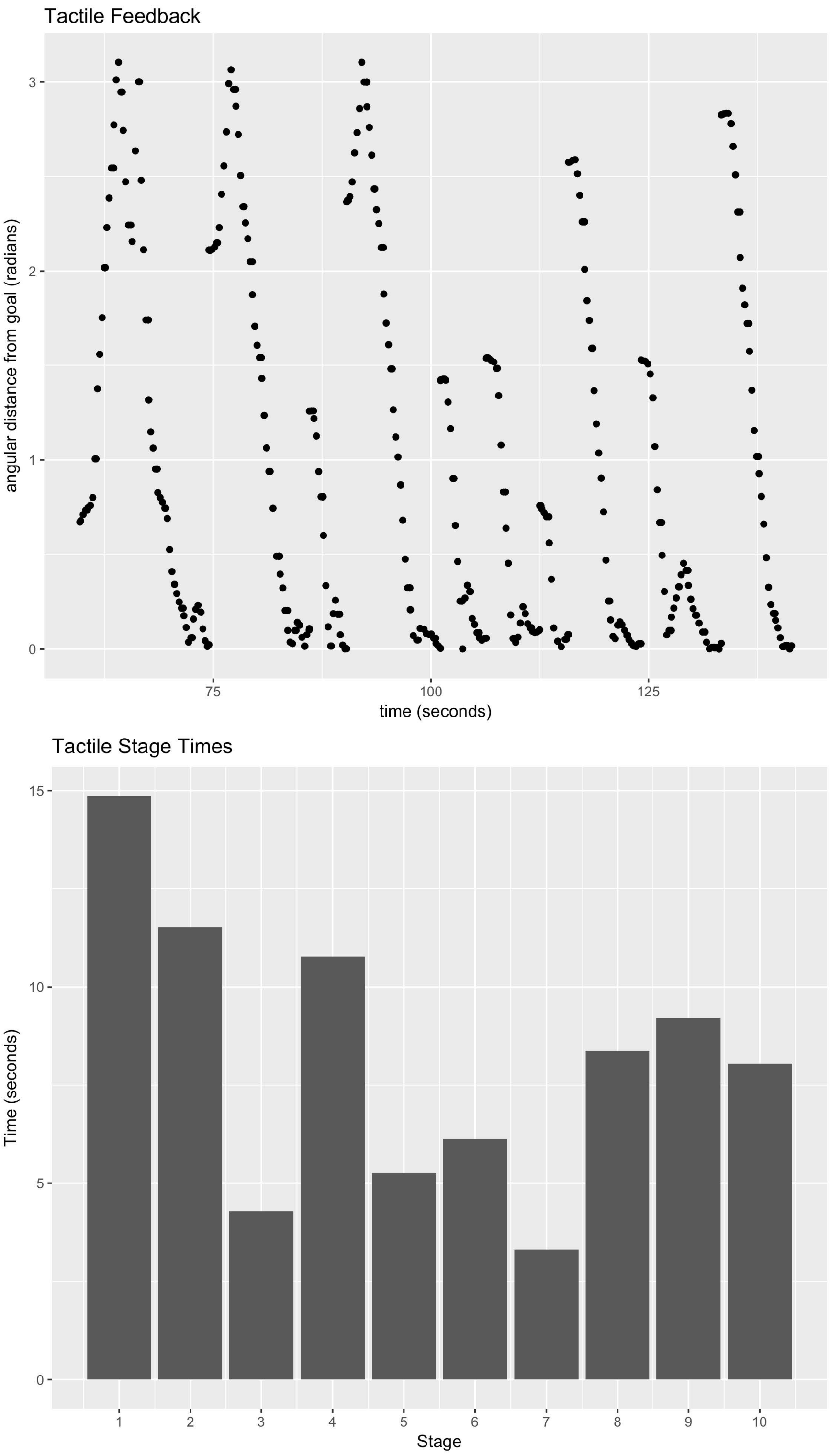


Figure 6.8: The graph at the top represents the angular distance of a user over time for the *Audible* test. The graph on the bottom corresponds to the time it took the user to guess each orientation for the same test as the graph on the top.


Tactile Feedback

Tactile Stage Times

Figure 6.9: The graph at the top represents the angular distance of a user over time for the *Tactile* test. The graph on the bottom corresponds to the time it took the user to guess each orientation for the same test as the graph on the top.

