STUDIES IN SPATIAL AURAL PERCEPTION: ESTABLISHING FOUNDATIONS FOR IMMERSIVE SONIFICATION

Ivica Ico Bukvic
Virginia Tech
SOPA, ICAT, C+I
Blacksburg, VA, USA
ico@vt.edu

Gregory Earle
Virginia Tech
ECE
Blacksburg, VA, USA
earle@vt.edu

Disha Sardana
Virginia Tech
HCD
Blacksburg, VA, USA
dishas9@vt.edu

Woohun Joo
Virginia Tech
HCD
Blacksburg, VA, USA
joowh@vt.edu

ABSTRACT

The Spatial Audio Data Immersive Experience (SADIE) project aims to identify new foundational relationships pertaining to human spatial aural perception, and to validate existing relationships. Our infrastructure consists of an intuitive interaction interface, an immersive exocentric sonification environment, and a layer-based amplitude-panning algorithm. Here we highlight the system's unique capabilities and provide findings from an initial externally funded study that focuses on the assessment of human aural spatial perception capacity. When compared to the existing body of literature focusing on egocentric spatial perception, our data show that an immersive exocentric environment enhances spatial perception, and that the physical implementation using high density loudspeaker arrays enables significantly improved spatial perception accuracy relative to the egocentric and virtual binaural approaches. The preliminary observations suggest that human spatial aural perception capacity in real-world-like immersive exocentric environments that allow for head and body movement is significantly greater than in egocentric scenarios where head and body movement is restricted. Therefore, in the design of immersive auditory displays, the use of immersive exocentric environments is advised. Further, our data identify a significant gap between physical and virtual human spatial aural perception accuracy, which suggests that further development of virtual aural immersion may be necessary before such an approach may be seen as a viable alternative.

1. INTRODUCTION

Human interfaces to the natural world are inherently multisensory [1]. In simulated environments we often mimic our interaction with the natural world by combining sensory mechanisms to broaden our cognitive bandwidth [2], and to reinforce comprehension [3] and learning [4], [5]. A 1997 report to the National Science Foundation [6] defines sonification as "the use of nonspeech audio to convey information". Simplistic examples of sonification include warning "beeps" that sound when a piece of heavy machinery backs up, and the click-frequency associated with Geiger counters [7], but the full potential of sonification and the multidimensionality of sound is only starting to be explored, particularly in the context of multidimensional datasets. This conceit echoes studies spanning the past two decades: in 1999 Hermann and Ritter suggested that sonification is an "underused perceptual channel for man-machine interaction" [8], and in 2007 Nasir and Roberts stated that "researchers have not fully utilized the maximum potential of spatial sound" [9]. More recently, a 2014 paper by Thomas Hermann suggests that sonification is still in its infancy [10], while in a 2018 publication Paul Vickers notes that "our knowledge of sonification design and theory is still fairly primitive" [11].

Unlike data visualization, which has a long history and a clear set of foundational guidelines [12], sonification is a nascent field [11] that has not yet produced a counterpart to Tufte's seminal work on data visualization [12]. The lack of such knowledge may be one of the major obstacles to the broader adoption of sonification. Sound is inherently multidimensional–each sound has multiple properties that can be assigned to independent variables, or combined to reinforce the perception of a single variable. Such dimensions include timbre, pitch, amplitude, psychoacoustic meaning, source location, and movement. This content richness, when coupled with the innate human ability to simultaneously detect and discriminate between multiple sound sources, supports the contention that sonification affords tremendous promise for analysis of large, complex, multidimensional datasets. Research into sonification may lead to new ways to understand and interact with data, and may significantly enhance and extend traditional data analysis techniques.

1.1. Immersive Exocentric Sonification

In the field of user interfaces the term exocentric environment refers to a virtual reality or other immersive environment that completely encompasses the user [13], [14]. In a previous publication we extended this definition into the aural domain to make a case for an environment that offers all the affordances of the way we interact with the real world [15]. Thus a live concert is experienced in an exocentric environment, but this element is lost when music is heard through headphones that attempt to mimic an exocentric environment but fail to account for user's change in location and orientation. A key difference is that head-motions, echoes, and phase and amplitude differences based on proximity, orientation, and environmental characteristics are fully experienced in an exocentric environment. This specific meaning of the term "exocentric environment" is used throughout this paper. Central to the exocentric environment is its focus on producing sounds whose qualities remain stable throughout the space. Our exocentric environment renders sound sources only around the space perimeter, with no at-

This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at http://creativecommons.org/licenses/by-nc/4.0

https://doi.org/10.21785/icad2019.017

tempt to emulate virtual sources inside or outside the physical volume. This environment is in contrast to an egocentric environment that prevents changes in sound due to head and body movement, rotation, and orientation. These disparate environments may lead to different conclusions pertaining to human spatial aural perception acuity. This is in part because the egocentric approach does not reflect our real-world capacity for processing sound and interacting with natural aural environments. It is worth noting that there are real-world scenarios in which the egocentric approach may be necessary, because significant head and body movement, rotation, and/or orientation is either not possible or is discouraged. As such, the exploration of both approaches continues to be relevant to the field of sonification.

