ArraYnger: New Interface for Interactive 360° Spatialization

Neal Andersen
andersne@indiana.edu

Benjamin D. Smith
bds6@iupui.edu

ABSTRACT
Interactive real-time spatialization of audio over large immersive speaker arrays poses significant interface and control challenges for live performers. Fluidly moving and mixing numerous sound objects over unique speaker configurations requires specifically designed software interfaces and systems. Currently available software solutions either impose configuration limitations, require extreme degrees of expertise, or extensive configuration time to use. A new system design, focusing on simplicity, ease of use, and live interactive spatialization is described. Automation of array calibration and tuning is included to facilitate rapid deployment and configuration. Comparisons with other solutions show favorability in terms of complexity, depth of control, and required features.

Author Keywords
Interactive Spatialization, Interface, Panning, Surround Sound, Immersive Audio.

ACM Classification
H.5.2 [Information Interfaces and Presentation] User Interfaces–Graphical user interfaces, J.5 [Computer Applications] Arts and Humanities–Performing Arts, H.5.5 [Information Interfaces and Presentation] Sound and Music Computing–Systems.

1.INTRODUCTION
Surrounding an audience with sound and video presents many exciting potentials for music performance and live intermedia art. The creation of such immersive environments for aesthetic exploration becomes ever more possible as the cost of computational power is continually reduced. However, the complexities of dynamically spatializing audio, moving sounds around a space fluidly and interactively, present significant control and interface issues to the performers, creators, and technicians. Technical solutions for working creatively in large-scale surround environments have been proposed and exist, however the learning curve, cognitive load, and sensitivity of these systems presents serious problems and limitations in live stage use. Based on extensive observation and experience in a surround environment designed for interactive, intermedia performance specific problems have been identified, primarily centering on rapid speaker array calibration, flexible and dynamic configuration, and interactive spatialization of live multichannel audio.

In the world of Acousmatics practices of real-time sound spatialization performance are rapidly evolving beyond conventional analog mixers as new interfaces are designed and developed. The potentials of touch and gestural control devices combined with advanced panning algorithms and acoustical modeling indicates a leap forward in real-time spatialization. However, current solutions present issues with learning curve and interface complexity when artists require a number of channels greater than 2 (stereo). After discussing the background of 360° surround environments, below, the leading live multi-channel mixing systems are examined. An original system design is presented and compared, with the objectives of expediting array setup and calibration [3], capturing and compensating for individual speaker characteristics, and providing intuitive yet extensive interaction with multichannel audio arrays in live performance settings.

2.BACKGROUND
Creating increasingly immersive performances in music and intermedia events is a common trend in the field, apparently seeking to deepen audience engagement and present new aesthetic experiences. Conferences and concerts see pieces and performances that mix video and sonic media with increasing frequency and depth. Given the continually growing computational power of readily accessible personal and mobile computing hardware this trend is likely to continue.

Large-scale institutional installations have led these explorations in recent decades, with the creation of fully surround video rooms and display environments. Primary cases include the Allosphere at U.C. Santa Barbara [5], the CAVE and CUBE environments at many academic institutions around the world, NASA's HIVE environment [3], and Big Tent at Indiana University [9]. While these environments have made progress in exploring the possibilities of 360-degree video and audio, access and expense stand as significant barriers. Working and creating in these spaces requires access only granted to a select few investigators resident at the appropriate institution, and audience sizes are frequently limited in these environments. The construction of these spaces requires extensive investment in technology, and continuing costs for technician maintenance and operation.

Further issues, limiting artistic use and exploration, center on the rigidity of the space and the aura and stigma of the surrounding institution. Many artistic expressions are not feasible in the brick-and-mortar surround environments, such as audience participation interactive performances, or contact improv dance events. Due to the space and audience size limitations even staging an instrument with live electronics performance is often infeasible. The typically elitist aura of the host institutions has a selective effect on audiences, preventing

NIME'17, May 15-19, 2017, Aalborg University Copenhagen, Denmark.

many segments of the populous from accessing the space and artistic works therein.

