Introducing Locus: a NIME for Immersive Exocentric Aural Environments

Disha Sardana
Virginia Tech
Human Centered Design
Blacksburg, VA, USA
dishas9@vt.edu

Woohun Joo
Virginia Tech
Human Centered Design
Blacksburg, VA, USA
joowh@vt.edu

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

Gregory Earle
Virginia Tech
Electrical and Computer Engineering
Blacksburg, VA, USA
earle@vt.edu

ABSTRACT

Locus is a NIME designed specifically for an interactive, immersive high density loudspeaker array environment. The system is based on a pointing mechanism to interact with a sound scene comprising 128 speakers. Users can point anywhere to interact with the system, and the spatial interaction utilizes motion capture, so it does not require a screen. Instead it is completely controlled via hand gestures using a glove that is populated with motion-tracking markers.

The main purpose of this system is to offer intuitive physical interaction with the perimeter-based spatial sound sources. Further, its goal is to minimize user-worn technology and thereby enhance freedom of motion by utilizing environmental sensing devices, such as motion capture cameras or infrared sensors. The ensuing creativity enabling technology is applicable to a broad array of possible scenarios, from researching limits of human spatial hearing perception to facilitating learning and artistic performances, including dance. Below we describe our NIME design and implementation, its preliminary assessment, and offer a Unity-based toolkit to facilitate its broader deployment and adoption.

Author Keywords

NIME, gestural control, motion capture, spatial sound, exocentric, interaction, immersive aural environment, mobile, wearable, screen-less

CCS Concepts

•**Human-centered computing → Gestural input;** *User studies;* Auditory feedback;

1. INTRODUCTION

“Screen-less wearable devices allow for the smallest form factor and thus the maximum mobility” [12]. Gustafson et al. state that current screen-less devices only support buttons and gestures; pointing is not supported because users have nothing at which to point.

In this paper we introduce Locus, a NIME created as part of the SADIE (Spatial Audio Data Immersive Experience) research project funded by the National Science Foundation. Its primary focus is to facilitate user interaction with spatial aural content in an immersive exocentric aural environment [7]. This is achieved via a pair of gloves with retroreflective markers, which leverage the intuitive way we interact with the world, particularly using finger pointing and other simple hand gestures. The environmental hardware and software infrastructure tracks these markers to monitor user input and generate immersive spatial aural content. In this paper we use the term “exocentric environment” to refer to a virtual reality or other immersive environment that completely encompasses the user [21] and allows exploration of the environment from multiple vantage points. Specifically, head-motions, echoes, and phase differences based on proximity and the listener’s vantage point are fully experienced in an exocentric environment, but are not evident with the headphone-only approach to aural immersion.

In everyday experience most if not all human interactions and auditory sensations occur in exocentric environments, so these are both natural and intuitive. Further, we commonly combine sensory inputs from multiple sources to understand the stimuli in the environment around us. By doing so we increase our cognitive bandwidth [1]. For instance, we can walk around to better pinpoint a particular sound source, and by doing so we combine our senses of location, motion, hearing, and sight to form a more accurate spatial image of our surroundings. Locus is built upon this specific meaning of the term “exocentric environment”.

1.1 New Interface for Musical Expression

In Figure 1, a user is wearing gloves fitted with reflective markers, and is making hand gestures to manipulate sounds. Using both hands, he is manipulating two sounds. He can turn around and use hand gestures to direct the sound to emanate from any direction. In addition to location, the user is also able to manipulate loudness and modulation. The ensuing aural fabric is centered on the manipulation of the spatial dimension, thus providing a foundation for the exploration of form and structure through the spatial dimension of sound.

The benefit of such a system is that it enhances the artistic expression in a performance, and does so using natural

NIME’19, June 3-6, 2019, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

Figure 1: User manipulating the sound field using gloves fitted with retroreflective markers

and intuitive gestures (e.g., pointing) [14]. The user is not entangled in wires, nor do they need to use buttons or a mouse. Instead they are immersed in the physical space while manipulating the sound field, and the audience experiences a desired spatial soundscape that is further reinforced by the user/performer's natural and easy to comprehend gestures. The intuitive nature of the interface yields a system that is easy to learn with minimal training, while the related body language provides opportunities for easy communication of the user's actions and their aural manifestations.

