Gestural Interactions with Object-Based Audio in an Internet of Sounds Ecosystem

1st Kurt Mikolajczyk
Creativity and Cognition Studios
University of Technolgy Sydney
Sydney Australia
kurt.mikolajczyk@uts.edu.au

2nd Sam Trolland
SensiLab
Monash University
Melbourne, Australia
sam.trolland@monash.edu

3rd Alon Ilsar
SensiLab
Monash University
Melbourne, Australia
alon.ilsar@monash.edu

4th Jon McCormack
SensiLab
Monash University
Melbourne, Australia
jon.mccormack@monash.edu

5th Sam Ferguson
Creativity and Cognition Studios
University of Technology Sydney
Sydney, Australia
samuel.ferguson@uts.edu.au

6th Oliver Bown
Interactive Media Lab
University of New South Wales
Sydney, Australia
o.bown@unsw.edu.au

***Abstract*—An Internet of Sounds (IoS) ecosystem for rendering object-based audio across wireless distributed devices powering lights and speakers is interfaced with wireless gestural controllers to support embodied interaction in audio spatialisation. The devices generate audio locally, supporting the design of abstract objects for affecting other media in a distributed system. Through contextualising ours with other gestural controllers interacting in spatial audio settings, we articulate how affordances of the IoS ecosystem and its components complement the embodied interaction that gestural control facilitates with object-based audio. The interactions we describe for controlling are imagined in two creative scenarios that we discuss in an envisioned application and future developments.**

***Index Terms*—Internet of Sounds, gestural interaction, object-based audio, interaction design**

I. Introduction

The motivation for our research lies in the observation that embodied interaction for audio spatialisation is underutilised in existing interfaces. Our vision is to explore the creative possibilities enabled through embodied interaction with object-based audio and Internet of Things (IoT) devices by using a wearable gestural controller to explore a spatial audio system. Leveraging the affordances of IoT technology through wireless communication and an edge-based approach to distributed audio generation, we map gestural actions to spatialised audio. The gestural controller and spatial audio system communicate wirelessly, supporting exploration of the sonic landscape from multiple listening perspectives. Generating audio on the device in our edge-based approach supports the design of abstract object-based audio effects we call *Manipulator Objects*.

Object-based audio is a speaker-agnostic approach to creating spatialised audio that abstracts the sounds as a region in space, which is then considered an object (Figure 1). Positional metadata attached to the objects is used to place a sound in the spatial audio environment [1]. Gestures can be described as a "motion of the body that contains information" [2], with gestural controllers being devices that capture this movement data and make it available in order to affect changes in a broader system. In this case the gestural movements are used as an interface to loudspeakers and LED lights distributed in a performance space controlled by a wireless network of ten Raspberry Pi computers shown in Figure 2.

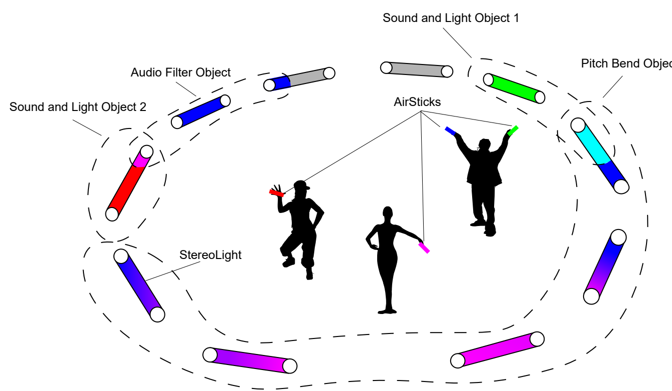

Fig. 1. Spatial Objects moved across networked devices (StereoLights) using gestural controllers (AirSticks).

The background review contextualises a description of our IoS ecosystem components in the literature, highlighting their novel aspects. We extend object-based audio concepts to different kinds of media such as light, sound, as well as effects and transformations. This is highlighted in Figure 1, which shows two objects of sound and light and two objects with effects: an Audio Filter Object and a Pitch Bend Object. The Audio Filter Object and Pitch Bend Object do not generate any audio, but instead apply transformations to overlapping objects. The Audio Filter Object applies a low pass filter to overlapping sound from Sound and Light Object 2. The Pitch Bend Object shifts the frequency of overlapping sound from Sound and Light Object 1. The amount of bend is mapped to the Pitch Bend Object's colour gradient, meaning if Sound and Light Object 1 were to proceed clockwise, the sound would

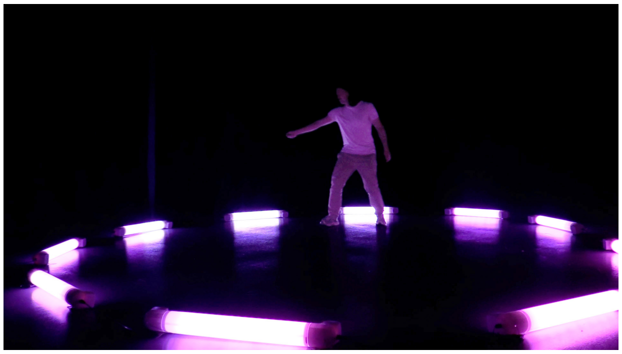

Fig. 2. Controlling Spatial Objects with an *AirStick*, where the position of sound is mapped to the orientation of the controller.

Fig. 3. Multiple users controlling Spatial Objects on the line configuration.

move to a lower frequency before returning to its original pitch.

