Some Considerations on Creativity Support for VR Audio

Anıl Çamcı*
Department of Performing Arts Technology
University of Michigan

ABSTRACT

As the consumer interest in VR grows, the community of content creators working in this domain expands accordingly. Given the intrinsic role of audio in VR experiences, this growth necessitates authoring tools for immersive audio that can cater to a wide range of designers regardless of their expertise in spatial audio. In this article, we discuss some of the modern considerations on designing interactive tools for creativity support in VR audio. We provide examples from existing spatial audio design software, and discuss areas in which new tools can facilitate, for expert and novice users alike, the use of immersive audio in compelling new VR experiences.

Index Terms: Human-centered computing—Interaction Paradigms—Virtual Reality—; Applied computing—Arts and Humanities—Sound and music computing

1 INTRODUCTION

The role of sound in creating plausible immersive experiences has been acknowledged since the early days of VR. While Heilig incorporated binaural audio into his prototype of a head-mounted display [7], Sutherland contemplated audio displays in his vision of an ultimate VR [14]. However, sound has since remained an afterthought in VR research, which has primarily focused on addressing technical challenges in the visual domain. Over the last decade, breakthroughs in low-persistence displays combined with modern spatial tracking systems have brought us closer than ever to the mass adoption of virtual reality as a mainstream entertainment medium, creating new avenues for research into other aspects of VR beyond the visual.

Although spatial audio research dates back to the early days of sound reproduction [5], the recent advances in VR technology have reinvigorated the interest in advanced spatial audio techniques, such as Ambisonics and binaural audio. Leading technology companies today are putting significant effort into developing tools that harness the potential of these techniques in VR applications. Most recent examples include Facebook's Spatial Workstation, and Google's Resonance Audio, which facilitate the use of immersive audio with modern VR platforms. However, these tools often cater to expert users: the user interactions offered in these tools are either adopted from expert-oriented sound spatialization software, or they adhere to the interaction paradigms of the host applications which they augment. These host applications, such as game engines and digital audio workstations (DAWs), are commonly geared towards expert users as well.

It is reasonable to assume that the consumer appeal of VR will continue to grow with new hardware iterations, which will prompt more content creators to work in this domain. Given the fundamental part that sound plays in engaging immersive experiences, authoring tools will need to enable expert and novice users alike to incorporate immersive audio in their VR designs. Regardless of one's level of expertise with spatial audio, the inherent ability of our auditory systems to perform highly complex sound localization tasks in everyday situations makes us adept at perceiving spatial sound. Creativity-support tools for immersive audio can leverage this ability with interfaces that empower designers to work with immersive audio. To achieve this, these tools should bridge the gap between our ability to reason about sound in space and the computational methods that are used to render such sounds.

In this article, we first offer an overview of current tools and platforms that enable immersive audio design. We then discuss some of the challenges and opportunities in designing user interfaces for creativity support in immersive audio in light of recent developments in VR technology. While doing so, we provide insights gained from studies undertaken with *Inviso*, [1] which is a browser-based spatial audio software that we designed as a creativity support tool for the rapid prototyping of immersive sonic environments.

2 CREATIVITY SUPPORT TOOLS FOR IMMERSIVE AUDIO

The interface of a creativity support tool is intrinsically informed by its active userbase. For instance, in the early 90s, DAWs were primarily geared towards professional recording engineers because the hardware systems needed to run these software applications were prohibitively priced for prosumers and consumers. Significant advances in digital computing since have made high-performance computers available to a wide array of users. As a result, modern DAWs reflect a similar variety ranging from applications that primarily serve professionals, such as Avid ProTools, to those that are intended for casual and novice users, such as Apple GarageBand. Similar trends in creativity support tools can be observed in other domains including photography, film, and design.