2. THE SADIE PROJECT

The Spatial Audio Data Immersive Experience (SADIE) is a project whose goal is to increase our capability to create, manage, and understand data and information, with an emphasis on immersive exocentric sonification of three-dimensional multivariate coupled systems. SADIE aims to study spatially distributed data by creating a natural aural environment that leverages intuitive affordances of the immersive exocentric sonification environment, including:

- Utilization of a physical space—recognition and utilization of acoustics, reverberance, and reflections;
- Location-based perception—sound amplitude that is dependent on the user's location within the acoustic field, creating an environment that builds upon natural human perception capabilities;
- Individual variance—aural perception is unique to each individual, and this limits the effectiveness of Head Related Transfer Functions (HRTF) that tacitly employ a one-size-fits-all approach.

2.1. Project Goals

The SADIE project has two primary goals. The first is to develop a powerful, flexible, and reproducible set of tools and techniques with minimal idiosyncrasies, through which it is possible to explore immersive exocentric sonification. The findings of this work will guide further research, including immersive sonification studies of inherently multidimensional spatial data, which may help to quantify its utility for research, teaching, and real-world applications. To minimize idiosyncrasies the project focuses on a geospatial environment model, which has inherently spatial qualities that are directly mapped onto the spatial domain around space perimeter. The second goal of the SADIE project is to develop intuitive approaches to interactions with aural data, including both scientific and artistic scenarios. The data obtained from test subjects have identified several fundamental findings relevant to both goals.

2.2. Side-stepping Idiosyncrasies

Cross-domain-mapping maps elements from a source domain onto a target domain to add an additional level of meaning to the target domain [16]. The approach described herein leverages the human capacity for cross-domain mapping while minimizing potential idiosyncrasies. For example, when we listen to sounds we use our vantage point, location, and motion to accurately pinpoint the sound source, thereby reinforcing our perception by using cross-domain-mapping. In contrast, in the existing immersive audio research literature we observe extensive work in studying human aural perception egocentrically, or in isolation from other senses [15]. One reason for this is the lack of access to infrastructure that is capable of rendering an immersive exocentric aural environment while also tracking users as they traverse the space. Some research has attempted to develop a simulated algorithm to address this problem [5], [17], [18]. Such implementations tend to introduce compounding idiosyncrasies [15] whose impact on the study data may be underestimated. For instance, consider the front-back confusion idiosyncrasy that is inherent to binaural virtual systems; it cannot be addressed without introducing a head-tracking system, but doing so creates latency issues that further compound the problem by introducing new idiosyncrasies.

Another area of concern that may hamper the ability to work with empirical data is the artificial way in which users interact with the system. Interaction that is complex and unnatural may yield biased findings. We posit that systems designed to identify foundational relationships need to be as natural and intuitive as possible. In the SADIE project we focus on allowing subjects to interact naturally with their surroundings, including the ability to freely navigate the space, and to manipulate the properties of spatial aural sources using simple and intuitive hand gestures. We achieve this by using a glove-based gesture interface that is tracked by a motion capture system.

2.3. Infrastructure

A key aspect of SADIE infrastructure is a unique Virginia Tech facility known as the Cube, an immersive cuboid audio facility that measures 50x40x32 feet [19]. The features of the facility that are most relevant to this study are the motion capture capability and the loudspeaker array. The latter includes 124 homogeneous speakers distributed across 5 layers within the facility (3 catwalks and 2 ceiling layers), 4 subwoofers in quad configuration covering frequencies down to 50 Hz, and two 17-inch subwoofers responsible for frequencies below 50 Hz. This configuration enables rendering of a cuboid hemisphere with listeners able to freely traverse the equatorial cross-section of sonified, inherently spatial data. The facility is conducive to all current spatialization algorithms, including both physical and virtual, and thereby allows for testing foundational assumptions and identifying the underexplored potential of immersive sonification.

3. IMPLEMENTATION

Central to SADIEs implementation are three components that constitute the Locus system [20]: the glove-based interaction interface coupled with a motion tracking system, Unity [21] middleware designed to translate captured data into easily interpreted and manipulated Open Sound Control (OSC)-like [22] network packets, and a MaxMSP [23] patch that renders spatial sound and responds to user interaction based on the Unity data stream. We discuss each component in greater detail below.

3.1. Interaction Interface

3.1.1. Prior Work

Within the ICAD community, Beilharz [24] proposed a gestural interaction interface designed to affect sonified data and en-

hance interactions with sound. There were different approaches to a gesture-control interface for use in sound mixing [25] and sound position adjustment of multi-track audio [26]. Sheridan et al. [27]. introduced hand gesture-based software called Soundstudio4D, which allows users to synthesize, spatialize and edit sound. Sterkenburg et al. [28] conducted research on how hand movement can be productively used in connection with auditory displays.