Chapter 7

Conclusion

Over the course of this text, there have been three main concerns related to the BSD project and SIS sound. The first was to give an introduction to BSD covering the history of the project and its exhibitions, the underlying research that forms the other side of the dialogue of the project, and the ways in which the project connected with SIS. The second concern covered SIS, from a definition that outlines the history of immersion in sound and technologies used to create sound localization, to prior research using similar systems. This background information allowed a detailed account of the SIS system developed for BSD, covering the multiple versions of development and design changes, the full hardware-software process was elucidated. The third concern was how to properly evaluate SIS systems, particularly when the system is used within a larger artwork such as BSD, and to also consider system evaluation at a more basic level to gauge the efficacy of SIS on participants. A two pronged approach became the method used here, in which standardized evaluation methods were juxtaposed with methods taken from ethnographic studies; providing both a controlled setting for experiments and in situ, public testing.

Following these primary concerns, considerations for the future of SIS systems and strategies in relation to Human Informatics as well as other applications was covered. A full account of an existing system, currently in development by the author, using headphones rather than parametric speakers for ubiquitous SIS over large scale regions was provided to highlight a combination of applications practical, the use of SIS for navigation, and poetic, giving voices to datacenters. While the disciplines and backgrounds at the heart of the project are not always seen hand in hand, it is significant to mention that without the glut of input from these cultures that the questions and demands needed to develop such a system as SIS would not have been posed.

7.1 Contributions of Bird Song Diamond to Human Informatics

The Bird Song Diamond project, as previously described, emerged from cross disciplinary collaboration, where both researchers and artists approached the collaboration with the intent to contribute their expertise in order to elevate and advance a common research question; though methodologies differed, the multifaceted approach allowed for oblique and novel approaches. As Human Informatics strives to draw together various disciplines with the aim of bettering the functioning, human-environmental relationship, and quality of life for human beings, it can look towards the BSD project, and the underlying bird song research project, as an instance of drawing on influences far outside of the standard avenues of cross-disciplinary collaboration. While travelling across the hallway can be a small step for collaborative efforts, some of the more novel approaches and unexpected ideas may only emerge after leaping across campus.

Another contribution which BSD can lend to Human Informatics is an outward looking position regarding advancements in human understanding. BSD begins with research on bird song, but very quickly appropriates this research for the sake of human understanding. One of the primary goals of BSD is to have installation visitors experience bird song, imagery of environmental and man-made perspectives, and participate in bird-like behavior, so that upon leaving the installation space, the visitor gains a new appreciation for the ecological implications of their own activities. Human Informatics often looks directly at the human for answers, through sensors, physiology, psychology, etc., but what it can gain from BSD is to also consider the non-human and the larger ecological implications of the goals of the field.

7.2 Empowering Capabilities of Spatially Immersive Sound

A number of potential applications for SIS have been proposed, whether already existing, in the form of enhancing artistic installations or adding realism to a projected creature, to theoretical applications, such as spatialized alerts when driving and added directionality in the context of teleconferencing. It is apparent that if SIS, and the proposed SIS system, functions to a high enough degree, that it can truly empower users in a wide variety of settings across a variety of fields. The experiments determining SIS performance and user perception regarding directionality and bimodal stimuli demonstrated that while there is still much to be improved with the system, it currently functions at an acceptable level for many different applications. While primarily concerned with the implementation of SIS in the BSD project, giving much more weight to the creativity extending capabilities of SIS, it was shown to produce very positive responses from individuals experiencing such a system for the first time. It will only be a matter of time before SIS can be added to a larger variety of artworks, so that the effectiveness in other artistic contexts can be assessed. SIS by itself has yet to be given its own solo platform, that is a completely SIS only work, which again will be a good test of the system's abilities to support immersive experiences.

Appendix 1:

Bird Song Diamond Questionnaire

On a scale from **1 to 5**, with **1** being "Strongly Disagree" and **5** being "Strongly Agree" please answer the following questions:

1. The sound coming from the interior speakers enhanced the experience.

1 2 3 4 5

2. The sounds were different depending on my location.

1 2 3 4 5

3. The locations of the sounds changed over time.

1 2 3 4 5

4. I could control the movement of the sounds.

1 2 3 4 5

5. I enjoyed interacting with sounds at different locations.

1 2 3 4 5

6. I felt a connection between bird song and the spatialized sound

1 2 3 4 5

7. The spatialized sound enhanced the overall experience.

1 2 3 4 5

8. The spatialized sound connected with the theme of the work.

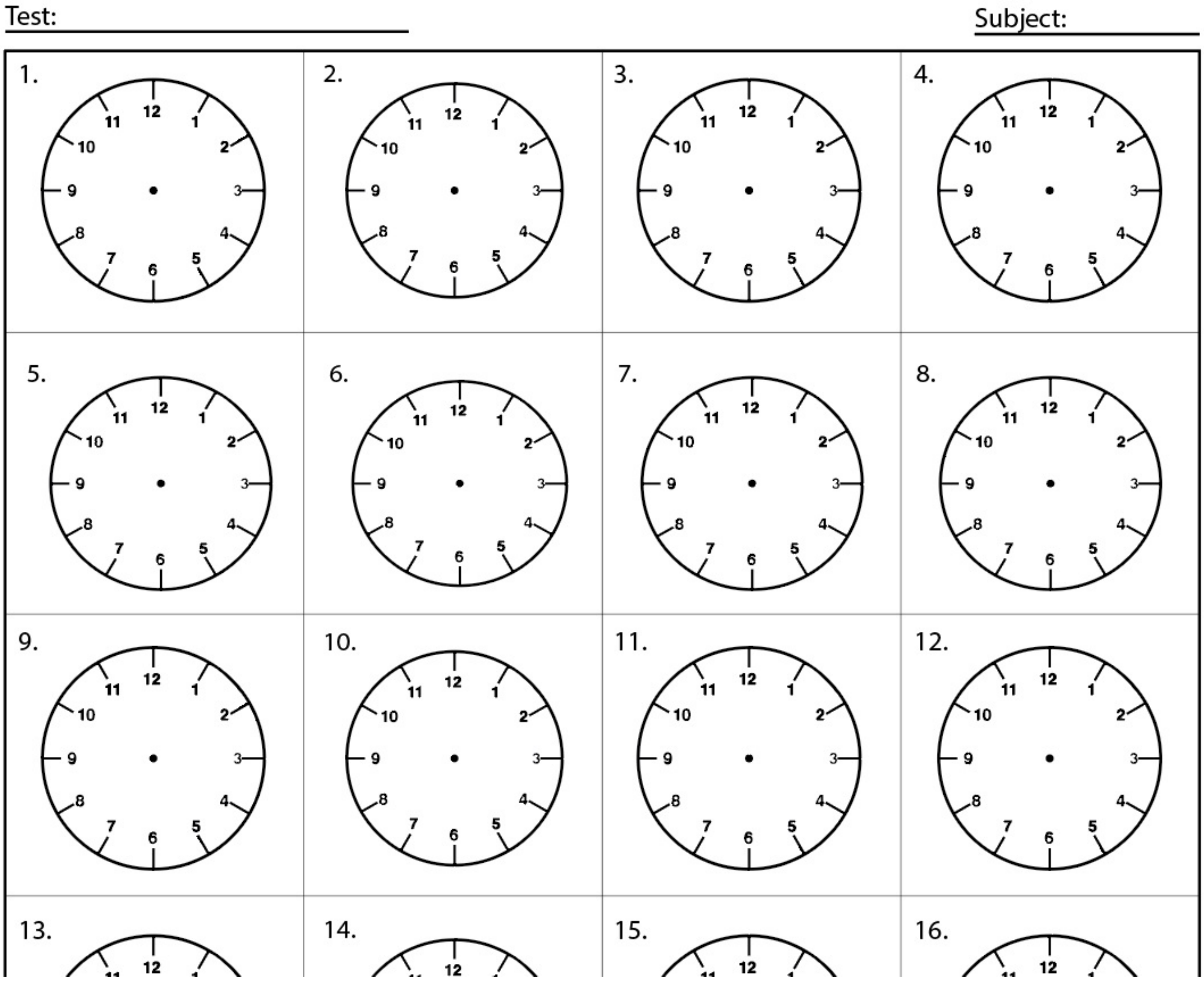
1 2 3 4 5

9. I was more engaged with the work because of the moving and spatialized sound.

1 2 3 4 5

10. Please write any additional impressions or comments regarding the sound: ________

Appendix 2:



Test: ______________________ Subject: ____________

1.	2.	3.	4.
5.	6.	7.	8.
9.	10.	11.	12.
13.	14.	15.	16.

References

Adobe, 2019. Adobe After Effects CC Visual effects and motion graphics software. Available at: https://www.adobe.com/products/aftereffects.html [Accessed January 5, 2019].

adtile, 2018. Contribute to adtile/Full-Tilt development by creating an account on GitHub. Available at: https://github.com/adtile/Full-Tilt [Accessed January 5, 2019].

