New Interfaces for Spatial Musical Expression

Ivica Ico Bukvic
Virginia Tech
Blacksburg, USA
ico@vt.edu

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

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

ABSTRACT

With the proliferation of venues equipped with the high density loudspeaker arrays there is a growing interest in developing new interfaces for spatial musical expression (NISME). Of particular interest are interfaces that focus on the emancipation of the spatial domain as the primary dimension for musical expression. Here we present Monet NISME that leverages multitouch pressure-sensitive surface and the D^4 library's spatial mask and thereby allows for a unique approach to interactive spatialization. Further, we present a study with 22 participants designed to assess its usefulness and compare it to the Locus, a NISME introduced in 2019 as part of a localization study which is built on the same design principles of using natural gestural interaction with the spatial content. Lastly, we briefly discuss the utilization of both NISMEs in two artistic performances and propose a set of guidelines for further exploration in the NISME domain.

Author Keywords

gestural control, comparative study, sound spatialization

CCS Concepts

•Applied computing → Sound and music computing; •Human-centered computing → Interaction techniques; Interaction devices;

1. INTRODUCTION

Although the earliest research in the spatial electroacoustic music can be traced back to mid-20th century Stockhausen's seminal work that proposes treating spatial sound as another dimension of musical expression [17], it is only in the last decade that we have finally seen a more widespread and systematic exploration of the spatial dimension as the primary driver of the musical form. Therefore, when considering the historical development of the musical language, the 21st century may be seen as the emancipation of the spatial sound. With the increased access to venues and spaces equipped with high density loudspeaker arrays (HDLAs) [5], there is a growing interest in new interfaces for spatial musical expression (NISME) that can facilitate the utilization of the spatial sound's full expressive potential. Although previous work with multitouch controllers has a rich history within the NIME community [8], only a relatively small subset focuses on spatialization, with the most recent example including ArraYnger [16]. Other examples of NIME research in the area of spatialization include gestural controllers, such as SoundMorpheus [3], innovative loudspeaker systems, e.g. spatial.motion [10], and gesture-based spatial sonification [2].

In this paper we present a new approach to manipulating spatial sound using the Monet NISME that consists of a multitouch controller with pressure sensing capacity, the D^4 library's spatial mask [4], and a new form of digital sound manipulation designed to amplify spatial dimension. Further, we compare it to a previously introduced NISME, Locus [15] that is inspired by the same design constraints and goals while approaching the implementation from a distinctly different perspective. Based on the user study data, we propose guidelines for the implementation of future new interfaces for spatial music expression.

2. DESIGN CONSTRAINTS

Several considerations drove the choice of Monet and its implementation. The same were also used in the development of the Locus NISME that will be used for comparison. Said considerations build on the general NIME guidelines [13], focusing on:

- Embodiment, expressivity, and accuracy: providing a level of embodiment (cognitive affordances) that can engage and empower the performer while clearly communicating the performer's actions to the audience and correlate them with the ensuing aural output, while doing so efficiently and accurately;
- Scalability and adaptability: able to scale to support high density loudspeaker array scenarios and different spatial configurations, while supporting a broad array of digital signal generation and processing approaches;
- Collaborative capacity: support for multiple concurrent agents (performers), and
- Intuitive/natural interaction: maximum possible use of natural and intuitive ways of interacting with the instrument that balances learning curve with challenge and engagement.

Given the shared design constraints, the comparison between Monet and Locus is deemed appropriate. Further, Locus has been already validated as a controller in a previous localization study [15] in respect to its embodiment, accuracy, scalability, adaptability, and intuitive and natural interface, thus leaving only artistic expressivity as an aspect we will revisit as part of this study. This allows us to use it as a foundational reference that may help also provide an insight in Monet's ability to address the same.

NIME'20, July 21-25, 2020, Royal Birmingham Conservatoire, Birmingham City University, Birmingham, United Kingdom.

3. MONET

Monet is a NISME focusing on effectively painting a sound across the perimeter-based spatial canvas denoted by the HDLA. It consists of two components, the controller and an instance of a spatial sound engine. Below we discuss each in greater detail.