Locus is a NIME that consists of several components: a pair of gloves with retroreflective markers, a motion tracking system, a software-driven sound engine realized in Max/MSP [17], a high density loudspeaker array (HDLA), and a Unity-based software toolkit that links all of these elements together. This ecosystem creates an opportunity for developing an intuitive interface for controlling and spatializing multiple immersive audio streams. This new NIME encompasses ideas and techniques common to wearable computing, gesture and spatial interaction, and other types of NIMEs associated with screen-minimum infra-maximum mobility [3] [11]. Several types of control interfaces in these areas have been studied to enhance the experience and interaction between a human, a computing system and an aural environment. These include a physical sound interface for real time performance [5] and a glove-type control device [8] [18] [13].

2. PRIOR WORK

Sound spatialization control interfaces are diverse. They typically rely on a custom visual widget that may be inspired by the physical controllers, and they frequently use one or more hardware controllers, such as faders, potentiometers, multidimensional and/or multi-touch surfaces, or simply a mouse and a keyboard (e.g., TouchOSC (`https://hexler.net/`) and Lemur (`https://liine.net/en/`)). Such controllers are often programmable and able to perform many different tasks, but they are not necessarily designed with sound spatialization in mind. An example of research that does focus on sound spatialization in this area is the work of Salvati et al., who developed a system [19] that allows a performer to localize the position of sounds that emanate from a musical instrument during a live performance. In an earlier study, Marshall et al. examined multiple approaches for gestural control of sound spatialization by comparing and analyzing existing software models to pinpoint the primary roles of gesture control for spatialization [15].

There is a growing interest in virtual, augmented, and mixed reality devices like Samsung's HMD Odyssey, which is designed to lower the cognitive load and make interaction with immersive sound more natural and intuitive. Deacon et al. created Objects VR [9], in which accessibility to their interactive music system is controlled by a 3D object-based control interface; their system can provide joyful interaction experiences at levels ranging from novice to expert. Locus' fundamental approach is to significantly diminish our cognitive load while providing an intuitive approach to interacting with sound source(s). For instance, one can simply and intuitively point with a finger to identify and then move a sound source to a desired location.

In this paper we describe the key elements of Locus and the early user studies performed with the system, in order to highlight the findings of the qualitative and quantitative data obtained. Much work remains to be done, but our initial studies confirm the promise of utilizing Locus as the primary control interface in an interactive design and performance studio environment that can create a new dimension for artistic expression in the performance-based arts.

3. IMPLEMENTATION

Locus was originally designed with two goals in mind: to study how humans perceive sound in an exocentric environment, and to expand the potential for artistic expression using immersive audio, with particular focus on the high-density loudspeaker arrays (HDLAs). It utilizes two off-the-shelf gloves fitted with six retroreflective markers per hand, and is designed for use with the environmental or perimeter-based 24-camera Qualisys motion capture system. The spatial marker data are extracted using the Unity gaming engine-based toolkit that includes raycasting and vector-based transformation capabilities. It is designed to facilitate data interpretation, as well as gesture and motion recognition. The system can detect simple continuous gestures, such as pinch, thumb trigger, and hand roll, which allows users to interact with the environment in a natural and familiar way. With its vector processing capabilities it is able to detect the location of a spot on the wall to which the user points with sub-millimeter accuracy, while simultaneously monitoring and responding to other hand gestures. Furthermore, it provides a convenient visualization designed to assist with system setup and troubleshooting.

Once the data are processed they are sent to a sound engine, in this case Max/MSP, that responds to the recognized gestures. Examples of simple interactions include changing the loudness of the sound, moving the sound from one location to another, or modulating the sound source. Below we discuss each component in greater detail.

3.1 Glove Controller

We chose to use optical tracking for our study so as to allow users to interact with the system as naturally as possible. The occlusion issue can be mitigated for optical tracking by using a large array of cameras that monitor the space from various angles. Spherical displaced retro-reflective markers integrated into gloves are preferred over reflective tape because they are less susceptible to occlusion [10]. We place these markers on the relevant fingers and joints to allow accurate interpretation of hand positions and motions. Joints in the human skeleton naturally possess different degrees of freedom (DoF) [16]. The distal phalange of index finger is stable, as compared to that of middle finger or pinky finger [20]. We therefore put one marker on the upper joint of the index finger, another on top of the thumb, two on the leftmost and rightmost knuckles, and two on left and right wrist joints. This configuration of passive markers allows the relevant hand gestures to be perceived accurately, while minimizing the total number of markers required.