A Gerzon, M., 1973. *Periphony: With-Height Sound Reproduction*,

Alais, D. & Burr, D., 2004. The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, 14(3), pp.257–262. Available at: https://doi.org/10.1016/j.cub.2004.01.029 [Accessed December 24, 2018].

Apollinaire, G., 1984. *The poet assassinated and other stories*, San Francisco: North Point Press.

Aristotle. & Ross, W.D., 1981. *Aristotle's Metaphysics*, Oxford [England]: Clarendon Press.

Arriaga, J.G. et al., 2014. Using Song to Identify Cassin's Vireo Individuals. A Comparative Study of Pattern Recognition Algorithms. In J. F. Martínez-Trinidad et al., eds. *Pattern Recognition*. Springer International Publishing, pp. 291–300.

Ashley, R., 1967. She Was A Visitor.

Autodesk, 2018. Cloud Powered 3D CAD/CAM Software for Product Design Fusion 360. Available at: https://www.autodesk.com/products/fusion-360/overview [Accessed February 10, 2018].

Battaglia, P., Jacobs, R. & Aslin, R., 2003. *Bayesian integration of visual and auditory signals for spatial localization*,

Beardsley, R. & Leech-Wilkinson, D., 2009. A Brief History of Recording to ca. 1950. Available at: http://www.charm.rhul.ac.uk/history/p20_4_1.html [Accessed January 5, 2019].

Benjamin, W., Arendt, H. & Zohn, H., 1968. *Illuminations*, New York: Harcourt, Brace & World.

Benjamin Piekut & George E. Lewis, 2013. *The Oxford Handbook of Critical Improvisation Studies, Volume 2*, Oxford University Press. Available at: http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199892921.001.0001/oxfordhb-9780199892921.

Berger, J. et al., 1973. *Ways of seeing*,

BlastTheory, 2007. Rider Spoke Blast Theory. Available at: https://www.blasttheory.co.uk/projects/rider-spoke/ [Accessed January 5, 2019].

Boursier-Mougenot, C., 1999. From Here to Ear.

Brâncuși, C., 1928. Bird in Space.

Brown, B. & Juhlin, O., 2015. *Enjoying machines*,

Bugler, C., 2012. *The bird in art*, London; New York: Merrell.

Burdick, A. et al., 2016. *Digital-humanities*,

Bush, V., 1945. As We May Think. *The Atlantic*. Available at: https://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/ [Accessed January 5, 2019].

Cage, J., 1960. Imaginary landscape, no. 4, or, March no. 2, for 12 radios. [Music].

Cage, J., 1961. *Silence : Lectures and writings*,

Cage, J., 1963. Variations IV.

Calder, A., 1971. *Eagle*,

Carlile, S. & Best, V., 2002. Discrimination of sound source velocity in human listeners. *Journal of the Acoustical Society of America*, 111(2), pp.1026–1035.

Carson, R. & Darling, L., 1962. *Silent spring*, Boston; Cambridge, Mass.: Houghton Mifflin ; Riverside Press.

Catuhe, D., Rousset, D. & Vandenberghe, S., 2018. Babylon.js demos & documentation. *Babylon.js*. Available at: http://www.babylonjs.com [Accessed January 5, 2019].

Chacin, A. et al., 2016. Bird Song Diamond: Call and Response and Phase Transition Work. *21st International Symposium on Artificial Life and Robotics*.

Chadabe, J., 1997. Electric sound : The past and promise of electronic music. Available at: http://books.google.com/books?id=J4nuAAAAMAAJ.

Charles Taylor et al., 2011. NSF Award Search: Award#1125423 - CDI-Type II: Acoustic

Sensor Arrays for Understanding Bird Communication. Available at: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1125423 [Accessed January 5, 2019].

Chertow, M.R., 2000. The IPAT Equation and Its Variants. *Journal of Industrial Ecology*, 4(4), pp.13–29. Available at: http://onlinelibrary.wiley.com/doi/10.1162/108819800052541927/abstract [Accessed February 9, 2018].

Colavita, F.B., 1974. Human sensory dominance. *Perception & Psychophysics*, 16(2), pp.409–412. Available at: https://doi.org/10.3758/BF03203962.

Cooley, M., 1996. On Human-Machine Symbiosis. In K. S. Gill, ed. *Human Machine Symbiosis: The Foundations of Human-centred Systems Design*. London: Springer London, pp. 69–100. Available at: https://doi.org/10.1007/978-1-4471-3247-9_2.