With our system we evaluate gestural interaction as an interface for object-based audio in the context of two creative scenarios: *composition* and *performance*. The *composition* scenario is defined through the act of *creating the system*, using the gestural interface to adjust parameters and define system constraints. The *performance* scenario is defined through the act of *creating with the system*, using the gestural interface to explore established parameters in real-time. We explore gestural interaction with object-based audio in each of these creative scenarios independently, highlighting some real-world examples in which these interactions could be employed. Designing a spatial audio mix for a film using a unique audio format, or planning a sequence from a triggered event in an interactive artwork, are examples of the *composition* scenario. An improvised dance with the gestural controller, or real-time interactions with a public art installation, would be considered in the *performance* scenario. The affordances of gestural interaction within each scenario are further explored in our discussion.

II. Background

This paper combines the authors' prior work in gestural interaction and distributed multimedia together in a cohesive IoS ecosystem, in which one or more people can manipulate a network of wireless media devices through the movement of wireless gestural controllers as shown in Figure 2 and 3. An IoS ecosystem is a structure of distributed, networked, sound making devices formed from connections of IoT technologies [3]. Object-based audio is used as a platform to bring the inherently spatialised gestural interaction and distributed media together into a IoS ecosystem that is grounded in a shared understanding of position and space.

A. Object-based audio

Object-based audio is an approach to spatial audio where a sound's position is represented as a region in space using metadata attached to each sound [4]. It stands in contrast to channel-based approaches for spatial audio, where the position of a sound is understood in relation to the speaker through which it is realised. For designing on distributed speakers, object-based audio is a way of abstracting speakers into the space they reside in. Activities in this approach allow the moving and manipulating of an object's position and size in relation to the room or environment in which it is realised.

Our object-based audio system does not implement the full extent of techniques for acoustic spatialisation used elsewhere such as [5] or [6]. Instead, our system uses vector-based-amplitude-panning (VBAP) to use the distance to the speaker as the way the Spatial Object is heard across a multi-speaker array. This simplified approach, similar to [7], does not employ filtering or capture metadata using room impulse recordings (RIRs) to position the sound in the most accurate way possible. In our system, VBAP mapped to a speaker's volume conveys a sound moving around the environment, though this is not an identical auditory experience to a physical loudspeaker moving itself [8]. However, for someone interacting with the system using a gestural controller (hereafter an *interactor*), our approach opens the creative design space to move, animate, and manipulate sound (and light) as an object in space in a perceivably believable spatial audio experience.

1) Taxonomy of Spatial Objects: We define "Spatial Objects" as the generic base for object-based media in our IoS ecosystem. As shown in Figure 4, Spatial Objects can be extended to represent sound, light, multimedia, or regions of space that can affect (manipulate) other Spatial Objects. We imagine Spatial Objects to encompass not only media like a sound, but changes to audio information using digital signal processing like filtering or pitch shifting, as well as changes to musical information such as activating rhythmic subdivisions or transforming harmonic content.

Spatial Objects in our IoS ecosystem are first defined by their position. Then, depending on what properties they are assigned, they can be considered a Sound Object, Light Object, Multimedia Object, or Manipulator Object. A Sound Object is a region of sound that is playing back an audio file or generating a synthesised tone. A Light Object represents a region of colour which, when combined with a sound, can be considered a Multimedia Object. A Manipulator Object represents a parameter that can affect other Spatial Objects. An example of this is shown in Figure 1, where a Manipula-

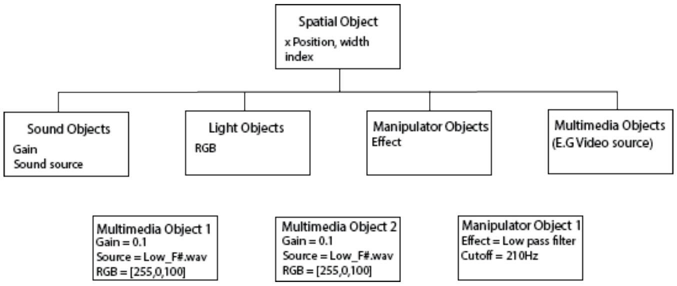

Fig. 4. An inheritance model of Spatial Objects.

tor Object mapped to a high pass filter overlaps a portion of Sound Object 1's region. In the overlapping region the audio generated by Sound Object 1 will be filtered by the Manipulator Object, but the rest of it's region remains with no tonal manipulation.

Manipulator Objects could extend to parameters to affect other types of media, for example, applying a strobe or pulsing animation to alter other Light Objects. For sound, Arpeggiator Objects, Reverb Objects, Delay Objects, are possible ways of parametising the sonic (and other media) manipulation. These examples of Manipulator Objects are all possible instantiations of Spatial Objects that can be positioned, combined, and overlapped in our IoS ecosystem. In this paper we will use the generalised term 'Spatial Objects' when referring to instances of our objects, noting that this term encompasses all of the examples described above.

B. Gestural controllers

Gestural controllers measure data about their movement in 3D space, and translate this information in order to interface with another system. Their role in spatial audio contexts is extensive. Marshall [9] has compared audio spatialisation approaches relating potential gestural interactions to musical applications. Elsewhere a comparison of touch interfaces and gestural control has been explored [10], where they demonstrate an interactor's ability to localise the position of a sound is improved with a touch interface but the gestural interactions were easier and more comfortable. Gestural controllers have also been evaluated as tools in musical performance [11] and for supporting composition [12]. In the IoS community gestures have been used as interface to 'air instruments'. These have been developed using computer vision [13]–[15], haptic [16], and other sensor technologies [17] [18] to capture movements and map the data to realise imagined musical instruments like air guitar. Gestural controllers have also been used as devices in telematic musical performance [19] as well as probes for understanding gesture in performance more broadly [20].