Alternatively, new interfaces in the form of flexible portable environments are now technologically more feasible and work is actively pursuing their aesthetic possibilities in performance and interaction. Big Tent [9] provides one such solution, with the goals of providing an aesthetically neutral canvas of surround video and audio, capable of holding audiences of up to 60 people. The highly portable nature of the construct allows it to be set it up in non-traditional environments, bringing what is commonly exclusive artistic content to non-typical audiences.

The challenges for a portable environment, such as Big Tent, center on rapid configuration and calibration. With every deployment the video projection and audio speaker array have to be calibrated uniquely. Expediting this process is a critical step to enable the use of these systems and facilitate creation and performance. Successfully setting up a multichannel audio system can be a meticulous process, as the location, orientation, and spectral characteristics of each speaker must be known in order to support accurate spatialization.

Surround sound setups in the realm of research and art music often seek to extend and explore extreme configurations. The number of speakers, such as in the BEAST array [10], often far exceeds that of commercial standards, such as 5.1 (6 channel) and 7.1 (8 channel) surround configurations. Along with maximizing the number of channels, artists in this world require methods of control over the sonic environment that feel natural and are inspiring to use. When dealing with multiple tracks of audio streaming through a large number of speakers a standard mixing board becomes impractical.

3.SPATIALIZATION INTERFACES

Currently available multichannel spatialization interfaces have many usability tradeoffs. More intuitive solutions for controlling multichannel sound typically have a number of limitations. On the other hand, solutions that are robust enough to handle professional level use cases present a significant barrier of entry requiring extreme expertise, and operation in live performance.

In Big Tent [9], performance audio is run through Ableton Live, due to its flexibility and robustness for real-time and fixed-media work. A promising solution for multichannel audio control exists in a MaxForLive (M4L) device, "ak.SendsPan."[1] One instance of the M4L device is placed on each audio track, routing to output send tracks, where each send is assigned to a different speaker. This setup is functional for up to 12 speakers but has a few drawbacks, and prevents use with larger arrays. The speakers are assumed to be in a ring, and irregular configurations cannot be supported. The control is limited to a simple dial control over the Azimuth, determining the location of the sound as a point on the periphery of the speaker ring. While an adjustable "spread amount" is provided, changing the span of the directed sound across speaker points, it does not allow audio to be spatialized inside the ring.

A typical research oriented solution exists in IRCAM's *Spatialisateur* (Spat), originally developed in the early 1990s [2]. This is a robust, modular system that provides myriad features for simulating acoustic spaces, multi-channel convolution reverberation, and spatialization algorithms (Ambisonics, VBAP, etc.). While the functionality is extensive, operation requires a high level of expertise to configure and setup for a new array, environment, or particular composition assumes the user has proficiency in Max programming. Spat boasts support for up to 512 inputs and outputs, however methods for organizing and controlling sound object movement in real-time is not a primary focus of development. Related work focuses on providing trajectory creation and editing interfaces for composers, but again requires a high degree of software knowledge [4] and does not support live manipulation of sound object location.

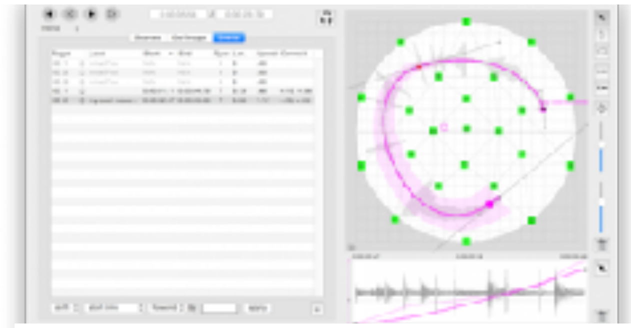

Figure 1a. Zirkonium - Desktop Spatialization Software

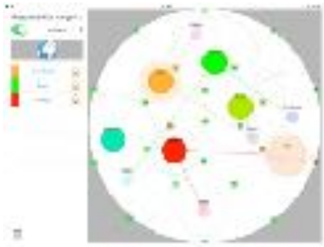

Figure 1b. Zirkpad(Zirkonium's iPad control interace)