Another benefit of the optical tracking system is that it does not require any user-specific finger model. Since the markers are fixed to a glove, which when coupled with acceptable deviations of marker positions (e.g. due to differences in hand size), provides a relatively unrestrained and easy-to-use environment [10].

For our initial implementation we have opted to use simple static gestures, like pointing with the index finger, analog and digital interpretation of thumb motion (pressing the thumb against the hand), and pinch (touching the thumb and index finger), as well as other analog gestures like roll. When coupled with arm motion and/or each other (e.g. pinch and roll together) even static gestures can become dynamic. The ensuing combination of variables and conditions can therefore generate a rich, intuitive, and easy-to-detect gesture vocabulary.

3.2 Hand Gesture Interpretation

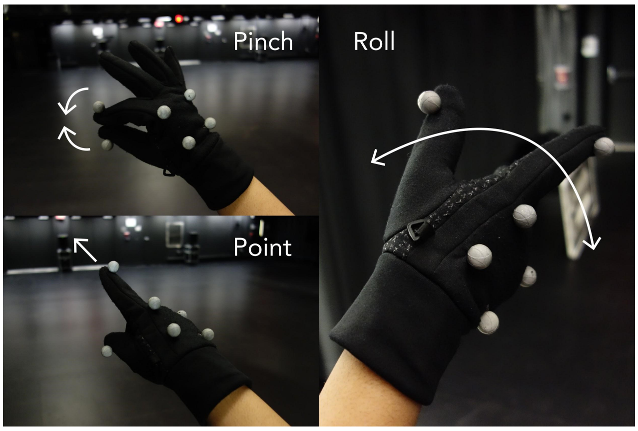

Figure 2: Hand gestures are used to manipulate the sound

Using the Qualysis Unity SDK, real time data representing the user's motion are sent to Unity with minimal latency (`http://www.qualisys.com/software/unity/`). In Unity, a virtual space is created with four walls and a ceiling that represent the true room dimensions. In the examples below we use marker abbreviations as provided in Table 1 for the right hand only. When applying the same values to the left hand, its markers are mirrored accordingly. The index finger knuckle (RKL) and index finger (RI) markers are used to calculate a vector, and the intersection between this vector and the walls in the room indicate where a user is pointing. To ensure that the user is indeed pointing we compare the vector of the index finger knuckle (RKL) and the thumb joint (RT) marker. When the angle between the two vectors is less than 20 degrees the system assumes that the user is pointing. This condition ensures that a user holding his index finger in a relaxed state that is typically not aligned with the palm will not erroneously be interpreted as pointing. In addition to carrying out the vector pointing calculations, the Unity environment allows us to easily visualize the observed markers, pointing direction, and intersection in a natural coordinate system (see Figure 3).

Pinch motions for musical expressions are detected using a proximity threshold between the thumb (RT) and index finger (RI) markers. A trigger motion is expressed using a proximity threshold between the thumb (RT) and pinky knuckle (RKR) markers, as previously described. This is more accurate than trying to detect proximity between the thumb and index finger knuckle, where overreaching with a thumb could result in a double trigger interpretation; markers are shown in Figure 2. Lastly, rolling motions of the hand are measured using a perpendicular vector from the index finger knuckle (RKL) to the pinky knuckle (RKR) and the thumb joint (RWL) to the index finger knuckle vector (RKL). The reason for the relative rotational nature of the roll is due to the human ability to place the hand in a way that, akin to a gimbal lock, may swap roles of the different axes (e.g. when pointing directly upwards), thereby making it extremely difficult to accurately detect the absolute roll without resorting to having another reference marker placed on a relatively static point on the user's torso. This choice allows the user to place their hand in any position, to possibly extend the roll range prior to initiating the actual roll (e.g. by using a trigger or some other binary action).