3.1.2. LOCUS

As part of Locus we developed a wearable device to facilitate natural user interaction with spatial aural content in the immersive exocentric environment. It uses two off-the-shelf gloves fitted with retroreflective markers, in conjunction with the 24-camera Qualisys Oqus 500+ motion capture system to allow users to point towards a perceived direction of a sound. To facilitate accurate tracking of both hands, including varying finger positions we used the AIM (Automatic Identification of Markers) model offered by the Qualisys QTM software. Once properly trained, the AIM model is capable of identifying the trained object regardless of hand size, finger or hand position, or orientation. Simple hand gestures are extracted from the Qualisys' spatial marker data using the Unity gaming engine-based toolkit. In this study we focus on the finger pointing gesture that offers a proven natural interaction [29] with minimal impact on the user performance [30]. It is coupled by a thumb trigger gesture consisting of thumb touching the side of the index finger that users can employ to mark the perceived location of the source. We use Unity's vector and raytracing processing capabilities to accurately detect the user's pointing location on the periphery of the domain with submillimeter accuracy, while simultaneously monitoring and responding to other gestures, e.g. thumb trigger. A visualization framework designed to accelerate system setup and troubleshooting illustrates these features on a computer screen.

Unity toolkit's OSC-like output formatting allows it to interface with a wide variety of network-enabled digital signal processing software. Once the motion-capture data are processed they are sent to MaxMSP that responds to the captured data and user's gestures. Doing so allows rapid prototyping by leveraging the functionality of the D4 audio spatialization library that was designed specifically for use with HDLAs in low-latency interactive scenarios with focus on sonification of multidimensional scalar arrays [31]. The resulting infrastructure allows us to distribute sound across the 124.6 HDLA with a high degree of control and interactivity. As described above, interactions employed in this study include pointing towards the perceived location of a spatialized sound source and marking such a location using a thumb trigger motion.

The system enables the support of both egocentric and exocentric environments for a wide array of creative scenarios. Its implementation in the facility further allows for a comparison of various spatialization techniques. Conversely, it allows validation of known sonification ground truths, as well as identification of entirely new ones. These unique affordances have inspired the following research questions.

4. RESEARCH QUESTIONS

The infrastructure described above allows us to study a number of key questions, with the highest priority being given to:

1. What techniques are best-suited to sonifying scalar arrays in an immersive space, and how can they be utilized to facilitate pattern perception in multivariate scenarios?
2. What is a normal user's spatial aural perception capacity, and for what idiosyncrasies must we compensate in interpreting our data?
3. How does the human ability to pinpoint sources and perceive patterns in an immersive exocentric sonification compare to that of more commonly studied egocentric and egocentric-like scenarios?
4. How does sonification in our immersive space compare to that of virtual systems, such as headphone-based binaural systems?

Observing our perception capacity limitations while addressing these topics may better inform the design process and the subsequent implementation of auditory displays. Consequently, the ensuing ground truths may help to create a foundation for a Tufte-like treatise in the audio domain.

5. EXPERIMENTAL APPROACH

Our initial case study focuses on system validation, and on testing the boundaries of human perceptive capabilities in the immersive environment described above. We seek to design tests from which we can infer the limitations of human interactions with the system. A simple example relates to the ability of users to locate sounds that move, are emitted from different locations, or are dispersed over a range of positions. We perform these studies via a sequence of game-like scenarios in which users are asked to identify the sound source location, while we create an anonymous database of user responses. Doing so allows us to determine the normal limits of human perception, and to assess the statistical significance of various tests. As evidenced by the existing body of research, casting our initial studies in game-form was expected to decrease stress among users, while providing a playful environment that may lead to improved retention for sequential studies [32].

5.1. Sonifying Data

Sonification studies can use synthesized and/or sampled sounds. Synthesis offers flexibility in how various parameters may be mapped to the sound generating properties, including simple data audification at a human-audible rate. This can result in widely varying and unnatural sounds. On the other hand, sampled sounds offer a sense of familiarity, and in some instances their psychoacoustic meanings can aid data interpretation. Furthermore, natural or familiar sounds may minimize fatigue and/or annoyance. A notable subset of the aforesaid two approaches are earcons and auditory icons [33] that have a proven role and value in notifying users. Of particular interest is faster response time associated with the auditory icons that, under the right conditions, can be seen as a form of sonification using sampled sounds. Consequently, in our study we opted for a sound that has the following qualities:

- Familiarity;
- Minimal fatigue and/or annoyance factor;
- Broad spectrum that enables greater spatial localization potential, allows for various processing/filtering techniques, and minimizes chances of the sound being masked by other sounds, and
- Consistent amplitude to enable detectable amplitude modulation.

A pre-recorded sound loop of cicadas meets all of these goals. This sound has been used throughout all studies conducted so far in conjunction with a low-frequency (4Hz) exponential inverted sawtooth waveform that modulates the sound's amplitude. Its use resulted in an impulse-like presence of a sound resembling pink noise with a short decaying envelope, followed by a near-silent moment that highlights dissipating reflections.

The study was further complemented by earcons that provided user feedback, which helps to promote a game-like experience for the user. Given the combination of sampled material, its manipulation through amplitude and pulse frequency modulation, and the use of earcons, our system is a hybrid of the aforesaid approaches.