Crist, E., 2013. The poverty of our nomenclature. *Environmental Humanities*, 3, pp.129–147.

Cycling74, 2018. Max Software Tools for Media. Available at: https://cycling74.com/products/max/ [Accessed February 10, 2018].

DeFanti, T.A., Sandin, D.J. & Cruz-Neira, C., 1993. A ”room” with a ”view”. *IEEE Spectr.*, 30(10), pp.30–33.

Desantos, S., Roads, C. & Bayle, F., 1997. Acousmatic Morphology: An Interview with François Bayle. *Computer Music Journal*, 21(3), pp.11–19. Available at: http://www.jstor.org/stable/3681010.

Ehnes, J., 2010. An audio visual projection system for virtual room inhabitants. *20th International Conference on Artificial Reality and Telexistence (ICAT2010).*

Espressif, 2018. ESP8266 Overview Espressif Systems. Available at: https://www.espressif.com/en/products/hardware/esp8266ex/overview [Accessed December 8, 2018].

Fechner, G.T. & Wundt, W.M., 1889. *Elemente der Psychophysik*, Leipzig: Breitkopf & H rtel.

Forum, E.C., 2019. ESP8266 core for Arduino. Contribute to esp8266/Arduino development by creating an account on GitHub. Available at: https://github.com/esp8266/Arduino [Accessed January 5, 2019].

Galilei, G. & Drake, S., 1990. *Discoveries and opinions of Galileo : Including The starry messenger (1610), Letter to the Grand Duchess Christina (1615), and excerpts from Letters on sunspots (1613), the assayer (1623)*, New York: Anchor Books.

Gibson, J., 1986. *The Ecological Approach to Visual Perception*, Lawrence Erlbaum Associates. Available at: https://books.google.co.jp/books?id=DrhCCWmJpWUC.

Godman, R., 2007. The enigma of Vitruvian resonating vases and the relevance of the concept for today. *The Journal of the Acoustical Society of America*, 122(5), pp.3054–3054. Available at: https://asa.scitation.org/doi/abs/10.1121/1.2942890 [Accessed December 7, 2018].

Harraway, D., 2016. Tentacular Thinking: Anthropocene, Capitalocene, Chthulucene. *e-flux*, (#75). Available at: http://www.e-flux.com/journal/75/67125/tentacular-thinking-anthropocene-capitalocene-chthulucene/ [Accessed February 9, 2018].

Head, M., 1997. Birdsong and the Origins of Music. *Journal of the Royal Musical Association*, 122(1), pp.1–23. Available at: https://doi.org/10.1093/jrma/122.1.1.

Hegarty, P., 2007. *Noise/music : A history*, New York: Continuum.

Hein, H.S., 1990. *The Exploratorium : The museum as laboratory*, Washington: Smithsonian Institution Press.

Hermann, T. et al. eds., 2011. *The sonification handbook*, Berlin: Logos Verlag.

Hofman, M. & Van Opstal, J., 2003. Binaural weighting of pinna cues in human sound localization. *Experimental Brain Research*, 148(4), pp.458–470.

Huhtamo, E., 2013. Illusions in motion media archaeology of the moving panorama and related spectacles. Available at: http://site.ebrary.com/id/10661915.

IDRE, U., 2013. Choosing the Correct Statistical Test in SAS, Stata, SPSS and R. Available at: https://stats.idre.ucla.edu/other/mult-pkg/whatstat/ [Accessed January 5, 2019].

Ikegami, T. et al., 2017. Life as an emergent phenomenon: Studies from a large-scale boid simulation and web data. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 375(2109).

Ishii, K. et al., 2007. A Navigation System Using Ultrasonic Directional Speaker with Rotating Base. In M. J. Smith & G. Salvendy, eds. *Human Interface and the Management of Information. Interacting in Information Environments*. Springer Berlin Heidelberg, pp. 526–535.

Iwata, H., 2015. . Available at: http://www.emp.tsukuba.ac.jp/english/environment/research.php [Accessed February 10, 2018].

J. Berkhout, A., Vries, D. & VOGEL, P., 1993. *Acoustic control by wave field synthesis*,

Kac, E. et al., 1996. Rara Avis. Available at: http://www.ekac.org/raraavis.html [Accessed December 1, 2018].

Kahn, D., 2013. Earth Sound, Earth Signal : Energies and Earth Magnitude in the Arts. Available at: http://dx.doi.org/10.1525/california/9780520257801.001.0001.