Software tools for immersive audio have so far shared a similar fate as those of early DAWs mainly because immersive media had not penetrated the consumer domain. As a result, these tools have been traditionally designed by and for research institutions and production companies that possess the means to instate state-of-the-art immersive systems. While the software tools developed by such institutions enable the design of complex spatial audio scenes, these tools mostly presuppose an existing understanding of spatial audio techniques. With the imminent mass adoption of VR technologies, immersive audio bears a potential to become a medium of creativity for a much wider userbase.

2.1 Stand-alone Software

Numerous software tools for immersive audio function as stand-alone systems that take an audio feed and spatialize it for different output configurations. For instance, Spat is a spatial audio toolkit developed at IRCAM [4]. Designed in the visual programming language Max, Spat supports a variety of spatial audio formats including stereo, binaural, VPAB, as well as B-Format and higher-order Ambisonics. The Spat Oper object offers a high-level UI with precise control over various parameters that pertain to the timbral qualities of the sound sources, and the reverberant characteristics of the virtual space that surrounds them. The UI also affords 2D (overhead) and 3D views of the sound field, where the user can position virtual speakers and sound sources around the visual representation of a stationary listener. Spat Revolution, which is the

*e-mail: acamci@umich.edu

[1] http://inviso.cc

2019 IEEE Conference on Virtual Reality and
3D User Interfaces
23-27 March, Osaka, Japan
978-1-7281-1377-7/19/$31.00 ©2019 IEEE

product of a collaboration between Flux:: and IRCAM, provides similar functionality in stand-alone and DAW plug-in formats.

The Ambisonics Tools for Max developed at ICST offers Ambisonic encoding and decoding of point sources [11]. In addition to parametric controls, the UI consists of separate overhead and lateral views of the sound field, where the user can move sound sources and virtual speakers with mouse interactions. Utilizing similar UI elements, The HOA library developed at CICM offers binaural and higher-order Ambisonic rendering of audio [12]. The HOA library can be deployed for a range of platforms including Max, PureData, Faust and openFrameworks.

Zirkonium developed by The Institute of Music and Acoustics at ZKM focuses on the notation and composition of spatial music [9]. Programmed in PureData, Zirkonium can work with internal and external spatialization algorithms. The primary UI element is an overhead view of the sound field where mouse interactions can be used to position sound objects and object subgroups, as well as to create motion trajectories. The UI also allows the parametric control of these elements. Similarly, SoundScape Renderer developed at the TU Berlin's Usability Lab [6], offers a UI that consists primarily of an overhead view of the sound field, and enables similar mouse-based and parametric controls. In addition to VBAP, Ambisonics and binaural audio, this tool supports wave-field synthesis.

Inviso developed by the author of this article is another stand-alone application that supports the design of immersive audio scenes in the web browser [3]. The main UI element of this tool is a 3D scene which can be populated with omni- and multi-directional sound objects that can be animated with motion trajectories, and non-directional sound zones with arbitrary shapes defined by the user. The tool combines common mouse-based interactions with contextual UI windows that offer additional control over sound parameters.

2.2 Game Engines

Game engines, which are commonly used for the design of immersive environments, offer basic spatial audio functionality that serves to augment the visual domain with such elements as point sources and reverb zones. The UI for such functionality primarily adhere to the interaction scheme of the game engine, which is, first and foremost, geared towards the design of 3D graphical environments. As a result, these UIs are often not optimized for audio applications.

2.3 Workstations

Audio event-buffer management workstations, such as Wwise and FMOD, also provide immersive audio functionality. These tools are often integrated with other production software, most commonly game engines, to achieve procedural audio. Furthermore, these tools can also be extended with spatial audio plug-ins in ways that facilitate hybrid interaction schemes.