	Stationary	Moving
Physical	Users stand in the middle of the room and are allowed to rotate head and body to locate the source, but are not allowed to move.	Users are encouraged to rotate their head and body, and to walk within the space to help locate the source. A boundary is maintained via a warning sound if the user leaves the motion tacked area.
Virtual	Users locate the sound sources using binaural rendition via motion tracked headphones. They are allowed to rotate their body and head but are not allowed to move.	Users wear headphones while moving around the room to locate the source. The motion tracking system monitors their position and orientation, modifying the sound accordingly.

Table 1: A 2x2 matrix of test scenarios and their variants.

5.2. Study Scenarios

Variables controlled during testing included room lighting, system calibration settings, speaker positions, and randomness of sound source locations and presentation order. The study consists of two scenarios. The first focuses on physical perception of point sound sources in an immersive exocentric environment. A virtual counterpart to using a headphone-based binaural implementation is the second focus, and both studies make use of the system's motion capture capability to account for changes in the user's orientation, head rotation, and position. In the binaural study we mount a rigid body onto the headphones and compensate for the difference between the location of the rigid body and the user's ears (15 cm downward offset against the local Y axis). The motion capture system records the user's head position and rotation and adjusts the output accordingly.

Two scenarios further explore two variants of exocentric environments, resulting in a 2x2 matrix shown in table 1. The first leverages the full potential of an immersive exocentric sonification environment in which users can move and orient freely. They can further invoke head rotation and motion though the space to improve their ability to locate the sound, thus mimicking the way we interact with real-world sound cues. In the second variant the user's location is fixed, but head and body rotation/orientation are allowed. The latter case is a hybrid that has elements of both exocentric and egocentric environments. Because the two scenarios were a part of a larger study, their order was kept consistent for the sake of minimizing the time overhead in reconfiguring the system, while the two scenario variants were presented in random order.

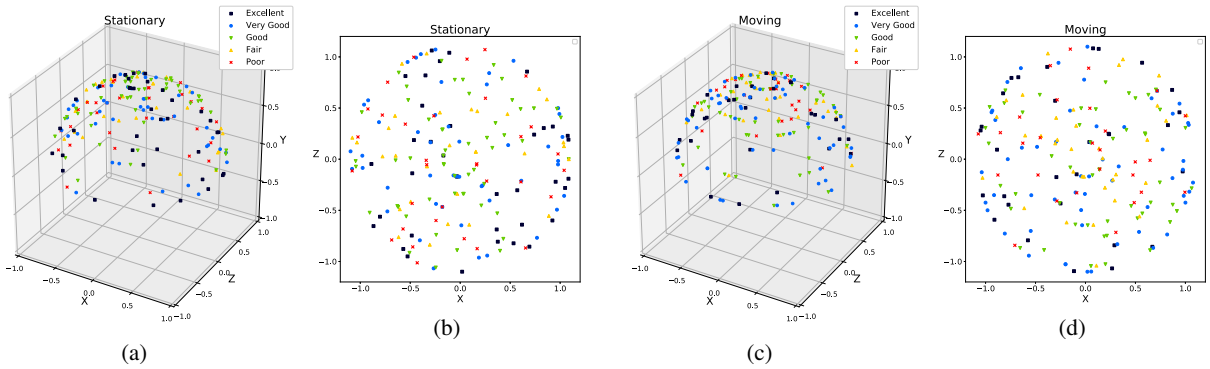
Each scenario consists of 10 trials per user. For each test question a sound is played through the speakers from a random location on the space perimeter, including the ceiling. In human hearing, spatial accuracy decreases with elevation of the sound source. To prevent potential data bias that may ensue from a batch of tests that may use a larger number of randomly generated higher elevations, the elevation choices were limited to 0-90 degrees in 10-degree increments that were consistently utilized in all scenarios, with each elevation being utilized only once per scenario. By phasing the sound sources we create both real (single speaker) and virtual sources, where in the latter case the sound appears to emanate from a region between speakers. Users are asked to find the location of the sound source under various conditions, using their dominant hand for both pointing and marking/triggering functions. With only a short practice session users became adept at interacting with the system.

5.3. Data Processing

Data processing involves calculating the miss-distance in spherical coordinates between the actual sound source and the locations to which the subjects point. The azimuthal and elevation angles of the pointing location are recorded during the study, along with the actual source angles, the time required for users to localize a perceived sound source (in ms), the relative accuracy on a scale of 1-5, and the final game score. Accuracy data are binned in 5 degree increments, so a user pointing to a location within 5 degrees of the correct azimuth and elevation receives the highest possible score for that test. Each lower level of accuracy corresponds to increasing the previous error radius by 5 degrees (10, 15, 20, and 25 or more). Each accuracy level is accompanied by a corresponding earcon. Total game scores allow us to track the best performers among our test subjects in both the stationary and moving scenarios. This competitive aspect of the game adds a degree of excitement, and encourages the test subjects to attempt to beat the all-time highest score, and/or be ranked among the top 10.