3.0.1 Controller

Monet uses Sensel Morph [11] pressure-sensitive multitouch controller in conjunction with MaxMSP [14] using D^4 library's spatial mask. Unlike prevalent approaches to touch-based audio spatialization that tend to focus on the source location and spread (a.k.a. radius), Monet explores unique affordances of the spatial mask which, akin to an image mask, allows for spatial projection using amorphous shapes that can mimic one or more simultaneous point sources, as well as other, more complex geometric shapes. As a result, Monet treats Morph's surface as an unfolded perimeter of the space within which the sound is being projected, resembling an inverse world map that is being painted from inside the space. At the center of the Morph's multitouch surface is the 0° azimuth and 0° elevation point, with the left and right edges denoting -180° and +180° azimuths, respectively. Elevation at the top of the surface is +90° and -90° at the bottom. Monet leverages jit.sensel external's ability to generate a jitter matrix of Morph's multitouch surface (using jit.sensel) that also reflects that pressure using single grayscale channel ranging from black (complete absence of pressure or 0) to white (maximum detectable pressure or 1) and the D^4 library's ability to use the same to dynamically calculate the spatial mask. The resulting controller allows for multitouch sound spatialization utilizing as many contact points as the human anatomy allows in order to theoretically generate any complex amorphous shape. When coupled with the optional jit.slide matrix delay of the Morph's pressure output that is toggled via a keyboard press, resulting in a shape and consequently the sound that fades over time, the ensuing system allows for complex manipulations of the mask with minimal training. While conceivably the same multitouch surface can be shared among two or more performers, due to its limited size it is primarily envisioned to allow for multiuser interaction using multiple Morph surfaces. Given the visual nature of the way Monet affects the spatial mask and the Morph's inability to display visual data, the interface is coupled with a visual display within the MaxMSP patch that was used as part of the user's familiarization with the interface.

3.0.2 Spatial Sound Engine

By itself the controller may seem superfluous considering the research on auditory spatial perception suggests, when projecting the same sound over multiple loudspeakers, humans are at best capable of detecting the perceived center (a.k.a. source position) based on the amplitude relationships between multiple loudspeakers, effectively resulting in a more complex example of a simple amplitude panning algorithm. The same may also allow for the listener to detect the overall size around the perceived center. Given, however, D^4 library's ability to assign a per-channel algorithm by which the sound output may be modulated based on the calculated spatial mask, the resulting system offers unique opportunities for perceivable spatial manipulation of sound regardless the sound source and the ensuing amorphous shape generated by the spatial mask.

Monet's default implementation of the per-channel spatial mask digital signal processing utilizes a normalized inverse sawtooth wave with values ranging from 0 to 1, resulting in sharp periodic attacks and varying rate of decay (modulated by an exponent) as the incoming signal modulator. As the spatial mask becomes stronger, the said sawtooth generator (phasor object) increases in its frequency, thereby increasing the number of periods per second, and changes its exponent from a value above 1 to that between 0 and 1, effectively inverting the exponential decay consisting mostly of silence into that of mostly signal with only a temporary notch-like decay before the consequent attack. Lastly, the output is further modulated by the overall mask value, effectively rendering dark areas entirely silent, the grayscale progressively louder, and the white loudest. As a result, dimmer (near dark) areas of the spatial mask introduce slow sporadic and generally soft-spoken impulse-like amplitude spikes of higher frequencies of the original signal. With the increase in the mask's brightness the spikes become louder, denser, and their decay approaches and eventually exceeds that of linear, resulting in decreased interruption in the overall sound between the attacks.

The instrument's default sound is a 30-second loopable sample of a complex Spectrasonics Omnisphere synth pad with multiple layers that, when spatialized using Monet's default implementation of the per-channel modulation, allow for a rich and dynamically changing fabric. The resulting sound manipulation can be previewed using the following link. Although specific in its default implementation, the resulting digital signal processing implementation provides a novel approach to interactive spatialization and allows for flexible substitution of both the input signal and the digital signal processing chain, including the possibility of introducing multiple variants of same or similar (related) sounds (e.g. that of dripping, running, and gushing water) and varying their simultaneous intensity based on the spatial mask.