1) AirSticks: Wireless Gestural Interface: The *AirSticks* [21] are the gestural controller used in this IoS ecosystem. The *AirSticks* are a wireless gestural instrument that translates participant's movements into discrete events and continuous control information in real-time. A 9-DOF IMU[1] (fusing data from an accelerometer, gyroscope, and magnetometer) is used to measure the device's 3D orientation and linear-acceleration. This data is sent wirelessly over Bluetooth Low Energy[2] to a custom application running on a nearby computer. The application interprets the participant's gestural actions to provide continuous [22] and discrete control events [23] that drive changes in the IoS ecosystem.

With the information from the *AirStick* we are able to detect the 3D orientation of the device to determine the direction in which it is pointing at any given moment. Additional information is gathered by aggregating movement data over a period of time to identify actions performed with the device. For example a short, sharp, gesture identified by a sudden change in linear acceleration is referred to as a 'strike', whereas consistent movement over a period of time followed by a moment of stillness is associated with the metaphor of 'building up energy' and 'release'.

C. Object-based audio within an IoS ecosystem

Within our IoS ecosystem, object-based audio provides a bridge linking the inherently spatialised interactions of both the distributed media and *AirStick*. An interactor wielding an *AirStick* can reposition themselves to hear and see from different perspectives. Likewise the distributed media can be picked up and moved to make and explore new spatial audio configurations. This is enabled by the wireless communication in our IoS ecosystem and supports embodied interaction by allowing control of the Spatial Objects while moving through the space and listening.

Communication and messaging systems for networking components is an important research area that IoS ecosystems such as ours leverage. Previous research has been directed at developing systems generalisable to a variety of interactive scenarios, which has, in addition to the contribution of reusable networking tools, also worked to identify commonly required utilities in IoS ecosystems such as device discovery and clock syncrhonisation. *Soundworks* [24] is web-based approach, while *O2* [25] and *LibMapper* [26] offer these features to extend and simplify the flexible Open Sound Control (OSC[3]). *MUSEPA* [27] similarly describes a communication through a semantic web of things architecture for networking heterogeneous distributed devices.

The loudspeakers and LEDs in our IoS ecoystem operate in a wireless 'edge' setting [28], generating the audio (and light) locally on the Raspberry Pi Zero in an IoT approach to spatial audio. Generating the audio locally on the devices massively reduces the network data requirements in wireless settings by sending instructions to the Raspberry Pis to make sound and light rather than actual raw audio data. The approach is also scalable as more outputs can be added to the network without requiring more channels from a central controller [29]. This combination of affordances makes available the

[1]https://www.bosch-sensortec.com/products/smart-sensors/bno055/
[2]https://www.bluetooth.com/specifications/specs/core-specification/
[3]https://opensoundcontrol.stanford.edu/index.html

design of object-based audio in wireless settings as it takes out the complexities of coordinating wireless networked audio data and reduces the problem to coordinating sets of simpler musical instructions.

1) StereoLights: Sound and/or Light Things in a Media Multiplicity: The *AirSticks* interfaced a system of distributed wireless light and sound devices. LEDs and speakers are contained in cylinders of light that have speakers on the end. The cylinders, dubbed a *StereoLight*, are driven by a Raspberry Pi Zero W that generates audio locally on the device. The *StereoLights* are an iteration of previous distributed IoT systems [30] that use homogeneous arrays of media to act in coordinated and collective ways in real-time contexts [31]. These kinds of works such as our arrangement of *StereoLights* we refer to as a *Media Multiplicity* [32]. Other multiplicitous media works include assemblages of robotic roller skates [33] arrays of speakers in projection rooms [34].

III. Method

We used a Research Through Design method to explore the design space of our IoS ecosystem through an iterative process of prototyping, reflection and evaluation [35]. The research aim was to explore embodied interactions using a gestural controller with an object-based multimedia system in an edge-based distributed audio environment. To understand the interactions and bring our insights forward into the community we describe our system as an artifact that demonstrates our approach and understanding of this unique interaction scenario [36].

We designed a prototype system over three days in a lab space, then moved to a large theater space where we explored our system and developed interactions across two more days. We documented the process through sound and video recording, with the first author logging daily activities and keeping a version history of the code development and sound design.

Our approach to present the interactions in relation to the creative scenarios is inspired by the framing done in the Embodicomp system [12], Marshall's [9], and Cohen [37]. Their work describes their designed gestural interactions in the context of a 'band metaphor' for composers in the EmobdiComp system, and Spatial Performers, Instrument Performers, and Spatial Conductors in Marshall. Cohen's work instead focuses on listening environments, such as a theater or cocktail bar. The rationale behind us framing the interactions as *composition* and *performance* is we believe they capture parts of the creative process across diverse artistic contexts, not only music.

A. Materials

Our IoS ecosystem is comprised of the devices and communication protocols shown in Figure 5 and detailed below.

1) Sound and Light Rendering: The *StereoLight's* are a distributed multimedia system controlled in real-time with *HappyBrackets*[4] using OSC over WiFi [38]. The Raspberry Pis inside the *StereoLights* run a program in parallel to control each of the speakers and lights connected to it. HappyBrackets is used to maintain synchrony across the ten distributed *StereoLights*.