5.4. Participant Demographics

After an initial round of beta-testing to identify and remove inconsistencies and biases, a total of 20 test subjects have participated in the study. A hearing test administered prior to the games allows us to screen out persons with hearing impairments. Participants to date are all adults, largely comprised of Virginia Tech students, faculty and staff, and several individuals unaffiliated with Virginia Tech. Test subjects were 30% female and 70% male, ranging in age from 18-55, with a mean age of 25.65. All participation was voluntarily, and no financial or other rewards were given to encourage participation. 90% of the test subjects are right-handed, 75% have had previous experience with a gesture-based device and 45% had been previously exposed to some form of spatial sound environment. Only one of the subjects reported that they were not very interested in music. Others classified themselves on a scale ranging from those who sometimes listened to music, to those who were music majors. All subjects were asked to confirm that they knew how to abort the test before the study began. Qualitative data are drawn from pre- and post-session questionnaires filled out by each user and archived to allow subsequent correlative studies.



(a) (b) (c) (d)

Figure 1: 3D (a and c) and top-down (b and d) color- and shape-coded projections of users' input in stationary and moving variants of the physical immersive scenario, showing consistent accuracy across the entire hemisphere.

6. HYPOTHESES

Several testable hypotheses are investigated using our infrastructure, although more users are needed to generate statistical significance. Prior to testing we formed several hypotheses that were tested using our group of 20 subjects:

1. The human aural perception in an immersive exocentric environment will produce more accurate sound source localization than the egocentric environment;
2. The physical immersive exocentric environment will produce more accurate localization of sound sources than the virtual (binaural, headphone-based) environment;
3. In both physical and virtual scenarios users will perform better in the variant that allows for movement.

7. DATA PRESENTATION AND DISCUSSION

7.1. Quantitative Data

Figure 1 shows the accuracy of our test subjects' ability to identify the source location for sounds emanating from random azimuths and elevations in both stationary and moving exocentric scenarios. Recall from section 5.3 that the color scale of the points corresponds to errors separated into 5 degree bins, ranging from 5 degrees of error to more than 25 degrees. Although the distance between stimuli and users in our study is significantly greater, a subset of this data can be compared to a subset from Figures 5 and 6 of the paper by Oldfield and Parker [29] regarding such errors in a well-controlled egocentric scenario that match the interaction technique while employing a similar sound source (white noise). Whereas both studies cover the full azimuthal plane with the cited study implicitly mirroring one side to another, the proposed comparison only makes sense within the elevation angles available in both studies. The consistent accuracy across the entire hemisphere evident in our data is likely associated with the ability of our users to orient themselves towards the source, thereby utilizing the strongest acuity of their anterior spatial aural perception while side-stepping biological limitations, such as the cone of confusion [29]. This preliminary comparison suggests that hypothesis #1 is correct. While seemingly obvious, this observation may be particularly important given the prevalent use of egocentric scenarios in auditory display research to drive the design and implementation decisions. As a result, we may need to carefully consider how the design of auditory displays can fully utilize the real-world human spatial aural perception capacity and cognitive bandwidth. The confirmation of the hypothesis #1 further warrants research into a more comprehensive exploration of egocentric immersive sonification and its comparison with the exocentric approach.

Figure 2 shows a bar chart that compares the physical and virtual scenarios. All the results are obtained in the same environment, so apart from the technology necessary to allow the virtual scenario to provide immersive exocentric capability, the physical conditions of the two are essentially identical. It is therefore noteworthy that the means, medians, and standard deviations of the errors in identifying the source of a sound are all significantly larger for the binaural data, confirming the hypothesis that human perception is enhanced in the physical immersive environment. The headphone-based tests simulate the immersive environment, but the measured performance results show that these simulations add a significant error, and could in fact lead to invalid conclusions about the utility of sonification as a data analysis tool.

The angular miss-distance (E) measured in our tests and shown on the left axis of the figure is defined as

$$E = cos^{-1}[\sin(\theta_1)\sin(\theta_2) + \cos(\theta_1)\cos(\theta_2)\cos(\phi_1 - \phi_2)],$$

where, θ_1= Perceived elevation angle, θ_2 = True elevation angle, ϕ_1= Perceived azimuth angle, ϕ_2= True azimuth angle

The data in Figure 2 confirm hypothesis #2 above. While expected, this result suggests that efforts to virtualize sonification and audio immersion may impede progress in sonification research by failing to utilize the full range of human auditory capacity, including cross-domain-mapping. This suggests that further development of the binaural approach to representing immersive aural content may be warranted before we begin relying on its economy and convenience, particularly in virtual/augmented/mixed reality scenarios that may benefit from heightened aural localization resolution.

Figure 3 demonstrates that the physical immersive environment data are indicative of better performance. Somewhat surprisingly, the immersive tests do not show marked improvement when subjects are allowed to move, as compared to the cases in which they were required to stand in the middle of the room. Even in

cases where headphones are worn, the increased volume and improved angular sensitivity that might be expected to improve performance are not evident in the data in any statistically significant way. These results are therefore mixed.