3.0.3 Interactive Spatialization

The resulting NISME allows for a novel manipulation of a single stem (a.k.a. sound) or a live audio signal in that one sound can produce dramatic variations across the HDLA perimeter without compromising perceivable relationship among the spatial-mask-modulated variations of the source sound projected over different loudspeakers. Depending on the frequency range of the modulating sawtooth wave, the resulting soundscape can be best described as anything from a sparse to a dense polytempo aural fabric that allows for the perception of amorphous areas and contours with greater and lesser intensities.

3.1 Locus

The design goal and implementation of the Locus NISME is described in a previous publication [15]. Below we summarize relevant details to provide context for a comparative study.

Locus's goal was to create an intuitive controller offering natural hand and arm-based sound source localization using perimeter-based audio spatialization. This was achieved through two off-the-shelf gloves with reflective markers, a 24-camera Qualysis motion capture system for gesture recognition, Unity's raycasting capabilities [9] to interpret user's hand gestures, and MaxMSP [14] providing spatial audio feedback. The user-worn technology was kept to a minimum to enhance the freedom of motion. Locus was tested as a user study controller in exploring limits of human spatial hearing perception [7, 6], as well as a NISME in artistic performances [1]. Consequently, its goals and design constraints are identical to that of Monet.

For the purpose of a comparative study, Locus was coupled with a sampled sound of cicadas and wind chimes that were assigned to the left and right hand, respectively. Although the potential glove-based vocabulary can be considerably more complex, in this study we utilized three basic hand gestures (see Figure 1 (b)): index finger pointing for changing the position of the respective sounds, thumb motion for loudness with amplitude decreasing as the thumb gets closer to the index finger knuckle (the metacarpophalangeal joint), and, similar to the Monet, generates a sawtooth-wave-driven amplitude modulation that can be controlled via a relative additive hand roll with index and middle fingers extended that mimics a manipulation of a radio dial. Modulated sounds are fed into the D^4 library that spatializes two sounds within the facility according to the motion capture data processed by Unity. In order for the sound movement to work, users had to point with their index finger with no bend between the distal, middle, and proximal phalanx. Further, pointing below the elevation of 3° where the lowest layer of speakers is located prevented any registered hand movement. Additional details regarding Locus gestures are provided in the Study Design section below.

4. USER STUDY

In order to assess the efficacy of the newfound NISME we implemented a study with 22 participants using a within-subjects study design in which users have an opportunity to shape spatial sound using Monet and Locus. The study took place inside a cuboid venue measuring 50x40x32 feet [12]. The facility consisted of a Qualysis 24-camera motion capture system enabling tracking of the users' motion and gestures. The facility offers homogeneous layers of 124 speakers at three different heights and a layer at the ceiling supporting elevations from 3° and 90°, and another 6 subwoofers, all of which provide user with an immersive audio world-stabilized (a.k.a. exocentric [7]) environment. The high-density loudspeaker array environment is compatible with all the current spatialization algorithms available including the Layer Based Amplitude Panning (LBAP) algorithm that is utilized by the D^4 library which also offers the spatial mask capacity.

4.1 Study Design

4.1.1 *Pre-Test Conditions*

The study described in this paper has been administered as a second part of a larger, Institutional Review Board-approved localization study in which Locus was used as a means of recording participant-perceived azimuth and elevation of a sound source. As a result, prior to this second component of the study, all participants had prior experience with the Locus as a controller. We deem this bias as insignificant given the Locus' role and interaction in this part of the study was fundamentally different. For the sake of efficiency, all tests started with the Locus system which users were already wearing before starting the expressive comparison study discussed in this paper. Each user was subjected to a hearing test, the demonstration of the Locus and Monet NISMEs and a training round, respectively. Further, in the case of the Locus there was a brief calibration phase to ensure that the glove markers are properly positioned and are being reliably and consistently detected by the system.