HappyBrackets, when used to manage multiple distributed outputs, has an abstract base class called a 'Renderer' that represents one instance of a light or speaker. The abstract base class in this case was built to manage the Spatial Objects. The Renderer program and HappyBrackets run using the *Beads*[5] audio library to generate audio on the device, using a custom Raspberry Pi hat developed by BitScope Australia[6]. The hat runs a LED driver splitting eight channels of data signals for the lights and a 3W amplifier to drive two 4 ohm Dayton Audio 2.5 inch loudspeakers.

2) Gestural Interaction: The central point of our system is a computer running *AirWare*, a custom built application that receives and analyses wireless data from one or more *AirStick* devices. *AirWare* processes the orientation and linear acceleration data from *AirStick* devices, identifying gestural features from the continuous movement of the devices that are used to manipulate sound objects. Both discrete and continuous gestural events are identified, which are made available over OSC. Discrete control is provided through precisely-timed, momentary, events that can signify an action or change to the system. Examples include triggering a sample or playing a note, and can include additional information including note velocity, length or pitch. Continuous control is provided through a constant stream of data that represents the current state of one or more dimensions of the *AirStick's* movement. This includes raw elements of the *AirStick's* movement such as 3D orientation and linear-acceleration, as well as higher-level gestural elements such as 'AirStick Energy' [21] calculated through accumulation of movement over a given period of time.

An internal loopback to pass OSC from *AirWare* to Cycling 74's MaxMSP[7] was used to bring the incoming data into a graphical programming environment, where MaxMSP became the software glue between our two IoT technologies. Alternatives to Max were considered, with Processing[8] narrowly missing out on account of Max's integration with Ableton Live[9] and our desire to use Ableton's interface to cue and sequence events. We also nominally considered alternatives such as Chataigne[10], PureData[11], TouchDesigner[12], but Max and Ableton were preferred as a previously used tool for sequencing and timing events and reusing components from older systems.

[4]https://www.happybrackets.net/
[5]http://www.beadsproject.net/
[6]http://www.bitscope.com
[7]https://cycling74.com/
[8]https://processing.org/
[9]https://www.ableton.com/en/live/max-for-live/
[10]https://benjamin.kuperberg.fr/chataigne/en
[11]https://puredata.info/
[12]https://derivative.ca/

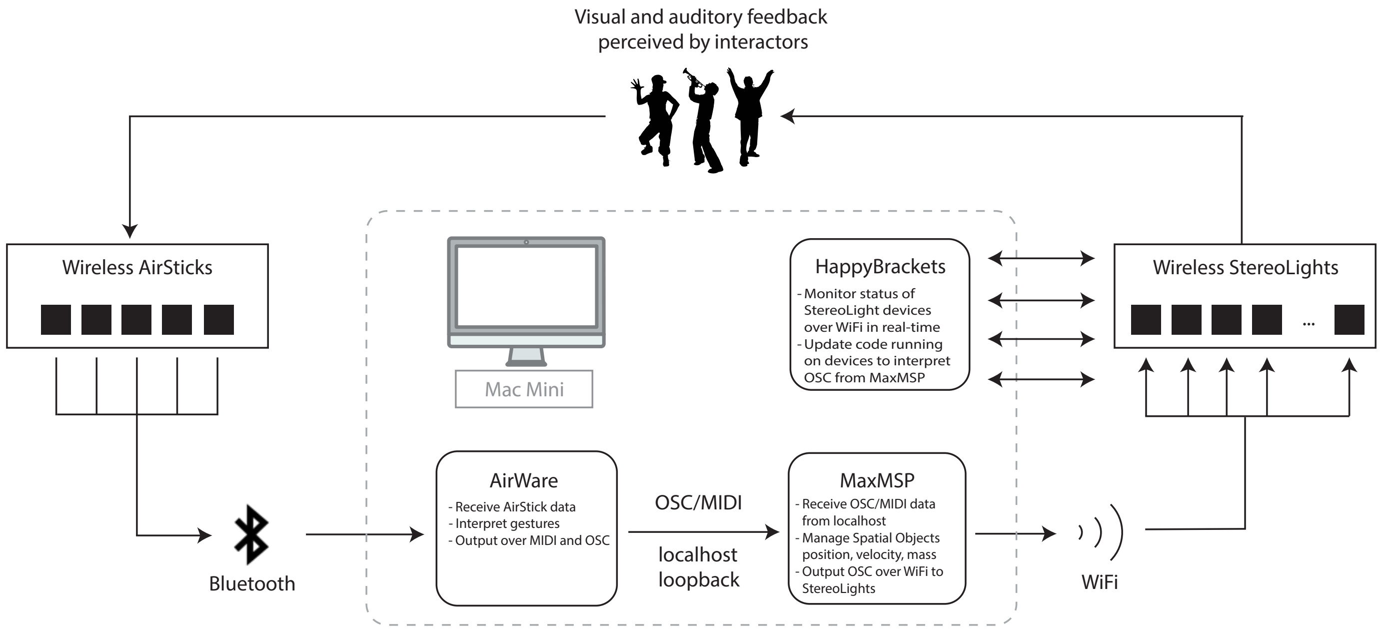

Fig. 5. Block diagram of communication in the IoS ecosystem.

Finally, the Mac Mini also ran IntelliJ IDE[13] with the HappyBrackets plugin installed to modify the code running on the *StereoLights*. Fitting with our Research Through Design approach, the rapid prototyping capability of this system meant we were changing the available parameters at different points in our system in a matter of seconds.