Zirkonium [7] presents another solution, intended to support complex arrays of up to 64 speakers in 3D configurations, providing myriad automation and customization options. However, customizing speaker layouts is a tedious process, comprising use of a seperate "SpeakerSetup" application, saving an XML file, and loading it into the main editor. The developers also created an iPad application, "ZirkPad" to mirror control of the primary "Zirkonium Trajectory Editor" application. The editor has rich automation capabilities to move sounds in different ways, which can then be played back during a performance. While Zirkonium is a well-featured multichannel audio editor, the learning curve is steep, requiring extensive expertise in spatialization and knowledge of the specific speaker array. Many composers and performers may find the 88-page manual intimidating, and the number of steps involved to move sounds across the array a prohibiting factor.

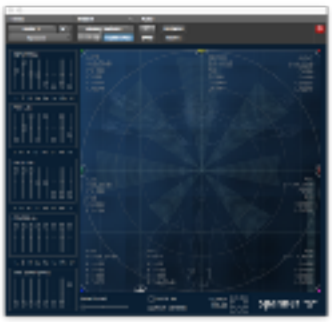

Figure 2a. Spanner(Pro Tools Plugin)

Figure 2b. SpanControl (Spanner's iPad control interface)

[1] http//www.maxforlive.com.

Similar to Zirkonium, Spanner[2] provides a multichannel spatialization iPad interface. The editor has intuitive feedback of audio levels, displayed as bars pulsating from a central point corresponding to the sound nodes' position. It also has grouping, hot key modifiers, and rotation features making the visual control elements significantly more engaging. One downside of Spanner is that it is currently only offered as a plugin for ProTools, preventing its use in live performance settings. Additionally, the plugin is optimized for use with 5.1 and 7.1 surround sound, making use with uniform or complex speaker arrays difficult. The plugin supports a maximum of 8 channels per instance. While film sound editors may find this solution satisfactory, it does not function for art music performance.

Finally, a new system, MIAM Spat, has been proposed [6] to solve problems closely related to those detailed herein. MIAM Spat primarily focuses on spatialization of Acousmatic music over acousmoniums, proposing a model of interpolation between "mixer states" through a simple touch interface. The states can be arbitrarily mapped to any area of the touch surface, allowing the user to spatialize sound by dragging a point between and across overlapping areas. However, currently MIAM Spat appears to support only a single sound input track and is a Windows OS software (limiting application in many setups). While promising an interesting solution, little evidence of applied success is available.

4.“ArraYnger"

Given the significant limitations of the current existing live spatialization interfaces a new interface design is proposed and described herein. The primary design goals of this new system, ArraYnger, are providing robust, fluid spatialization of any number of sound sources over any number of speakers, and simple, strong calibration and configuration of the speaker array. Before sound can be accurately spread over the array, the locations and sonic characteristics of each speaker must be understood (i.e. the array, as an instrument, must be 'tuned'). This is accomplished in ArraYnger with a polar grid of microphones set up in the middle of the array and acoustically "pinging" each speaker to detect its angle and distance [1]. In this way a virtual map of the locations of all the speakers is created, eliminating the tedious process of physically measuring and checking each speaker location.

A secondary step involves capturing the impulse response (IR) of each speaker in order to understand the spectral characteristics of the array and the physical environment. Typically this is accomplished with a sine-wave sweep across the audible spectrum through each speaker individually. Both the location and the frequency response can be recalculated at any time automatically, based on user need.

These steps provide ArraYnger with an estimated virtual map (Fig. 4) and the ability to tune each speaker variably between their natural sound and a de-convolved state (using the captured IR signature) (Fig. 3). It is also possible to bypass the calibration process and manually configure the positions of the speakers through a simple click-and-drag interface.