4.1.2 *Free-Play Testing*

The users were first asked to explore the Locus NISME and use it to manipulate the spatial soundscape consisting of the two aforesaid perimeter-based sounds. The choice of sound was driven by ensuring that the ensuing soundscape is appealing, while minimizing auditory fatigue. Although there was no time restriction for this test, most users took no more than 5 minutes in which they could manipulate the sound as shown in the Figure 1.

Following the Locus free-play test, users were given an opportunity to use the Monet NISME. Its interface allowed for sound manipulation using either or both hands. To move the sound around, the users placed their finger on the device surface, and moved it along the desired direction. Given the optional built-in jitter matrix delay that gradually dissipated over time the mask generated by the user hand pressure, users could also make a more complex drawing, including amorphous shapes (Fig.1). The resulting sound emerged out of the speakers in the room based on where they pressed on the surface, and gradually dissipated once such a pressure was gone based on the assigned decay level and the jitter framerate. Considering the test venue did not have any loudspeakers at elevations below 3°, spatial mask values generated by Monet below 3° were used to simulate lower hemisphere by panning across the space. Like in the case of the Locus system, there were no time restrictions and users took up to 5 minutes to play with the device.

4.2 Data Collection

At the end of each of the two scenarios (Locus and Monet) users were given a post-evaluation survey. The questions sought feedback regarding each NISME's level of comfort, intuitiveness, easiness, engagement and expressive potential. The questions were repeated to check for consistency in users' responses. All study questions, responses, and response highlights are provided online via the following link.

4.3 Participant Demographics

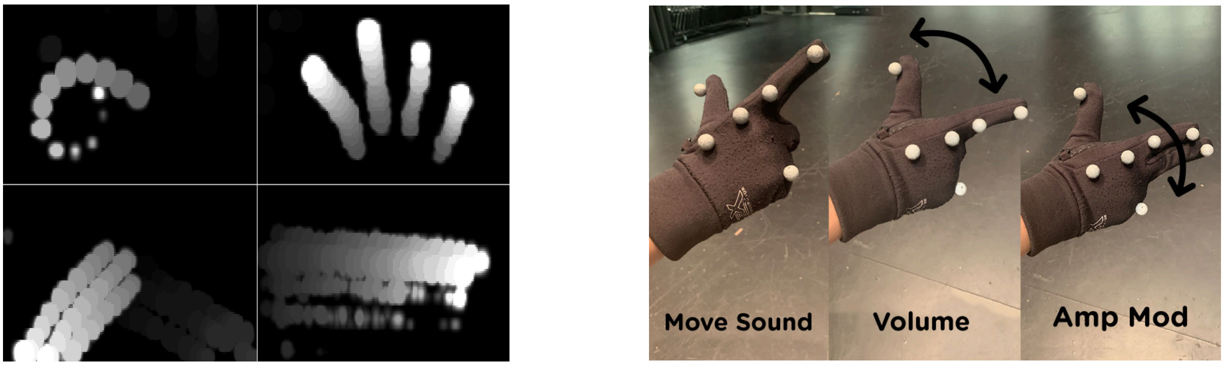
Figure 2 shows the demographics of the 22 study participants. Six of the participants were female, and most of the participants were between the age range (20-32). Two participants were left-handed. None of them had a known physical disability or mobility issues that would prevent them from hearing spatial sound, making hand gestures, and/or moving around the space. Only five users did not have a prior experience with gesture-based interaction devices. All users prior experience with music, either through casual listening, or being a music enthusiast. 45% of the participants had prior exposure to an immersive spatial audio environment.

4.4 Participant Feedback

Figure 3 shows a summary of user responses collected through the questionnaire. We discuss them further below. In addition, we offer a summary of the subjective feedback that offers richer detail of user experiences with the respective NISMEs, including their intuitiveness and expressiveness, the most challenging aspects faced during the experience, usability, and suggestions on how to improve them.