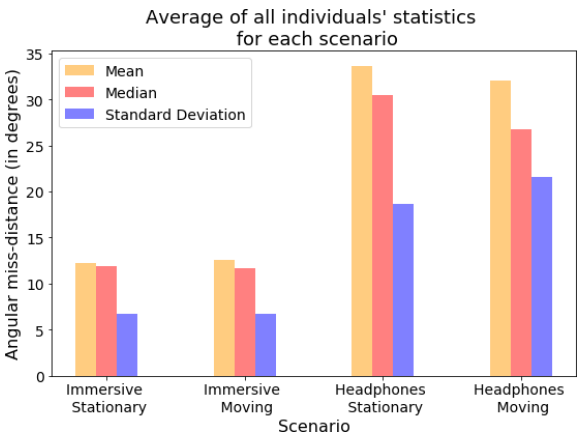
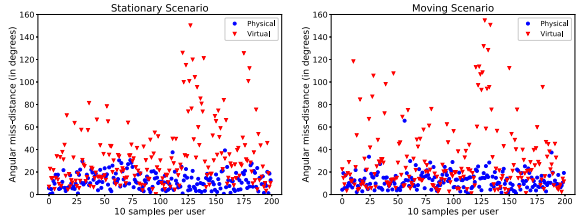


Figure 2: Comparison of the mean, median and standard deviation of the error in localizing a sound source from a group of 20 test subjects. The immersive exocentric environment consistently yields better performance than the virtual, headphone-based approach.

The preliminary data from 20-users suggests that hypothesis #3 is not true, despite the logical assumption that as a listener approaches the source their accuracy is expected to increase. While unexpected, this may be also seen as an advantage in terms of the applicability of the exocentric scenario, whereby head and body orientation may be sufficient to capitalize on the additional perception resolution afforded by the exocentric environment. We conclude that further study in this area may be warranted due to several factors:

- While accurate, human finger pointing at a distance in a space may result in deviations in the perceived location of the sound source for which the current dataset does not accurately account. We aim to address this in follow-on studies by providing more focused training of participants, which should improve their pointing accuracy. Further, the moving component may require a larger space to fully realize its impact and therefore separate the data from the two environments in a statistically significant way.
- In the exocentric scenario that allows for motion, users were confined to the central 20x20-foot space, where the motion capture worked most reliably. In the follow-on studies we intend to expand this to the edges of the space to allow for better-resolved comparisons of the two scenarios.
- Our preliminary qualitative data suggest that users who have more experience with sound and music generally perform better at localizing sound sources. We also observe that participants in general did not feel as comfortable moving around the space as they did when standing in the center, as if they were preconditioned to the stationary scenario. This interesting result warrants further study.



(a) (b)

Figure 3: The left plot shows data from 20 test subjects (with 10 points each) who were asked to locate the source of a sound from a location in the center of the Cube, and the right plot shows the result when the same subjects were allowed to move around the room during the test. Blue dots are results obtained in the physical immersive exocentric environment, and red dots in the virtual counterpart using headphones.

7.2. QUALITATIVE DATA

In addition to the quantitative data shown above, study participants were asked to answer a series of questions both before and after their experience. These data have not yet been fully analyzed, but for a few particular questions a consensus appears to emerge. These are enumerated below:

1. The overwhelming majority of test subjects experienced no discomfort or disorientation as a result of the testing process;
2. The single hearing-impaired individual among our test group became very frustrated and terminated the headphone-based test, but experienced no such effects in the physical immersive environment. Further study is warranted here, as this single data point raises intriguing questions relevant to whether immersive environments are demonstrably better for teaching hearing- and/or vision-impaired individuals who may prefer not to have their ears occluded by headphones [34].
3. The majority of the responses indicate that "Interacting with the gloves was comfortable." (Strongly agree 14, Somewhat agree 5, Strongly disagree 1). The single person who strongly disagreed mentioned in the feedback that "gloves are too tight for large people!"
4. 17 out of 20 people reported that the headphone-based test was the most challenging part of the game. A few were more specific, observing that they faced difficulties in locating sound elevations. One of the users specifically mentioned that finding the right elevation was the most challenging part of the test.

In terms of responses pertaining to improving the system, a few users mentioned that it would have been better if they were allowed to move more, and if cameras were tracking a larger area, thereby allowing them to move farther from the center. One of the users said "sometimes the glove wouldn't respond to gestures in certain places," indicating limited trackable area to move around".

8. SEEKING PATTERNS IN GEOSPATIAL DATA

Much of what is discussed here revolves around ground truths and primitives. We consider these essential elements to use as we work toward a larger goal of sonifying geospatial data. Geospatial data from the low-Earth orbit environment is a prime example of big

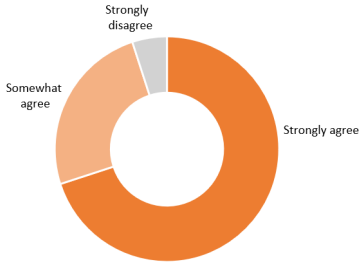


Figure 4: "Interacting with the gloves was comfortable."