Kahn, D., 1999. *Noise, water, meat : A history of sound in the arts*, Cambridge, Mass.: MIT Press.

Kasahara, S. et al., 2012. Second surface: Multi-user spatial collaboration system based on augmented reality. In ACM Press, pp. 1–4. Available at: http://dl.acm.org/citation.cfm?doid=2407707.2407727 [Accessed February 18, 2018].

Koppen, C. & Spence, C., 2008. *Audiovisual asynchrony modulates the Colavita visual dominance effect*,

Koppen, C. & Spence, C., 2007. Spatial coincidence modulates the Colavita visual dominance effect. *Neuroscience Letters*, 417(2), pp.107–111. Available at: http://www.sciencedirect.com/science/article/pii/S0304394006011566.

Kraft, D., 2013. *Birdsong in the music of Olivier Messiaen*, London: Arosa Press.

Krause, B., 1987. Bioacoustics, habitat ambience in ecological balance. *Whole Earth Review*, 57, pp.14–18.

Kuka, D. et al., 2009. DEEP SPACE: High Resolution VR Platform for Multi-user Interactive Narratives. In *Proceedings of the 2Nd Joint International Conference on Interactive Digital Storytelling: Interactive Storytelling.* ICIDS '09. Berlin, Heidelberg: Springer-Verlag, pp. 185–196. Available at: http://dx.doi.org/10.1007/978-3-642-10643-9_24.

Kuutti, J., Leiwo, J. & Sepponen, R., 2014. Local Control of Audio Environment: A Review of Methods and Applications. Available at: https://aaltodoc.aalto.fi:443/handle/123456789/25352 [Accessed January 5, 2019].

LabLua, 2018. The Programming Language Lua. Available at: https://www.lua.org/ [Accessed January 5, 2019].

Lanier, J. & Heilbrun, A., 1988. VR Interview, Whole Earth Review. Available at: http://www.jaronlanier.com/vrint.html [Accessed January 5, 2019].

Lazar, J., Feng, J.H. & Hochheiser, H., 2017. *Research methods in human-computer interaction*,

Legrady, G., Pinter, M. & Bazo, D., 2013. Swarm Vision. Available at: https://www.mat.ucsb.edu/g.legrady/glWeb/Projects/sv/swarmvision.html [Accessed December 1, 2018].

Lombardo, V. et al., 2009. A Virtual-Reality Reconstruction of Poème électronique Based on Philological Research. *Computer Music Journal*, 33(2), pp.24–47. Available at: https://doi.org/10.1162/comj.2009.33.2.24 [Accessed December 5, 2018].

Lynxmotion, 2018. Lynxmotion - SSC-32U USB Servo Controller. Available at: http://www.lynxmotion.com/p-1032-ssc-32u-usb-servo-controller.aspx [Accessed January 5, 2019].

Malm, A., 2018. *The Progress of This Storm: On Society and Nature in a Warming World*, Verso. Available at: https://books.google.co.jp/books?id=7xZZMQAACAAJ.

Martin, G., 2011. *Through birds' eyes: Insights into avian sensory ecology*,

Maruyama, N. et al., 2014. Evolution of Artificial Soundscape in a Natural Environment. *Exploiting synergies between biology and artificial life technologies: tools, possibilities, and examples at ALIFE*, 14.

Maruyama, N., Oka, M. & Ikegami, T., 2013. Creating space-time affordances via an autonomous sensor network - Semantic Scholar. *2013 IEEE Symposium on Artificial Life (ALife)*, pp.67–73.

Massumi, B., Fish, S. & Jameson, F., 2002. *Parables for the Virtual: Movement, Affect, Sensation*, Duke University Press. Available at: https://books.google.co.jp/books?id=93S7aCK0AP8C.

Meese, T.S., 1995. Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Perception & Psychophysics*, 57(3), pp.267–281. Available at: https://doi.org/10.3758/BF03213053.

Microsoft, 2014. High definition face tracking. Available at: https://docs.microsoft.com/en-us/previous-versions/windows/kinect/dn785525(v%3dieb.10) [Accessed January 5, 2019].

Microsoft, 2016. Kinect - Windows app development. Available at: https://developer.microsoft.com/en-us/windows/kinect [Accessed January 5, 2019].

Miles, S., 2008. Objectivity and Intersubjectivity in Pauline Oliveros's "Sonic Meditations". *Perspectives of New Music*, 46(1), pp.4–38. Available at: http://www.jstor.org/stable/25652374.

Milk, C. et al., 2012. The Treachery of Sanctuary.