5. DISCUSSION

While the two NISMEs fulfill similar roles, their approach is fundamentally different. Locus, focusing on the natural and intuitive pointing gesture, offers a straightforward sense of embodiment–user simply needs to point in a direction where they wish the sound to emanate and then use one of the two simple gestures to control and communicate changes to sound's amplitude and modulation. Monet, being bound to a relatively small multitouch surface and requiring users to exaggerate their otherwise subtle hand movements in order to be able to visibly communicate them to others by default was expected to offer slightly lesser sense of embodiment. In this case, visual projection may play a critical role in helping demystify the relationship between the NISME, user's



(a) Monet (b) Locus

Figure 1: Interfaces

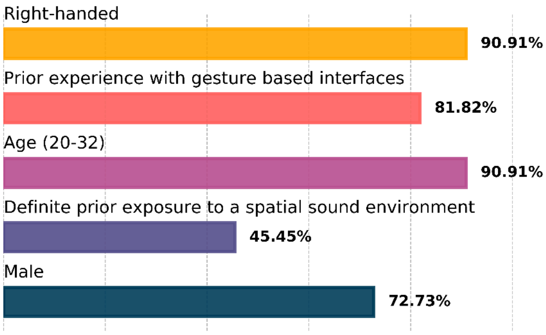


Figure 2: Participant Demographics

gestures, and the spatial aural output. One the other hand, Locus is tied to the traditional notion of spatialization, using two point sources (one per hand) that are being moved around the space. Monet challenges this notion by introducing "painting" of spatial sound across the HDLA canvas, thus enabling users to produce complex motions and amorphous shapes that are difficult if not impossible to generate using the point source approach, even when coupled by its more advanced features, such as the point source spread or radius. For instance, by moving multiple fingers from the top of the Morph's surface downwards, a user can generate multiple trajectories of the same, yet differently modulated sound source that appear to descend from the ceiling and envelop the listener from all sides, resulting in a complex combination of trajectories that can be only recreated by a large number of point sources. Further, to sidestep the limits of human spatial perception, when it comes to the projection of the same sound from multiple loudspeakers, Monet utilizes D^4 library's spatial mask to modulate the sound per each physical channel, thereby creating a unique soundscape that is simply impractical to recreate using the point source approach.

Despite the Monet and Locus' fundamentally different approaches, the survey data suggests they offer strikingly similar levels of usability, expressivity, comfort, learning curve, and intuitiveness. Digging deeper into the data, there are a few subtle but nonetheless important differences. On average users felt Monet was slightly less frustrating to use, while some reported control issues with Locus. This may be in part due to lesser complexity of Monet's implementation which led to fewer situations where it failed to accurately respond to user's input. Such issues were associated predominantly with Locus due to its reliance on the motion tracking system that requires optimal coverage of the user's hand in order to accurately detect hand position, motion, and gestures. The same may be also due to the fatigue we address further below.

Despite introducing added cognitive load, requiring users to map the 2D multitouch surface onto the 3D perimeter and compensating for potential changes in their orientation, data suggests there is no significant difference in the intuitiveness of the two NISMEs. Monet did feel slightly easier to use, while Locus felt more slightly intuitive. Locus did have somewhat larger reported level of discomfort and fatigue due to the fact that users had to hold their arms up in order to project the sound in a particular direction. The same may be also attributed to the fact that Locus was used immediately before for the first part of the study, as noted above. It is worth noting that the question of which interface is better referred to all NISMEs, not just the two used in this comparison. The analysis of this question proved somewhat problematic in respect to other NISMEs, as a large number of users reported neutral responses following the experience with Locus which may have been in part because they had no such prior experience. That said, the survey of Monet still provided a valuable comparison with Locus.

In their qualitative response, users also noted collaborative potential in both NISMEs in artistic and scientific scenarios, particularly the Locus. The overall slight preference for Locus, while not statistically significant, may in part stem from the bias of the same being used exclusively in the first part of the study that focuses on the research of the foundational scientific concepts [15] which may have helped users facilitate their understanding of the NISME and its capabilities.