B. Process

1) Prototype System: Figure 6 shows the initial prototype system developed in our lab space. This preliminary design established the communications from Figure 5 and developed a basic object-based media system using code from previous HappyBrackets projects.

Fig. 6. Prototype IoS Ecosystem in our lab space

[13]https://www.jetbrains.com/idea/

The gestures used were continuous data mappings of orientation to position a Spatial Object, as well as having discrete events for triggering sounds as notes. We also explored accumulated data strategies through AirWare, using the metaphor of continuously shaking the *AirStick* to build up ‘mass’, followed by a moment of stillness in which the Spatial Object is ‘released’ to traverse the installation space with a velocity modified by the mass value. In this prototype we experimented with the mappings, for example inverting the relationship of mass to velocity.

2) Sound design and Spatial Object code for the StereoLight: The object-based approach required creating sounds to be represented as Spatial Objects, a process of objectification described in [4]. For designing the sounds, a 17-voice harmony played on piano was sampled playing through a 3 chord harmonic progression. The notes of the chord were divided into three pitch ranges to became the sounds tied to the Spatial Objects. Effects processing such as reverb and delays gave an ethereal quality to the recordings that was then rendered as two channel versions, one for each side of the *StereoLight*. This is the same as stereo recording but without any implied direction, the two channels are just considered ‘a’ and ‘b’. It creates subtle phasing artifacts as the Spatial Objects move, a similar approach to Cohen’s *filtears* [37] to help identify, but also give character to, the Spatial Objects’ animation.

Overall the sound design gave the Spatial Objects a sustained, ambient sound, encouraging exploration of their harmonic blending as they gradually overlap. It was also organised in a way so the Spatial Objects could remain on the same part of the chord progression, change chords automatically, or change from a discrete gestural event. The sound design for the Spatial Objects was happening while the code for the object-based audio system in HappyBrackets was being written. This was a back and forth between assembling samples in a Digital Audio Workstation and hearing them on a single *StereoLight*.

3) Gestural interactions with Spatial Objects: The mapping of detected gestures from the *AirSticks* to sonic and light output on the *StereoLights* was modified in an iterative process in an open performance space. We made two spatial configurations of the *StereoLights* to explore interactions on: a straight line and a circle shown in Figures 7 and 8. Topologically, the key difference between them is their continuity: In the circle configuration, the Spatial Objects wrap around the circle shape like a planet orbiting the center, while the line has a boundary at each end.

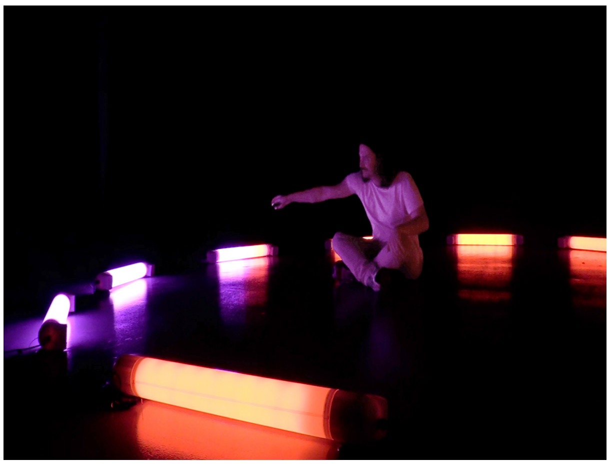

Fig. 7. The circle spatial configuration

The identification of discrete and continuous gestural events from the *AirSticks* and their mappings was used to control parameters of the Spatial Objects. This iteratively built up capabilities and features, moving from programming new interactions on the central computer to experimenting, playing, and evaluating the designed interactions on the system.

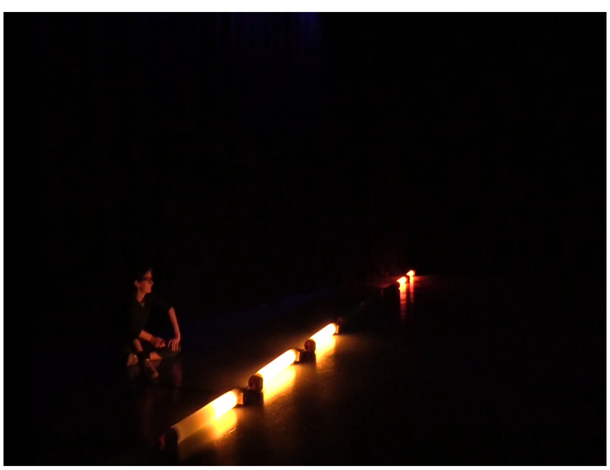

Fig. 8. The line spatial configuration

IV. Findings

We explored the IoS ecosystem in both spatial configurations, iteratively designing and experiencing the following interactions:

- Direct Control: Move and scale the Spatial Objects with the *AirStick's* orientation.
- Indirect Control: Accumulate movement data over time to influence the Spatial Object's 'mass'.
- Blend and combine Spatial Objects with one another.
- Dial-in precise values for Manipulator Objects.
- Distribute *AirSticks* across multiple participants.
- Give multiple participants multiple *AirSticks* each.

The interactions are described under two general creative scenarios that we envision the gestural controller supporting. *Composition* here refers to the user with gestural control *creating the system*, designing parameters, sequencing events, planning, exploring, and evaluating aesthetic decisions in the assembly of a work. *Performance* refers to the user with gestural control *creating with the system* evaluating in real-time, where interaction becomes like a dialogue with the system.