Table 1: Identifiable markers used

Marker	*Tag*	*Marker*	*Tag*
Left Index Finger	LI	Right Index Finger	RI
Left Thumb	LT	Right Thumb	RT
Left Left Knuckle	LKL	Right Left Knuckle	RKL
Left Right Knuckle	LKR	Right Right Knuckle	RKR
Left Left Wrist	LWL	Right Left Wrist	RWL
Left Right Wrist	LWR	Right Right Wrist	RWR

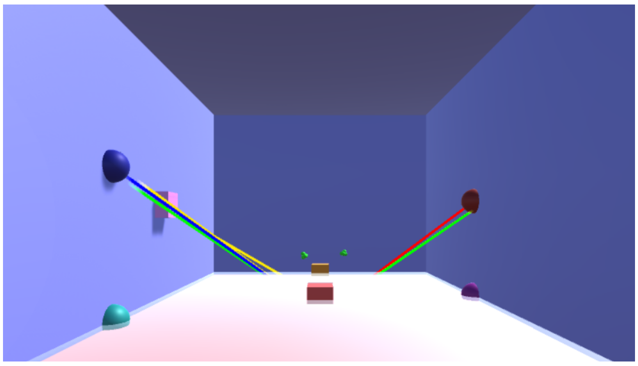

Figure 3: The image shows virtual room as seen on the Unity platform. The red ray shows where the user is pointing with his right hand, and the blue ray shows the same for the user's left hand. The blue and red spheres are the points where these rays intersect with the room's wall/ceiling. The cyan and purple spheres show the projection of blue and red spheres respectively, on the x-z plane (floor)

User motions of interest are detected and the resulting data are sent to Max/MSP over a network socket using an OSC-like format (`http://opensoundcontrol.org/`) via User Datagram Protocol (UDP) packets designed to be easily parsed using Max/MSP's built-in [route] object.

The commands sent from Unity are formatted hierarchically, as follows: <left, right> <pinch, thumb, roll, handazel, pointazel> <value 1, value 2,...,value n>

The first argument defines which hand the data packet refers to, and the second argument defines the kind of data being sent. It is followed by one or more floating point values, depending on the type of data. Table 2 offers additional details for each of the commands for a specific example of sound manipulation.

4. USER STUDIES

Table 2: Recognized Commands

Command	*Arguments*	*Description*
handazel	Absolute angles in degrees expressed as floating-point values with respect to the spatial origin	Offers azimuth and elevation angle at the collision point on the wall extended by vector (RKL - RWL). Determines whether hand is above ground or not.
pointazel	Absolute angles in degrees expressed as floating-point values with respect to that spatial origin	Offers azimuth and elevation angle at the collision point on the wall extended by vector (RI - RKL). When compared to the handazel value, it can be used to determine whether a user is pointing or not.
pinch	Distance in meters expressed as a floating-point value	Offers distance between RI and RT. It can be used to construct a threshold-based on/off value for detecting pinch, or as a dynamic range value between the two markers.
thumb	Distance in meters expressed as a floating-point value	Offers distance between position vectors RT and RKR. It can be used to construct a threshold-based on/off value for detecting trigger, or as a dynamic range value between the two markers.
roll	Relative angle in degrees expressed as a floating-point value	Offers relative roll defined by the RKL-RKR vector in respect to the RWL-RKL vector. Due to lack of additional markers, it is relative and is typically associated with a triggering command, such as threshold-driven pinch and thumb gestures.

We conducted user-studies to investigate users' ability to interact intuitively with the glove interface and the spatial aural content. The goal of the study was to test the functionality of the Locus system, including how accurately the design and implementation are working, and to analyze the ease and expressivity of human subjects' interaction with the spatial sound. The experiments required people to be in a room containing a HDLA and a motion capture system that would minimize any spatial fidelity issues commonly associated with sparser multichannel loudspeaker arrays. The wearable controller (gloves) with its retroreflective markers conveys hand position and motion data to the Qualisys motion capture system that is used for tracking movements. The hand-position data are extracted using a custom toolkit built in Unity, which creates a virtual environment in which hand gestures can be recognized and interpreted. Once this information is processed, it is sent to another software element (Max/MSP) that responds to the gestures detected.

The whole system requires a careful calibration. The Qualysis Streaming follows a left-handed Cartesian coordinate system, Unity software interprets data in a right-handed coordinate system (requiring a simple vector transformation), and the D4 library is implemented so as to follow the left hand rule. Calibration is done with respect to world stabilized axes, not body or head stabilized systems[4]. The active spatial region in the room is calibrated so that the center of the room is the origin of a fixed-axis, Cartesian, three-dimensional space. The speakers within the room populate the walls and ceiling area at or above the horizon, and the origin for the vertical (Y) axis is defined to be at floor level in the geometric center of the room.