One of the notable areas of improvement were the aforesaid reliability of the Locus system. Similarly, it appears that the Monet's default implementation could be further improved by better scaling of the grayscale image information onto a logarithmic nature of the human perception of loudness. Due to computationally heavy jitter patch, Monet also had some lag and users indicated that the use of a tablet instead of Sensel would lower the cognitive load by integrating the multitouch surface over the visual display of the spatial mask, thus providing a visual feedback akin to that of a painted canvas. Doing so at this time would sacrifice Morph's unique ability to sense pressure. Other suggestions included the integration of virtual buttons, like the one for toggling of the decay for which the test relied on a physical keyboard. This is something that Sensel Morph supports through overlay inserts and may be worth further exploration. Some user comments suggested the lack of understanding of the different modes of operation, such as Monet's disabling of the decay which would stop the sound as soon as finger(s) lost contact with the multitouch surface,

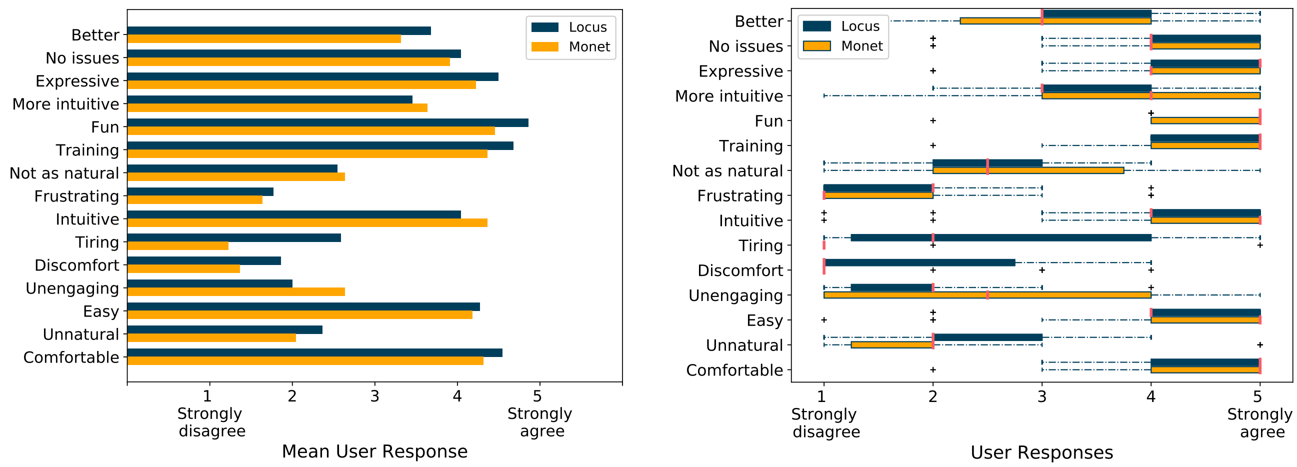


Figure 3: Survey Response Statistics: (a) Mean of the user-responses, (b) Box plots for the user-responses where the spread of the box suggests the values of the discrete user-responses between 25th and 75th percentiles of the total responses, the red vertical line is the median value of the user-responses and the + sign indicates outliers.

which to them felt was less controllable. Some went as far as expressing preference for the "tail", or the decay envelope which felt more natural and organic. The same mode felt somewhat limiting as the fingers had to maintain contact with the surface at a desired location in order to ensure that the sound remained on. Users have also expressed interest in additional gestures for the Locus, something that the interface has been designed to accommodate and which was further explored in the artistic performances described below. A complete qualitative user feedback can be accessed via the following link. A majority of issues reported by the users have been alleviated following the study through iterative improvement. This was particularly evident in the artistic scenarios where both systems were pushed to their limits. We will discuss those further below.