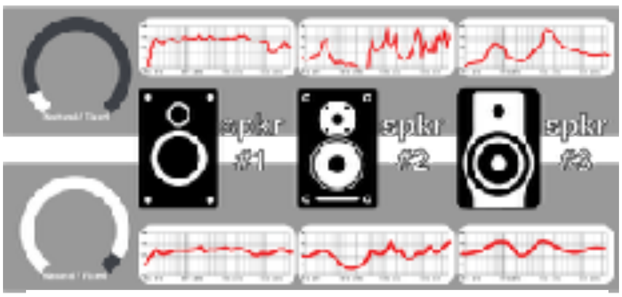

Figure 3. ArraYnger's IR state changer

4.1.Graphical User Interface

In ArraYnger, sound inputs (playback, synthesized, or live audio) are represented as circular nodes indicating the position and spread of each sound object. Mixing and spatializing is possible after assigning any number of nodes to an audio input stream routed through ArraYnger (see sec. 4.2). Manipulating the position of each node moves the associated sound source around the array, employing a distance based amplitude-panning algorithm. The size of a sound node controls the spread of the sound across the array. Node color and other appearance parameters can be customized in the settings panel, which is opened via the cogwheel icon (fig. 5, top right). The interface guides the user to move sound nodes primarily as the method of 'play.' Speaker positions can be displayed as a background layer as black circles, and if required, speaker relocation can be accessed from the estimation menu identified by the magnifying glass icon (fig. 5, bottom right). The help menu (the question mark icon) walks a user through the process of getting up and running with their surround sound setup, in addition to providing specific information about mapping and routing.

4.1.1.Move

Sound object nodes are the primary focus of control in the interface and provide immediate access to spatialization. After the speakers positions are estimated (fig. 4), and one or more sounds are routed through ArraYnger (sec. 4.2), a player can drag the sound objects around on the screen. Multitouch drag gestures move the sound nodes around within the array arena.

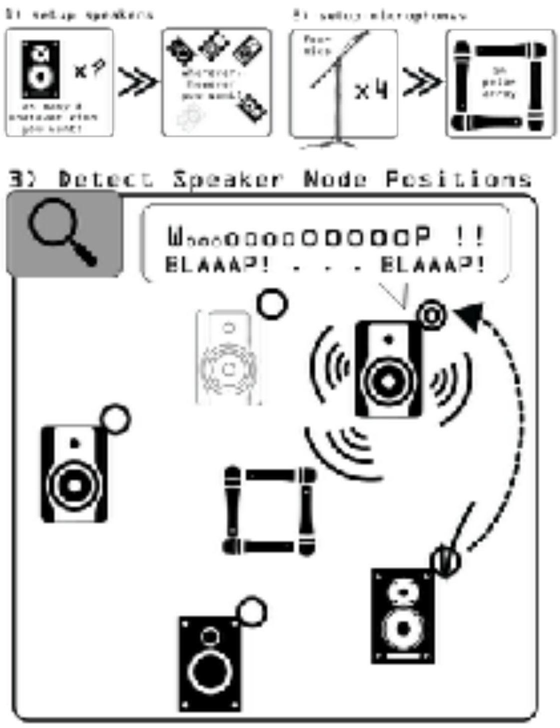

Figure 4. ArraYnger's Speaker mapping procedure

Multiple instances of the same sound node can exist and are treated as independent nodes. Multi-touch allows the performer

[2] http://www.thecargocult.nz/spanner.shtml

Figure 5. ArraYnger U.I. Panel Icons

to move many nodes simultaneously. The result of dragging a sound over a speaker node results in a falloff in volume as it moves further from the center. The very edge of the speaker node circle will have the lowest volume, and the center of the speaker will have the highest.

4.1.2.Scale

Resizing sound nodes is achieved with the conventional multi-touch ‘pinch’ gesture (i.e. two fingers moving together or apart). Only one node can be resized at a time in this way, however an option to resize all nodes simultaneously is available through sub menu options.