4.1 Apparatus

Motion Capture

For this study, we used 24 Qualisys Oqus 500+ motion tracking cameras and an AIM (Automatic Identification of Markers) model offered by the Qualisys QTM. The advantage of using an AIM model over a rigid body is that it allows for finger motion without compromising recognition of the body, as long as the system is properly trained prior to its use. Once trained, the AIM model is capable of identifying the trained object regardless of hand size, finger or hand position, or orientation.

Immersive Sound

Unity toolkit's OSC-like output formatting allows it to interface with a wide variety of network-enabled digital signal processing software. In our case, we used Max/MSP for all our sound-related processing. This is in part to leverage the functionality of the D4 system [6], which is an audio spatialization library designed specifically for HDLA low-latency scenarios. This setup allows us to distribute sound across the 128 loudspeakers with a high degree of control.

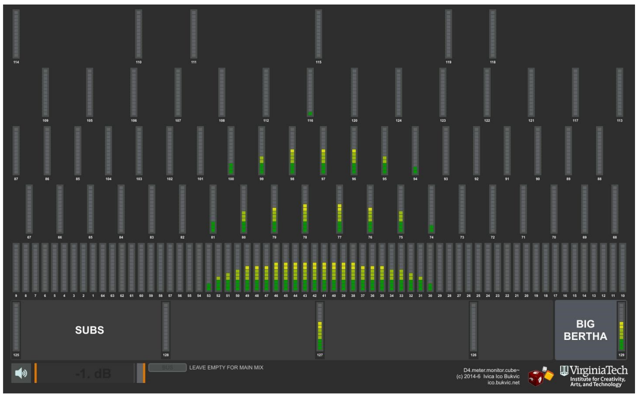

Figure 4: D4 monitor

4.2 Tasks and Procedures

Subjects were asked to interact with the glove controller to manipulate two sounds. We selected two easily distinguishable sounds that were perceived as pleasing, and therefore easy to listen to over extended periods of time. One was the noisy sound of cicadas and another a spectrally-distinct collage of pitched wind chime sounds. While these are not interpretable as a musical composition, they are designed to promote musicality. The primary goal of this scenario is to allow users to focus on natural spatialization, including sound source location, trajectory, loudness, and modulation. Before the users begin interacting with the glove, clear instructions are given and one of the investigators demonstrates available gestures and their perceived outcomes. The first and most intuitive gesture is 'Pointing'. Pointing is to be done such that the markers on the index finger and hand are in a straight line (see Figure 2). Lowering the pointing finger while keeping the hand up will stop pointing. Users can use this to instantaneously relocate the sound to a different location by moving the hand without pointing and then resuming the pointing stance. They can also move the sound in a continuous manner from one point to another by moving their index finger in pointing form to the desired location. In the thumb-trigger gesture users bring their thumb closer to the index finger. This is used to change the loudness of sound. Moving the thumb closer to the finger decreases the loudness, and moving it away from the hand increases it. When used in its extreme positions, one can also use this as a toggle switch, since minimum loudness is zero, and the user can turn on and off sound at any point. Rapid triggering motions create low frequency amplitude modulation, which proved as a useful derivative gesture in musical expression scenarios. Another allowable interaction includes two gestures implemented together. For example, pinch and roll produces periodic swells or sound modulation; these work regardless of where the hand is pointing (above or below the horizon). Akin to an infinite knob, it is relative in nature.

After trying out the glove controller for about 5 minutes, users were asked to give both numerical ratings and qualitative feedback on their experiences. We ask the following specific questions and record the users' responses. The first section of the questions is based on the Likert scale [2] from 1 to 5, and the second section is devoted to short answer questions.