Mills, A.W., 1958. On the Minimum Audible Angle. *The Journal of the Acoustical Society of America*, 30(4), pp.237–246. Available at: https://asa.scitation.org/doi/10.1121/1.1909553 [Accessed January 5, 2019].

Nagel, T., 1974. What Is It Like to Be a Bat? *The Philosophical Review*, 83(4), pp.435–450. Available at: http://www.jstor.org/stable/2183914.

Nees, M.A. & Walker, B.N., 2011. Auditory Displays for In-Vehicle Technologies. *Reviews of Human Factors and Ergonomics*, 7(1), pp.58–99. Available at: https://doi.org/10.1177/

1557234X11410396 [Accessed January 4, 2019].

Nelson, T.M. & Nilsson, T.H., 1990. Comparing headphone and speaker effects on simulated driving. *Accident; Analysis and Prevention*, 22(6), pp.523–529.

Node.js, 2018. Node.js. *Node.js.* Available at: https://nodejs.org/en/ [Accessed January 5, 2019].

NodeMCU, 2018. Overview - NodeMCU Documentation. Available at: https://nodemcu.readthedocs.io/en/master/ [Accessed December 8, 2018].

Nyman, M., 1999. *Cage and beyond.*, [Place of publication not identified]: Cambridge University Press.

Ochiai, Y., Hoshi, T. & Suzuki, I., 2017. Holographic Whisper: Rendering Audible Sound Spots in Three-dimensional Space by Focusing Ultrasonic Waves. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* CHI '17. New York, NY, USA: ACM, pp. 4314–4325. Available at: http://doi.acm.org/10.1145/3025453.3025989.

Oliveros, P., 2005. *Deep listening : A composer's sound practice*, New York, NY: iUniverse.

OptiTrack, 2018. OptiTrack Unity Plugin - NaturalPoint Product Documentation Ver 2.0. Available at: https://v20.wiki.optitrack.com/index.php?title=OptiTrack_Unity_Plugin [Accessed January 5, 2019].

Ouzounian, G. & DeMarinis, P., 2010. An Interview with Paul DeMarinis. *Computer Music Journal*, 34(4), pp.10–21. Available at: http://www.jstor.org/stable/40962937.

Pandas, 2019. Pandas.DataFrame.Rolling — pandas 0.23.4 documentation. Available at: https://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.rolling.html [Accessed January 5, 2019].

Phua, K.S. & Gan, W.-S., 1999. *Spatial speech coding for multi-teleconferencing*,

Pick, H.L., Warren, D.H. & Hay, J.C., 1969. Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, 6(4), pp.203–205. Available at: https://doi.org/10.3758/BF03207017.

Pijanowski, B.C. et al., 2011. Soundscape Ecology: The Science of Sound in the Landscape. *BioScience*, 61(3), pp.203–216. Available at: http://www.jstor.org/stable/10.1525/bio.2011.61.3.6.

Plant, S., 2006. *The most radical gesture the situationist international in a postmodern age.*, London [etc.: Routledge.

Plato. & Allen, R.E., 2008. *The republic*, New Haven, Conn.; London: Yale University

Press.

Pompei, F.J.J., 2019. *Sound from ultrasound : The parametric array as an audible sound source /*,

Reas, C. & Fry, B., 2014. Processing a Programming Handbook for Visual Designers and Artists, Second Edition. Available at: http://www.books24x7.com/marc.asp?bookid=73658.

Renaudot, E. et al., 1664. *A general collection of discourses of the virtuosi of France : Upon questions of all sorts of philosophy, and other natural knowledg*, London: Printed vor Thomas Dring; John Starkey,; are to be sold at their shops, at the George in Fleet-street neer Clifford's-Inn,; the Miter between the Middle-Temple-Gate; Temple-Bar.

Rozier, J. & Karahalios, K., 1999. Hear&There. Available at: http://smg.media.mit.edu/projects/HearAndThere/ [Accessed January 5, 2019].

Russolo, L. et al., 1967. *The art of noise : Futurist manifesto, 1913*, New York: Something Else Press.

Sato, H., 2018. Human Informatics Research Institute National Institute of Advanced Industrial Science and Technology. Available at: https://unit.aist.go.jp/hiri/en/ [Accessed January 5, 2019].

Schaeffer, P., North, C. & Dack, J., 2012. *In search of a concrete music*,

Schmeder, A., Freed, A. & Wessel, D., 2010. Best Practices for Open Sound Control. In *Linux Audio Conference.* Utrecht, NL.