data, and its inherent 3D mapping helps us side-step any potential idiosyncrasies that may be associated with arbitrary assignment of the spatial dimension in the process of sonification. As part of this project we have developed a pipeline that allows for importing such data from an empirical model and converting it into the D4 library's time-based changing spatial mask. We have extracted the model data and mapped the ensuing spatial mask to a sonification model that combines amplitude and pulse modulation of the sound source. This combination has shown greatest perception accuracy for the 20 users in our study. The geophysical data vary widely over the spatial domain, and have temporal, latitudinal, longitudinal, seasonal, and solar cycle variability. Our goal is to determine whether such large and complex data sets can be better understood using sonification techniques, and if so, to identify the sonification approaches that yield the best results. While a preliminary demonstration and a production pipeline has been implemented, a number of challenges remain in terms of appropriate sonification techniques in multilayered, multivariate scenarios in conjunction with the aforesaid spatial mask.

9. UNKNOWNS

A facet of this research that may require further attention is identifying the accuracy of the pointing technique. While clearly natural and intuitive, there is a need to further refine the interaction interface to potentially amplify the differences between static and moving scenarios. There are also other considerations, such as occlusion of the ears by long hair, and whether this may also have an effect on the observed data. The Locus system is also easily adapted to accommodate purely egocentric scenarios. Doing so will allow for a more accurate comparison of the two environments and may reveal additional ground truths. The ensuing data will serve as a foundation for a model that capitalizes on the cross-domain mapping to study human sonification capacity in the context of how we interact with the real world. Lastly, in respect to rendering point sound sources, further comparison of the algorithm utilized by D4 to other known spatialization approaches may be warranted.

10. CONCLUSIONS & FUTURE WORK

The results of our initial study confirm some of our initial hypotheses and refute others, but in almost all cases they suggest a need for additional research. We look forward to continuing these studies with significantly larger groups of test subjects in order to add statistical significance to our results. While it is too early to reach solid conclusions, our results so far suggest that the human ability to localize point-based sound sources is significantly better in physical immersive exocentric environments than in its virtual counterpart. Surprisingly, we have been unable to confirm the seemingly obvious prediction that an ability to move around in a space is helpful to the act of localizing a sound source. This suggests the potential for broader applicability of exocentric environments, including scenarios where movement is not an option. Perhaps most importantly, we show that the acuity in the immersive exocentric environment is far greater than that of the egocentric environment, which may warrant a rethinking of how we study spatial aural perception and its use in sonification and other real-world scenarios.

Future work on these and other topics is advised. In particular, a wide variety of sounds and modulation techniques should be studied to determine ground-truths that can be broadly applied to sonification of real-world data. Our work suggests that erroneous conclusions could be reached if poor choices are made in the modulation techniques applied. We have not yet tested a number of key questions, including:

1. How many distinct sounds can a user identify and/or correlate with one another in the immersive exocentric environment?
2. At what point does sonification reach the limits of human perception capabilities, and how can we recognize when this occurs?
3. How can we use the infrastructure developed for our sonification studies to enhance a user's understanding of complex multidimensional datasets?
4. Do immersive environments offer more promise than binaural techniques for teaching individuals with hearing and/or vision impairments?
5. Can we develop techniques to represent vector quantities using sonfication, and if so, how do perceptive abilities change in such cases?
6. What are the opportunities and advantages of collaborative sonification?
7. How does the conditioning in one sonification scenario translate into better performance in that specific case, and in others?
8. What role do time and stress play in localizing sources in both exocentric and egocentric scenarios?

We believe the unique infrastructure at our disposal may allow significant progress to be made on many of these topics. We look to continue this research to gain additional knowledge about the nascent field of sonification, and how it can be used to improve understanding and/or pedantic techniques. It is exciting to imagine a future in which facilities such as ours are common, and are used routinely to explore complex problems and discover new relationships in large and complex datasets. Finally we note that the infrastructure described here may be attractive to performing artists, opening doors to new means of artistic expression. There is surely much more to be learned from sonification studies, and new discoveries awaiting those with the ability to explore its potential. To facilitate this progress our goal is to make the software infrastructure and the supporting documentation publicly available to promote reproducibility and hasten progress towards the sonification of large multi-dimensional datasets and/or a Tufte-like treatise in the audio domain. This publication is an early step toward those goals.

11. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1748667.

12. REFERENCES

[1] G. L. Young, "Human ecology as an interdisciplinary concept: a critical inquiry," in *Advances in ecological research*. Elsevier, 1974, vol. 8, pp. 1–105.

[2] P. A. Kirschner, "Cognitive load theory: Implications of cognitive load theory on the design of learning," 2002.

[3] K. Hussein, E. Tilevich, I. I. Bukvic, and S. Kim, "Sonification design guidelines to enhance program comprehension," in *2009 IEEE 17th International Conference on Program Comprehension*. IEEE, 2009, pp. 120–129.

[4] P. Lennox and T. Myatt, "Concepts of perceptual significance for composition and reproduction of explorable surround sound fields." Georgia Institute of Technology, 2007.

[5] N. Mariette, "Mitigation of binaural front-back confusions by body motion in audio augmented reality." Georgia Institute of Technology, 2007.