Considering the two NISMEs have generated similar results, with Locus being the only that has received validation in terms of its usability as part of a prior scientific study, we may infer that Monet has offered similar levels of embodiment, accuracy, and expressivity, thus validating its implementation. This inference is further supported by the overwhelmingly positive feedback captured through the user study.

6. NISMES IN ARTISTIC SCENARIOS

In addition to the study, both NISMEs were used in two artistic works. Monet was used in a work titled "Envelop" (2018) and Locus in the previously cited "Traces" [1]. "Envelop", originally written for the San Francisco Envelop venue, utilizes the same digital signal processing algorithm with dynamically changeable sounds that can be layered on top of each other together with other fixed media stems, all of which are spatialized, and some of which are in constant automated motion. The ensuing soundscape is meditative and predominantly tonal in nature without strict timing cues. Through a series of performances the system proved consistently reliable, scalable, and uniquely expressive, with the greatest challenge being the projection of the performer's embodiment towards the audience. Future iterations of the system may benefit from a larger multitouch surface coupled by a screen, or the use of depth perception cameras that may offer new gesture-based ways of interacting with the spatial mask.

Similarly, Locus has been further extended and utilized in the aforesaid Traces work that was premiered in the summer 2019. The glove was enhanced with additional markers to allow for new gestures, including thumb, index, and middle fingers on each hand. The relative dial-like hand roll algorithm received further improvements to improve its reliability and accuracy, as was the MaxMSP processing of the incoming data. Traces in many ways represents the most ambitious implementation of both the Locus and the D^4 library utilizing over 8,000 audio streams within MaxMSP that are interactively mixed down to 130 physical outputs with sub-20ms latency. The implementation is heavily threaded using the MaxMSP poly~ object and uses dynamic switching of various instances to maintain a manageable CPU overhead. In its opening, it features a real-time sonification of the Earth's atmosphere density data during a quiescent solar day, utilizing a variant of the spatial mask-based per-channel digital signal processing similar to that of Monet's implementation used in the study. A notable difference is that the incoming signal is split into two components, each being assigned a spectral subset of the original sound. By default, the low spatial mask values retain only higher frequencies of the original sound while the higher values introduce the remaining spectra. The work then transitions to a more accessible tonal musical language inspired by the data using prerecorded and minimally live-processed stems. The choice of limited live processing was to encourage listeners to shift their attention to the spatial dimension as the primary driver of the musical structure without sacrificing the accessibility of the musical language. The Locus' gestural vocabulary was expanded to include pinch with thumb and index finger, thumb and middle finger, a three-finger pinch and roll, and their permutations with roll and other existing gestures. This allowed for sounds to be grabbed and thrown around (thumb-middle finger pinch while pointing and release while moving), as well as for triggering of events and changing of sections (radio-dial-like three finger pinch and roll). Given its greater sense of audience perceived embodiment and consequently cognitive affordances, the overall complexity of the system, and most importantly strictly timed aspects of the work that require carefully choreographed execution, this NISME has resulted in greater tension and intensity from the performer's perspective, requiring extended rehearsals. In that respect Locus may have gotten closer to the proposed NIME sweet spot where the sense of challenge is in balance with embodiment, and engagement, although this distinction is undoubtedly in great part due to the stylistic differences between the two works described here. The video of the resulting 9-minute work can be found at [1].

6.1 Towards NISME Guidelines

The study and the utilization of the two NISMEs in the artistic scenarios have resulted in several emerging design guidelines future NISMEs may want to consider. In addition to the design constraints summarized in the second section, namely embodiment, expressivity, and accuracy, scalability and adaptability, collaborative capacity, and intuitive and natural interaction, NISMEs need to manifest these qualities first and foremost in respect to the spatial dimension. They need to do so while balancing the challenge with the embodiment, and engagement. NISMEs need to empower performers to make meaningful actions that clearly communicate their relationships to the spatial and other expressive dimensions (e.g. sound modulation), so that both the performer and audience can build a sense of rapport with the newfound spatialization-centric paradigm. In that respect, until audiences develop greater comfort and familiarity with the notion of spatial dimension carrying the musical structure, NISME aesthetics should be treated as an even more frail subset of NIMEs that requires an etude-like clarity. Whereas NIMEs may be seen as vehicles to facilitate the emancipation of timbre, we expect that only under the aforesaid proposed conditions we are to expect for a NISME to appropriately support the proposed paradigm shift and the anticipated emancipation of the spatial dimension.