Section 1

1. Interacting with the interface was comfortable.
2. Interaction with the system was unnatural.
3. I felt the system was easy to use.
4. The experience for me was not as engaging as other approaches to interactively spatializing sound.
5. I experienced discomfort during the experiment.
6. It was a physically tiring experience.
7. I found the interaction intuitive.
8. I found the experience frustrating.
9. This approach to interactive spatialization is not as natural than other approaches to interactively spatializing sound.
10. I felt the training was helpful.
11. The experience was fun and interesting.
12. This approach to interactive spatialization is more intuitive than other approaches to interactively spatializing sound.
13. I found the interface to have expressive potential
14. I experienced no issues with the process and the infrastructure (e.g. controller interface, acoustics in the room, preliminary briefing material, etc).
15. This approach to interactive spatialization is overall better than other approaches to interactively spatializing sound.

Section 2

1. What aspect of the experience did you find most challenging, and why?
2. Please briefly state any comments pertaining to fatigue, usability, the potential of the system, expressiveness, and how it may compare with existing techniques.
3. Please list other approaches to interactively spatialize sound you have previously used and briefly describe them.
4. Please share any other feedback you may have for improving the interface below.

4.3 Participants

As of writing this paper a total of seven test subjects participated in the user study. The study was open to adults who did not have any hearing or mobility issues. Participants were recruited from students, faculty, and staff. Three of the test subjects were female and the rest were male, ranging in age from 26-39. All the participants came voluntarily, and no financial or other rewards were given to encourage participation. Most of the subjects that came were right-handed except one. Six of them had previous experience with a gesture-based device. Only one of them wasn't sure if they had been previously exposed to some form of spatial sound environment, but the rest of them had prior exposure. All subjects were asked to confirm that they knew how to abort the test in case of any discomfort before the study began. The study was granted an Institutional Review Board approval.

5. PRELIMINARY OBSERVATIONS

The most notable observations during our preliminary evaluation of the system are linked to its intuitive nature. In preliminary tests the system was observed to have a minimal learning curve, making it possible for any user to immediately engage in the system, regardless of prior experience with immersive sound and/or music-making. The simple and natural motions, such as finger pointing, thumb trigger, pinch, and hand rolling motions are skills that normal humans attain and master at an early age. The ability to concurrently move two sounds around in three dimensions by simply pointing to a wall while simultaneously modifying each sound's parameters generated a sense of excitement, immersion, and a seamless connection with the spatial aural material. Further, in their exploration of the interface's vocabulary and mapping of gestures to sound, the users were able to quickly discover new and unanticipated ways of interacting.

One of the users mentioned it as "*a really cool experience*". The same user gave us feedback that the interface "*feels magical because I'm pointing at something and the interaction components are invisible.*"

One of the challenges as described by a user: *"Volume control was challenging, but I believe that is due to limited exposure. Were I "playing" the gloves as an instrument regularly that would no longer be an issue (I would gain muscle memory)."*

An issue that users reported in the survey was fatigue. Users had to keep their arms in the air for performing specific gestures. This may be in part because only some gestures could be performed while keeping the hands down. For long performances fatigue could be a factor, as mentioned in the feedback. One subject mentioned that "*There is potential for fatigue if you choose to spatialize a often and over longer distances (which I did), but as with any other instrument, that can be fixed with practice. The potential for fatigue and expressivity is on par with all other analog instruments I currently play, some of which are fatiguing by their sheer physicality (cello).*" Another person mentioned, "*I did notice slight arm strain during use. this might limit adoption but I don't believe it should be considered an issue, more a feature*".

One suggestion that we received was *"Keep going in this direction. It's already shaping up to be a highly intuitive instrument. I would strongly recommend bringing in more performers to "play" the gloves, from as wide a variety of backgrounds as possible to make it truly human-centered"*.

Based on the questionnaire's results, 100% of the subjects agreed that they found the interface to have expressive potential.

6. CONCLUSION

In this paper we present the design and implementation of Locus, a NIME designed specifically for an immersive exocentric aural environment. It is driven by a simple glove interface in combination with an environmental motion capture technology and a 128-loudspeaker HDLA. Locus' ability to efficiently leverage a HDLA makes it also more broadly applicable, including sparser and more readily available loudspeaker arrays whose computational overhead is significantly lower. The ensuing ecosystem of tools and toolkits allows for near seamless transition between various HDLA configurations. The observed interaction suggests that Locus and the supporting toolkit has the potential to significantly lower the cognitive load in both artistic and research scenarios utilizing immersive exocentric aural environments. Further, its intuitive design pushes the boundaries of interactive spatialization, allowing for one hand to control multiple dimensions of a single sound, while also offering clear and natural interaction and its projection onto the audience. Encompassing the spatial component is arguably the last frontier of sound that remains to be exploited as a primary driver of musical structure and expression. 21st century advances in computing power and reduced latency have set the stage for this transformation. Locus may serve as an important catalyst in this process. The supporting toolkit designed in Unity offers a rich platform for future augmented and virtual reality implementation that may further enhance the multisensory experience of the spatial component of sound, and its under-explored potential as the structural driver of musical expression. Further, Locus has the potential to enhance learning in educational scenarios, with particular focus on immersive and inherently spatial data. Our goal is to build upon these preliminary findings with more in-depth studies.