1) Creative Scenario 1 - Composition: The *composition* scenario imagines the user designing the system for future interactions. The wireless nature of the interface supported this by allowing the interactor to reposition themselves in order to listen and observe from multiple perspectives. This frees the interactor to move throughout the installation space and experience how the acoustics are perceived at different physical locations. This is because the *AirStick's* orientation maintains the Spatial Object's position as the designer moves about the installation space. In this way the designer can evaluate the system while imagining themselves as a future interactor with the system.

The *AirStick*, when worn on the forearm, allows for fine-grain control of sound-design parameters within the Spatial Objects. This was evaluated by mapping the roll of the wrist to the frequency cutoff of a low-pass filter Manipulator Object. Turning the *AirStick* on the wrist's axis gently rolled the filter with precision comparable to a typical dial feel. This style of control supports real-time parameter adjustments to explore system constraints while removing the interface and bringing attention to the listening and sensory information as it would be perceived in the spatial audio environment.

Lastly, the gestural interaction provided a way to rehearse sequenced events. This was evaluated in a multi-user experience in which two *AirSticks* were mapped to Spatial Objects that automatically moved through the harmonic progression described earlier. The resolution of the harmonic sequence became timed, where the interactor's movements sought to synchronise the overlap of Spatial Objects with the harmonic resolution. The harmonic progression could be retriggered, allowing us to rehearse and practice the timing of our gestures.

2) Creative Scenario 2: Performance: The performance scenarios envisioned users working as interactive artists *creating with the system*. For example, an interactive multimedia dance performance or an immersive art installation responding to a musician's gestures while playing. These interactions were similarly explored in the two spatial configurations.

The continuous data from the *AirStick* mapped to the position of the Spatial Object had a low latency. This interaction gave the sonic experience of moving a recorded sound in real-time. Rhythms could be made with gestures in timed movements. Chord changes could be triggered through discrete controls of one or more *AirSticks*, or through automation, and suggested the Spatial Object's animations could be indicative of the tension and release in other aesthetic parts of the system. This was also the case controlling Spatial Objects with multiple *AirSticks* and working collaboratively with other interactors.

Discrete events were used to set things in motion or trigger notes which led to rhythmic coordination with other Spatial Objects and the interactors controlling them. This supports the *performance* scenario because the immediate feedback enables a dialogue-like communication between the interactors and their Spatial Objects. We noticed that being able to see the entire installation space as in the line configuration made it easier to set Spatial Objects in motion with discrete events. This was evaluated through controlling the release timing and velocity of the Spatial Objects, rather than the position. Timing discrete events naturally became easier when we were visible of each other, which we were not always in the circle configuration.

Lastly, in the line configuration, the participation between collaborators with Manipulator Objects supported a real-time interactive experience representative of a performance scenario. With three participants and three *AirSticks*, different roles were established between those controlling Spatial Objects with sounds tied to them, and those controlling Manipulator Objects. We used a Manipulator Object tied to Pitch Bend, so Spatial Objects overlapping would be bent harmonically. The interaction between the Spatial Objects and the participants supported a performance scenario by creating a kind of ensemble between the interactors who were responding and reacting to each other and the system in the moment.

V. Discussion

The interactions provided by the gestural controller are supported by the IoT technology to give meaningful control over the spatial audio system. They allow for embodied interaction for working with Spatial Objects. Not only in the obvious mapping from movements of the body to movements of the Spatial Objects, but also in the way the gestural controller allows an interactor to be immersed in the spatial audio environment. They can embody the future listener to make sense of the spatial audio and move as the listener might through the space, while adjusting the parameters of the system, and while observing the actual system and not an interface representing it. The alternative is to use a graphical interface such as an Ipad to mix spatial audio wirelessly. Similar, except the user continually refers to some graphical or hardware interface. Why is an alternative required?

A. Gestures for object based media

Gestural controllers are not a complete alternative. Instead, gestural interfaces for object-based media can be a supportive tool for developing spatial audio works and performances. Within both scenarios users can explore the animation and movement of sound in a spatial audio environment. This activity in relation to object-based audio is already a highly abstract task, and while there are plenty of objective assessments that can be made about balancing a spatial audio mix, the rules and possibilities, that we also think of now in relation to Manipulator Objects, are not written. Gestures as interface, supported by wireless communication in our IoS ecosystem, allows the user to move and re-configure the spatial audio environment. In doing so they are able to imagine themself a future listener, embodying them in the spatial audio environment. We think this is clearly a useful perspective to have in creative projects requiring design for spatialised media.

However, we note that the development of new features and interactions within our IoS ecosystem depended on the chain of tools running together on the central controller computer in Figure 5. Building up the features in each interaction required us to enter various problem solving modes, for example, determining how the objects would wrap around the structures. These activities, as for many in sound design, are supported by graphical representations of systems that are useful for allowing a full view of the system, numerical representations of what is happening, and timelines for organising and assembling ideas into larger works possible. So while the gestural controllers gave a natural and intuitive control over the parameters of Spatial Objects, the interaction design for developing systems for full creative works required additional tools that were not controlled through gesture.

B. Envisioned applications

The two described scenarios can apply to diverse creative contexts, but a commonality of the interactor in both is they are using the gestural controller to help them make sense of the system. Being able to make sense of the system allows the interactor to make meaningful judgements about what to do next (in response to feedback from the system or another interactor). In addition to the fact that a performance implies an audience is observing the system with the performer, the main difference for making sense of the system through interaction in the two scenarios is temporal. In the performance scenario the system is in dialogue with the interactors in a constant back and forth. In the composition scenario observations happen over longer periods of time: a sequence is triggered, it unfolds, and the interactor evaluates and then decides what needs to change.