4.2.Routing

An audio input channel into ArraYnger becomes a sound object that can be turned on, turned off, doubled, mixed, and spatialized. Routing audio through ArraYnger is modeled on common audio software conventions, attempting to leverage user familiarity with commercial digital audio workstations. For those who are unfamiliar with digital audio routing the global help menu provides a walkthrough. In the settings panel (accessible through the cogwheel icon; see fig. 5), the user first selects the audio input device from their hardware audio interface or virtual audio interfaces (such as soundflower, loopback, jacktrip, rewire, etc.).

Once a device with the desired channels is selected, tracks may be added or removed in the mixing/routing panels (from the “faders” icon; see fig. 5). Since mixing in ArraYnger happens via node movement rather than traditional linear faders, the main purpose of the mixing/routing panel is to make adjustments fitting the player’s personal preferences in this non-traditional environment. This allows for local and global configurations of preferred interactions with the nodes, and lets the player assemble simple or complex routings.

An example of one routing scenario might be 2 instances of a node, which would stream the same audio input channel and carry the same color & label. Another possibility would be the ability to remove an instance of a node or add a copy of an input channel. Or, perhaps there are 16 channels coming in from the audio interface, but the user only wants to use channels 1, 5, and 6 for one piece and channels 2, 6, 7, 7 (second instance) and 11 for another piece. These

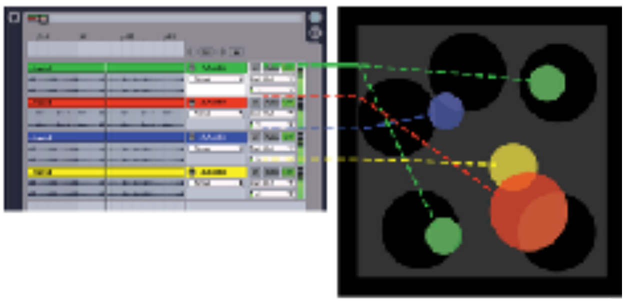

Figure 6. Routing in ArraYnger

configurations can be saved in the routing panel through the presets object, provided by the Max environment. The maximum number of inputs is only limited by the hardware or virtual audio device employed (the maximum of most systems appears to be 64 channels).

5.COMPARISON

The complexity, depth, and required features of spatialization software inform this comparison of current solutions. A performer, spatializing sound in a live context, can benefit from a balance of complexity and ease of use, as well as the depth of control over the available features. Complexity is evaluated in terms of what minimum level of experience/expertise is required for use. Depth is categorized by the level of control one may have over the software. Ideally, the software should allow easy entry, but still allow the user to dive into an expansive and powerful set of features. The final component is simply to evaluate if the spatialization software offers features necessary for live performance.

Touch surfaces provide an ideal interface for multi-source spatialization, as moving many objects with conventional computer interface devices (keyboards, mouse, etc.) is highly constrained. Zirkonium, through ZirkPad, and Spanner, with the iPad app SpanControl, take advantage of this control paradigm with their dedicated touch control applications. ArraYnger primarily focuses on conventional multi-touch gestures to enable real-time spatialization and mixing.

Figure 7. Public gallery installation with ArraYnger

Economy of design, and minimal design, in interfaces can reduce cognitive load, allowing the user to focus on specific tasks. Simplicity in the interface is seen as supporting the goal of easy use, especially under the pressures of live performance settings. The ArraYnger interface is minimal by design, especially in comparison to the cluttered screens of Zirkonium and Spanner. A novice user can begin spatializing and manipulating sound very quickly in ArraYnger, without needing to learn all the intricacies of the more complex systems seen in Zirkonium and Spanner.

Not all surround sound systems offer the ability to configure the speaker location. The program, Zirkonium [7] allows very complex, dense speaker configurations in 3D space. However the trade-off here is the level of expertise/experience required to get up and running, actively spatializing sound and exploiting the capabilities of the software. The nature of needing to manage and pass files between 2 separate applications each time the user needs a new speaker configuration is a detractor for improvisatory set-ups. A feature similar to ArraYnger automatic speaker calibration and IR deconvolver (fig 3.) has not been seen in any of the compared systems.