Locus was designed for use with the Cube's environment. However, its toolkit is designed to easily adapt to other spaces. We aim to make the toolkit publicly available by the summer of 2019.

7. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No.(1748667). The authors would also like to acknowledge Virginia Tech's Institute for Creativity, Arts, and Technology (ICAT) for the project support.

8. REFERENCES

[1] K. P. A. Cognitive load theory: implications of cognitive load theory on the design of learning. *Learning and Instruction*, 12(1):1–10, Feb 2002.

[2] I. E. Allen and C. A. Seaman. Likert scales and data analyses. *Quality progress*, 40(7):64–65, 2007.

[3] T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, 36(7):28–35, 1993.

[4] M. Billinghurst, J. Bowskill, N. Dyer, and J. Morphett. Spatial information displays on a wearable computer. *IEEE Computer Graphics and Applications*, 18(6):24–31, 1998.

[5] B. Bongers. Physical interfaces in the electronic arts. interaction theory and interfacing techniques for real-time performance. *In M. Wanderley and M. Battier, eds. Trends in Gestural Control of Music. Ircam - Centre Pompidou*, 2000.

[6] I. Bukvic. 3d time-based aural data representation using d4 library's layer based amplitude panning algorithm. *International Conference on Auditory Displays. Canberra, Australia.*, 2016.

[7] 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.

[8] G. Costantini, M. Todisco, and G. Saggio. A wireless glove to perform music in real time. In *8th WSEAS International Conference on APPLIED ELECTROMAGNETICS, WIRELESS and OPTICAL COMMUNICATIONS, Malaysia*, 2010.

[9] T. Deacon, T. Stockman, and M. Barthet. User experience in an interactive music virtual reality system: an exploratory study. In *International Symposium on Computer Music Multidisciplinary Research*, pages 192–216. Springer, 2016.

[10] K. Dorfmuller-Ulhaas and D. Schmalstieg. Finger tracking for interaction in augmented environments. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pages 55–64, 2001.

[11] M. Fukumoto and Y. Tonomura. "body coupled fingerring": wireless wearable keyboard. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 147–154. ACM, 1997.

[12] S. Gustafson, D. Bierwirth, and P. Baudisch. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 3–12. ACM, 2010.

[13] C. Lai and K. Tahiroglu. A design approach to engage with audience with wearable musical instruments: sound gloves. *in Proceedings of the international conference on new interfaces for musical expression, Ann Arbor, Michigan*, 2012.

[14] 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*, 72(2):454–469, Feb. 2010.

[15] M. T. Marshall, J. Malloch, and M. M. Wanderley. Gesture control of sound spatialization for live musical performance. In *International Gesture Workshop*, pages 227–238. Springer, 2007.

[16] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):677–695, Jul 1997.

[17] M. Puckette and D. Zicarelli. Max/msp. *Cycling 74 (1990): 1990-2006.*

[18] F. G. H. P.-W. S. M. S. Serafin, S. Trento and T. Mitchell. Controlling physically based virtual musical instruments using the gloves. *in Proceedings of the international conference on new interfaces for musical expression, London, United Kingdom*, 25(2):521–524, 2014.

[19] D. Salvati, S. Canazza, and A. Roda. Sound spatialization control by means of acoustic source localization system. In *Proceedings of the 8th Sound and Music Computing Conference*, pages 284–289, 2011.

[20] L. Shao. Hand movement and gesture recognition using leap motion controller. *Virtual Reality, Course Report*, 2016.

[21] A. L. Shelton and N. Yamamoto. Visual memory, spatial representation, and navigation. pages 140–177, 2009.