The two scenarios, and the sense-making that happens in them, are not a binary for how we see our system being brought forward into an application. Instead, the scenarios represent envisioned ways of working and exploring the potentials of object-based media. The immediate feedback of real-time interaction would provide a way for an interactor to quickly explore how different Manipulator Objects could work together. When exploring a largely unknown creative space such our Spatial Objects system, our envisioned application using gestural controllers would allow embodied interactions for both moving and manipulating Spatial Objects while observing and in dialogue with the system, as well as setting objects in motion to make sense of something unfolding over a period of time.

VI. Future

Our system could be brought to a more realistic and believable spatial audio experience through improving the design of the algorithms that manipulate the Spatial Objects to position them. As discussed in Marshall [9], different spatial audio approaches to position the sound in the environment can afford different interactions and sonic possibilities. This is bounded of course by constraints from the computational limits of small single board devices like the Raspberry Pi Zero, but there are undoubtedly optimisations to grant more flexibility in the system.

Our envisioned application could consider the IoS ecosystem as research probe into creative practice for technologies within the IoS community. This research would support the development of models of creative practice with emerging technologies. Furthermore, building and understanding the needs of the artists in this domain would help developers and engineers understand the needs of creative applications and the technology research directions to take.

What would happen if a Spatial Object's positional meta data was substituted with something that was not positional at all? What if the metadata only contained a device's *temporal* information, or perhaps only an arbitrary identifier like a *name*. It is also conceivable to base the metadata on abstract information that might make use of sensory inputs, for example a weather system. While an abstract thought for now, this and the other creative possibilities of working in novel ways with distributed sound are places we want to take our system.

VII. Conclusion

The nexus of our IoT technologies has resulted in an object-based media system running on a distributed audio system controlled by a gestural interface. We developed new concepts for combining Spatial Objects with different media and effects (Manipulator Objects) that we can flexibly explore using the edge-based audio. Our understanding of the interactions for the performance and composition scenarios describes an envisioned future application of how we would bring new creative works to the IoS ecosystem and use gestures to explore object-based media. As such embodied interaction using gestures to control Spatial Objects is, in our IoS ecosystem, supported by its underlying IoT technology, allowing interactors to make sense of a re-configurable spatial audio environment from multiple perspectives.

VIII. Acknowledgements

Sincerest thanks to Caitlin Dubler and Vanessa Gawan-Taylor for bringing their expertise and insight to the design of the *StereoLights* and for all their assistance experimenting, deploying, designing, and exploring our interactive system.

Funding for this research was received from the Australia Research Council: ARC Linkage LP180100151, Artistically Rethinking Creative Coding.

References

[1] F. Rumsey, "Spatial audio Channels, objects, or ambisonics?" *AES: Journal of the Audio Engineering Society*, vol. 66, no. 11, 2018.

[2] G. Kurtenbach and E. A. Hulteen, "The art of human-computer interface design," *B. Laurel (ed.), Gestures in Human-Computer Communication*, pp. 309–317, 1990.

[3] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Ostergaard, F. Font, T. Backstrom, and C. Fischione, "The Internet of Sounds: Convergent Trends, Insights and Future Directions," *IEEE Internet of Things Journal*, 2023.

[4] P. Coleman, A. Franck, J. Francombe, Q. Liu, T. De Campos, R. J. Hughes, D. Menzies, M. F. Galvez, Y. Tang, J. Woodcock, P. J. Jackson, F. Melchior, C. Pike, F. M. Fazi, T. J. Cox, and A. Hilton, "An Audio-Visual System for Object-Based Audio: From Recording to Listening," *IEEE Transactions on Multimedia*, vol. 20, no. 8, 2018.

[5] R. Hoy and D. Van Nort, "A technological and methodological ecosystem for dynamic virtual acoustics in telematic performance contexts," in *Proceedings of the 16th International Audio Mostly Conference*, 2021, pp. 169–174.

[6] J. Cadavid, M. B. Møller, S. Bech, T. Van Waterschoot, and J. Østergaard, "Performance of Low Frequency Sound Zones Based on Truncated Room Impulse Responses," in *ACM International Conference Proceeding Series*, 2022.

[7] M. Gurevich, D. Donohoe, and S. Bertet, "Ambisonic Spaitialization for Networked Music Performance," *International Conference on Auditory Display (ICAD) 2011*, 2011.

[8] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst, "Auditory distance perception in humans: A summary of past and present research," *Acta Acustica united with Acustica*, vol. 91, no. 3, 2005.

[9] M. T. Marshall, J. Malloch, and M. M. Wanderley, "Gesture Control of Sound Spatialization for Live Musical Performance," *International Gesture Workshop*, 2007.

[10] G. Marentakis and S. A. Brewster, "A study on gestural interaction with a 3D audio display," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3160, 2004.

[11] B. D. Donato and J. Bullock, "gSpat: Live sound spatialisation using gestural control," *International Conference on Auditory Display*, 5 2020.

[12] D. El-Shimy, S. Cowan, and J. R. Cooperstock, "Embodicomp: Embodied interaction for mixing and composition," in *ICMC*, 2014.

[13] D. Cocchiara and L. Turchet, "Democratizing access to collaborative music making over the network using air instruments," in *ACM International Conference Proceeding Series*, 2022.