Spanner, does not let the user configure individual speaker positions. This software, designed for a certain niche of the audio community, is optimized towards traditional 7.1, 5.1, stereo, or mono speaker set-ups. With custom locations not available the features of this program extend more into spatialization techniques like pairing, rotation, and other special features native to ProTools.

Perhaps, the least deep example would be the MaxForLive device, ak.SendsPan, which requires speakers to be in a ring, and provides no additional features or ways of synchronizing and grouping sound objects.

6.FUTURE DIRECTIONS

Beyond the scope of the current design, more features in ArraYnger's interface are planned to streamline and motivate the immersive spatialization experience. As briefly discussed in the design considerations of Perez-Lopez [8], an interactive experience might call for variances dependent upon number of performers, intended users, and whether the interface implements shared control or not. These concepts match the intended direction of ArraYnger's interface, being easy to connect and configure according to the number of participants. Towards this end a web-based version is planned to allow many performers to interact with the spatialization concurrently through any mobile computing device. This could provide the opportunity to integrate an easy-to-connect shared sonification experience.

7.REFERENCES

1. N. Anderson and B. D. Smith. Relative Sound Localization for Sources in a Haphazard Speaker Array. In *Proceedings of the 42nd Annual International Computer Music Conference* (ICMC 2016) (Utrecht, The Netherlands, Sep 12th -16th, 2016).
2. T. Carpentier, M. Noisternig and O. Warusfel, Twenty years of Ircam Spat: looking back, looking forward. In *Proceedings of the 41st International Computer Music Conference* (ICMC 2015) (Denton, TX, Sep. 25th to Oct. 1st, 2015).
3. "DEVELOP's HIVE: Redesigning and Redefining the 3-D Virtual Environment." 2012. Earthzine. August 13. http://earthzine.org/2012/08/13/develops-hive-redesigning-and-redefining-the-3-d-virtual-environment/.
4. J. Garcia, J. Bresson, M. Schumacher, T. Carpentier, and X. Favory, Tools and applications for interactive-algorithmic control of sound spatialization in OpenMusic. In inSONIC2015, Aesthetics of Spatial Audio in Sound, Music and Sound Art. (Vancouver).
5. T. Höllerer, J. Kuchera-Morin, and X. Amatriain. 2007. "The Allosphere: a Large-scale Immersive Surround-view Instrument." In Proceedings of the 2007 Workshop on Emerging Displays Technologies: Images and Beyond: The Future of Displays and Interacton, (San Diego, California).
6. G. Le Vaillant, and R. Giot, 2014. Multi-touch Interface for Acousmatic Music Spatialization.In Proceedings of the 40th International Computer Music Conference (ICMC 2014) (Athens, Greece, Sep. 14th to Sep. 20th, 2014).
7. C. Miyama, G. Dipper and L. Brümmer. Zirkonium 3.1 - a toolkit for spatial composition and performance. In *Proceedings of the 42nd Annual International Computer Music Conference* (ICMC 2016) (Utrecht, The Netherlands, Sep 12th -16th, 2016).
8. A. Perez-Lopez. 3DJ: A supercollider framework for real-time sound spatialization. In *Proceedings of the 21st International Conference on Auditory Display* (ICAD 2015) (Graz, Austria, July 8th-10th, 2015).
9. B. D. Smith, and R. Cox. Big Tent: A Portable Immersive Intermedia Environment. In *Proceedings of the 42nd Annual International Computer Music Conference* (ICMC 2016) (Utrecht, The Netherlands, Sep 12th -16th, 2016).
10. S. Wilson, and J. Harrison, 2010. Rethinking the BEAST: Recent developments in multichannel composition at Birmingham ElectroAcoustic Sound Theatre. *Organised Sound*, 15, no. 03 (2010): p. 239-250.