7. CONCLUSIONS

In this paper we have proposed a new categorization of NIMEs with specific focus on the emancipation of the spatial dimension of the musical expression. Further, we have presented the design and implementation of Monet NISME that offers a unique approach to spatializing sound. Through a 22-participant user study, we assessed its design and implementation by comparing it to a second NISME (Locus) that has been designed with identical goals and design constraints, while using a fundamentally different approach. We leveraged Locus validation as a controller in a prior study and through a comparative study evaluated Monet, while also exploring the artistic expressivity of both NISMEs through free-play testing scenarios. The study data was further complemented by qualitative reflections from the performer's perspective using two artistic works specially written for the two NISMEs. The resulting findings have served as a foundation for the NISME design guidelines presented above. Informed by the user feedback we aim to continue refining and improving upon their design while also exploring their collaborative dimension.

8. ACKNOWLEDGMENTS

This work is in part supported by the National Science Foundation under Grant No. 1748667 and Virginia Tech's Institute for Creativity, Arts, and Technology (ICAT).

9. REFERENCES

[1] Traces. https://www.youtube.com/watch?v=jPBzkMkQlq4. [Last accessed: 11 January 2020].

[2] N. Barrett. Creating tangible spatial-musical images from physical performance gestures. In E. Berdahl and J. Allison, editors, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 191–194, Baton Rouge, Louisiana, USA, May 2015. Louisiana State University.

[3] C. Benson, B. Manaris, S. Stoudenmier, and T. Ward. Soundmorpheus: A myoelectric-sensor based interface for sound spatialization and shaping. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 332–337, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University.

[4] I. I. Bukvic. 3d time-based aural data representation using d4 library's layer based amplitude panning algorithm. International Community on Auditory Display, 2016.

[5] I. I. Bukvic. *Introducing D4: An Interactive 3D Audio Rapid Prototyping and Transportable Rendering Environment Using High Density Loudspeaker Arrays.* Ann Arbor, MI: Michigan Publishing, University of Michigan Library, 2016.

[6] I. I. Bukvic, G. Earle, D. Sardana, and W. Joo. Studies in spatial aural perception: establishing foundations for immersive sonification. Georgia Institute of Technology, 2019.

[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] L. Engeln, D. Kammer, L. Brandt, and R. Groh. Multi-touch enhanced visual audio-morphing. In T. M. Luke Dahl, Douglas Bowman, editor, *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 152–155, Blacksburg, Virginia, USA, June 2018. Virginia Tech.

[9] J. Halpern. Introduction to unity. In *Developing 2D Games with Unity*, pages 13–30. Springer, 2019.

[10] B. Johnson, M. Norris, and A. Kapur. speaker.motion: A mechatronic loudspeaker system for live spatialisation. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 16 of *2220-4806*, pages 41–45, Brisbane, Australia, 2016. Queensland Conservatorium Griffith University.

[11] K. Lui-Delange, S. Distler, and R. Paroli. Sensel morph: Product communication improvement initiative.

[12] 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*, 40(4):62–78, 2016.

[13] G. Paine. Towards unified design guidelines for new interfaces for musical expression. *Organised Sound*, 14(2):142–155, 8 2009.

[14] M. Puckette. Max at seventeen. *Computer Music Journal*, 26(4):31–43, 2002.

[15] 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, June 2019. Federal University of Rio Grande do Su.

[16] B. Smith and N. Anderson. Arraynger: New interface for interactive 360 spatialization. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pages 291–295, Copenhagen, Denmark, 2017. Aalborg University Copenhagen.

[17] S. Williams. Osaka expo '70: The promise and reality of a spherical sound stage. In *Insonic*, November 2015.