[14] C. T. Tolentino, A. Uy, and P. Naval, "Air Drums: Playing Drums Using Computer Vision," in *2019 International Symposium on Multimedia and Communication Technology, ISMAC 2019*, 2019.

[15] X. Fan, H. Ave, A. Arbor, G. Essl, H. Ave, and A. Arbor, "Air Violin : A Body-centric Style Musical Instrument," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2013.

[16] I. Hwang, H. Son, and J. R. Kim, "AirPiano: Enhancing music playing experience in virtual reality with mid-air haptic feedback," in *2017 IEEE World Haptics Conference, WHC 2017*, 2017.

[17] M. Karjalainen, T. Mäki-Patola, A. Kanerva, and A. Huovilainen, "Virtual air guitar," *AES: Journal of the Audio Engineering Society*, vol. 54, no. 10, 2006.

[18] J. E. Tamani, J. C. B. Cruz, J. R. Cruzada, J. Valenzuela, K. G. Chan, and J. A. Deja, "Building guitar strum models for an interactive air guitar prototype," in *ACM International Conference Proceeding Series*, vol. Part F137694, 2018.

[19] S. D. Thorn, "Telematic wearable music: Remote ensembles and inclusive embodied education," in *ACM International Conference Proceeding Series*, 2021.

[20] R. I. Godøy, E. Haga, and A. R. Jensenius, "Playing "air instruments": Mimicry of sound-producing gestures by novices and experts," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3881 LNAI, 2006.

[21] S. Trolland, A. Ilsar, C. Frame, J. McCormac, and E. Wilson, "AirSticks 2.0: Instrument Design for Expressive Gestural Interaction," in *Proceedings of the International Conference on New Interfaces for Musical Expression*, 2022.

[22] J. Rovan and V. Hayward, "Typology of Tactile Sounds and their Synthesis in Gesture-Driven Computer Music Performance," *Trends in Gestural Control of Music*, 2000.

[23] S. Holland, T. Mudd, K. Wilkie-McKenna, A. McPherson, and M. M. Wanderley, "Understanding Music Interaction, and Why It Matters," in *Springer Series on Cultural Computing*, 2019.

[24] B. Matuszewski, "A web-based framework for distributed music system research and creation," *AES: Journal of the Audio Engineering Society*, vol. 68, 2020.

[25] R. B. Dannenberg, "Communication for real-time music systems: An overview of o2," *Computer Music Journal*, vol. 45, 2023.

[26] J. Malloch, S. Sinclair, and M. M. Wanderley, "Libmapper: (a library for connecting things)," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '13. New York, NY, USA: Association for Computing Machinery, 2013, p. 3087–3090. [Online]. Available: https://doi-org.ezproxy.lib.uts.edu.au/10.1145/2468356.2479617

[27] L. Turchet and F. Antoniazzi, "Semantic web of musical things: Achieving interoperability in the internet of musical things," *Journal of Web Semantics*, vol. 75, 1 2023.

[28] J. Lin, W. Yu, N. Zhang, X. Yang, H. Zhang, and W. Zhao, "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications," *IEEE Internet of Things Journal*, vol. 4, no. 5, 2017.

[29] S. Ferguson, A. Rowe, O. Bown, L. Birtles, and C. Bennewith, "Networked pixels: Strategies for building visual and auditory images with distributed independent devices," *C and C 2017 - Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, pp. 299–308, 2017.

[30] K. Mikolajczyk, O. Bown, and S. Ferguson, "A Study of Creative Development with an IoT-Based Audiovisual System: Creative Strategies and Impacts for System Design," in *Proceedings of the 15th Conference on Creativity and Cognition*, ser. C&C '23. New York, NY, USA: Association for Computing Machinery, 2023, pp. 139–149. [Online]. Available: https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3591196.3593057

[31] O. Bown, A. Fraietta, L. Loke, and S. Ferguson, "Creative Coding and Interaction Design for Media Multiplicities: Challenges, Paradigms and Frameworks," in *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 877–880. [Online]. Available: https://doi-org.ezproxy.lib.uts.edu.au/10.1145/3374920.3374966

[32] O. Bown and S. Ferguson, "Understanding media multiplicities," *Entertainment Computing*, vol. 25, no. October 2017, pp. 62–70, 2018. [Online]. Available: https://doi.org/10.1016/j.entcom.2017.11.001

[33] W. Marynowsky, S. Ferguson, A. Fraietta, and O. Bown, "'the ghosts of roller disco', a choreographed, interactive performance for robotic roller skates," in *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*, ser. TEI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 631–637. [Online]. Available: https://doi.org/10.1145/3374920.3375284

[34] "David Hockney: Bigger & Closer (not smaller & further away)," accessed: 2023-31-08. [Online]. Available: https://lightroom.uk/whats-on/david-hockney/

[35] J. Zimmerman, J. Forlizzi, and S. Evenson, "Research through design as a method for interaction design research in HCI," in *Conference on Human Factors in Computing Systems*. ACM, 2007, pp. 493–502.

[36] B. Dixon, "Experiments in experience: towards an alignment of research through design and john dewey's pragmatism," 2019.

[37] M. Cohen and L. F. Ludwig, "Multidimensional audio window management," *International Journal of Man-Machine Studies*, vol. 34, no. 3, pp. 319–336, 1991.

[38] A. Fraietta, O. Bown, and S. Ferguson, "Transparent communication within multiplicities," in *2020 27th Conference of Open Innovations Association (FRUCT)*, 2020, pp. 61–72.