Shi, C. & Gan, W.S., 2010. Development of Parametric Loudspeaker. *IEEE Potentials*, 29(6), pp.20–24.

Simon, T., 2014. Birds of the West Indies.

Sosolimited, Hypersonic & Design, P., 2016. Diffusion Choir.

Stengers, I. & Goffey, A., 2015. *In Catastrophic Times: Resisting the Coming Barbarism*, Open Humanities Press. Available at: https://books.google.co.jp/books?id=0oXRjgEACAAJ.

Sutherland, I.E., 1965. The Ultimate Display. In *Proceedings of the IFIP Congress.* pp. 506–508.

Suzuki, R. et al., 2017. HARKBird: Exploring Acoustic Interactions in Bird Communities Using a Microphone Array. *Journal of Robotics and Mechatronics*, 29(1), pp.213–223.

Suzuki, R. et al., 2018. *Field observations of ecoacoustic dynamics of a Japanese bush warbler using an open-source software for robot audition HARK*, Takatori, H. et al., 2016. *Development of A Large-Immersive Display “LargeSpace”*,

Taylor, C.E. et al., 2017. Sensitivity of California Thrashers (Toxostoma redivivum) to song syntax. *Bioacoustics*, 26(3), pp.259–270. Available at: https://doi.org/10.1080/09524622.2016.1274917.

Technologies, U., 2018. Unity. *Unity.* Available at: https://unity3d.com [Accessed January 5, 2019].

Thrift, N., 2008. *Non-Representational Theory: Space, Politics, Affect*, Taylor & Francis. Available at: https://books.google.co.jp/books?id=0LM6UedgqA0C.

Thrift, N., 1997. Re-imagining places, re-imagining identities. In *Consumption and Everyday Life.* Culture, Media and Identities. SAGE Publications Ltd, pp. 159–212.

TriState, 2015. TriState ! . Available at: http://www.tristate.ne.jp/parame.htm [Accessed January 5, 2019].

Tuchman, M., 1971. *Art & technology: A report on the Art & Technology Program of the Los Angeles County Museum of Art, 1967-1971*, Los Angeles County Museum of Art; distributed by the Viking Press, New York.

Valle, A., Tazelaar, K. & Lombardo, V., 2010. *In a concrete space. Reconstructing the spatialization of Iannis Xenakis’ Concret PH on a multichannel setup*,

Verdaasdonk, M.A., 2007. *Living lens: Exploring interdependencies between performing bodies, visual and sonic media in immersive installation.* PhD thesis. Queensland University of Technology.

Vitruvius Pollio. & Morgan, M.H., 1960. *Vitruvius : The ten books on architecture*, New York: Dover Publications.

Wagner, B., 2018. C# Programming Guide. Available at: https://docs.microsoft.com/en-us/dotnet/csharp/programming-guide/ [Accessed January 5, 2019].

Wagner, R. et al., 1964. *Wagner on music and drama : A compendium of Richard Wagner’s prose works*,

Wainwright, J. & Mann, G., 2018. *Climate Leviathan*, Verso Books. Available at: https://books.google.co.jp/books?id=qNEPDgAAQBAJ.

Wark, M., 2016. *Molecular Red: Theory for the Anthropocene*, Verso Books. Available at: https://books.google.co.jp/books?id=J-5qjwEACAAJ.

Watanabe, J., Tavata, T. & Verdaasdonk, M.A., 2006. Toward individualized audience experience in performance installation. *DIME.*

Watanabe, J. et al., 2007. *Practical Approach of Vibro-scape Design in Multi-media*

Performing Art,

Wernicke, J., POLII. Available at: https://www.johanneswernicke.com/pol2 [Accessed January 5, 2019a].

Wernicke, J., POLYUS. *polyus*. Available at: https://www.johanneswernicke.com/polyus [Accessed January 5, 2019b].

Wickham, H., 2014. Tidy data. *The Journal of Statistical Software*, 59(10). Available at: http://www.jstatsoft.org/v59/i10/.

Yoneyama, M. et al., 1983. The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design. *The Journal of the Acoustical Society of America*, 73(5), pp.1532–1536. Available at: https://doi.org/10.1121/1.389414 [Accessed January 5, 2019].

Yu, K. et al., 2016. Wireless Sensor Array Network DoA Estimation from Compressed Array Data via Joint Sparse Representation X.-B. Jin, ed. *Sensors (Basel, Switzerland)*, 16(5), p.686. Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4883377/.