[6] G. Kramer, B. Walker, T. Bonebright, P. Cook, J. H. Flowers, N. Miner, and J. Neuhoff, "Sonification report: Status of the field and research agenda," 2010.

[7] J. P. Bliss and R. D. Spain, "Sonification and reliability-implications for signal design." Georgia Institute of Technology, 2007.

[8] T. Hermann and H. Ritter, "Listen to your data: Model-based sonification for data analysis," *Advances in intelligent computing and multimedia systems*, 1999.

[9] T. Nasir and J. C. Roberts, "Sonification of spatial data." Georgia Institute of Technology, 2007.

[10] T. Hermann, "Taxonomy and definitions for sonification and auditory display." International Community for Auditory Display, 2008.

[11] M. Quinton, I. McGregor, and D. Benyon, "Investigating effective methods of designing sonifications." Georgia Institute of Technology, 2018.

[12] E. R. Tufte, *The visual display of quantitative information*. Graphics press Cheshire, CT, 2001, vol. 2.

[13] "Exocentric environment," Jan 2013. [Online]. Available: https://en.wikipedia.org/wiki/Exocentric_environment

[14] A. L. Shelton and N. Yamamoto, "Visual memory, spatial representation, and navigation," *The visual world in memory*, pp. 140–177, 2009.

[15] I. I. Bukvic and G. D. Earle, "Reimagining human capacity for location-aware aural pattern recognition: A case for immersive exocentric sonification." Georgia Institute of Technology, 2018.

[16] G. Lakoff and M. Johnson, *Metaphors We Live By*, 1st ed. University of Chicago Press.

[17] G. Kramer, *Auditory display: sonification, audification and auditory interfaces*. Addison-Wesley Longman Publishing Co., Inc., 2000.

[18] M. J. Morrell and J. D. Reiss, "Inherent doppler properties of spatial audio," in *Audio Engineering Society Convention 129*. Audio Engineering Society, 2010.

[19] E. Lyon, T. Caulkins, D. Blount, I. Ico Bukvic, C. Nichols, M. Roan, and T. Upthegrove, "Genesis of the cube: The design and deployment of an hdla-based performance and research facility," *Computer Music Journal*, vol. 40, no. 4, pp. 62–78, 2016.

[20] D. Sardana, W. Joo, I. I. Bukvic, and G. Earle, "Introducing locus: a nime for immersive exocentric aural environments," in *Proceedings of the International Conference on New Interfaces for Musical Expression*. Porto Alegre, Brazil: Federal University of Rio Grande do Su, June 2019.

[21] J. Halpern, "Introduction to unity," in *Developing 2D Games with Unity*. Springer, 2019, pp. 13–30.

[22] M. Wright, A. Freed, *et al.*, "Open soundcontrol: A new protocol for communicating with sound synthesizers." in *ICMC*, 1997, pp. 101–104.

[23] M. Puckette, "Max at seventeen," *Computer Music Journal*, vol. 26, no. 4, pp. 31–43, 2002.

[24] K. Beilharz, "Wireless gesture controllers to affect information sonification." Georgia Institute of Technology, 2005.

[25] M. Lech and B. Kostek, "Gesture-controlled sound mixing system with a sonified interface." Georgia Institute of Technology, 2013.

[26] M. J. Morrell, J. D. Reiss, and T. Stockman, "Auditory cues for gestural control of multi-track audio." International Community for Auditory Display, 2011.

[27] J. Sheridan, G. Sood, T. Jacob, H. J. Gardner, S. Barrass, *et al.*, "Soundstudio 4d: A vr interface for gestural composition of spatial soundscapes." in *ICAD*, 2004.

[28] J. Sterkenburg, S. Landry, and M. Jeon, "Influences of visual and auditory displays on aimed movements using air gesture controls." Georgia Institute of Technology, 2017.

[29] S. R. Oldfield and S. P. Parker, "Acuity of sound localisation: a topography of auditory space. i. normal hearing conditions," *Perception*, vol. 13, no. 5, pp. 581–600, 1984.

[30] P. Majdak, M. J. Goupell, and B. Laback, "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, Perception, & Psychophysics*, vol. 72, no. 2, pp. 454–469, Feb. 2010. [Online]. Available: https://doi.org/10.3758/APP.72.2.454

[31] I. I. Bukvic, "3d time-based aural data representation using d4 librarys layer based amplitude panning algorithm." International Community on Auditory Display, 2016.

[32] M. Krause, M. Mogalle, H. Pohl, and J. J. Williams, "A playful game changer: Fostering student retention in online education with social gamification," in *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*. ACM, 2015, pp. 95–102.

[33] M. P. Bussemakers and A. De Haan, "When it sounds like a duck and it looks like a dog... auditory icons vs earcons in multimedia environments," in *Proceedings of the international conference on auditory display*, pp. 184–189.

[34] H. Petrie, V. Johnson, T. Strothotte, A. Raab, R. Michel, L. Reichert, and A. Schalt, "Mobic: An aid to increase the independent mobility of blind travellers," *British Journal of Visual Impairment*, vol. 15, no. 2, pp. 63–66, 1997.