2.4 Plug-ins

Numerous spatial audio software applications are designed as plug-ins for a range of host applications including DAWs and game engines. For instance, Facebook's Spatial Workstation extends DAWs with binaural and Ambisonic output features. The UI of this application puts a similar emphasis on a top-down view of a sound field surrounding a stationary listener. Google Resonance Audio offers a similar suite of plug-ins that can also extend game engines such as Unity 3D and Unreal Engine. The Resonance Audio objects take the form of rudimentary game objects native to the host platform with additional parametric fields that pertain to spatial audio. As a result, the UI to control such objects primarily adheres to the interaction scheme of the host application.

Other modern spatial audio plug-ins include VisiSonic's RealSpace 3D Audio,[2] New Audio Technology's Spatial Audio Designer,[3] AudioEase's 360Pan Suite,[4] SSA's a7 Ambisonics Suite,[5] MNTN Production Suite,[6] and the Ambisonic Toolkit.[7] Similar to the aforementioned applications, the most common UI element in these plug-ins is a 2D overhead view of the sound field. Furthermore, these interfaces often employ extensive parameter spaces that enable precise design of spatial audio scenes.

3 Challenges and Opportunities

The design of a creativity support tool for immersive audio presents many challenges in terms of usability, cross-modal representations, and the mapping of 2D interfaces to 3D sounds. These challenges necessitate modality-specific approaches to interaction design for immersive audio. Furthermore, modern VR platforms afford opportunities to address some of the issues in this domain.

3.1 Catering to a Range of Users

While the level of parametric detail offered by modern spatial audio design tools enables the precise authoring of immersive audio scenes, it can also prohibit novice users who are not versed in the parameter spaces common to spatial audio from making creative use of these tools. With wider adoption of VR technologies, authoring tools for immersive audio will need to cater to a wider range of content creators. Following Shneiderman's design principles for creativity support tools, these tools should present a low barrier of entry for novice users, while offering high ceilings for experts [13].

In *Inviso*, we utilized common mouse-based interactions for object manipulation and 2D drawing to facilitate the rapid prototyping of immersive audio scenes. Besides these interactions, which have been reported by users to enable intuitive control of an audio scene, contextual UI windows give access to parametric control of the output for applications that require a higher level of precision. Furthermore, the interface affords continuous modes where the user can manipulate the output at various scales ranging from object level to the scene level. Although the browser-based implementation of this tool inhibits the use of computationally-intensive acoustic modelling techniques, it offers an easily accessible cross-platform operation. For such a tool to appeal to expert use cases besides prototyping, users emphasized the need for a means to transfer their designs to different platforms, which can be considered a fundamental feature for a creativity support tools in general.

3.2 Visual Representation of Spatial Audio

The UI of an audio software necessitates visual representations of sounds and audio-related concepts. For instance, in DAWs, audio channels are often represented as horizontal tracks that represent the temporal progression of a sound in a similar fashion to tape tracks.

A common element of spatial audio is the point source, which is often represented in software UIs as a dot or a sphere. Point sources, which can be moved around the listener, define omnidirectional sound objects that emanate sound in all directions. Sound sources with limited directionality are commonly represented with cones. Similar to the cone of a speaker, the sound cone demarcates a limited area of dispersion. Sound cones in spatial audio are often defined as a combination of two cones: while an inner cone articulates the range within which a sound will be heard directly, an outer cone serves as a cross-fade region for smooth transitions.

Sound-producing objects in our daily lives can rarely be categorized as simply as omni- or uni-directional. Instead, such objects set off complex sound propagations in multiple directions. Emulating this complexity, sound objects in *Inviso*, can be populated with an arbitrary number of sound cones with unique range and orientations.

[2] https://realspace3daudio.com

[3] https://newaudiotechnology.com/products/spatial-audio-designer/

[4] https://audioease.com/360/

[5] https://ssa-plugins.com

[6] https://mntn.rocks/production-suite

[7] http://ambisonictoolkit.net

3.3 Controlling 3D Sound with a 2D Interface

Besides the keyboard used for text entry, the mouse (or the trackpad) remains the most common hardware interface for modern computers. The mouse allows the user to interact with software elements laid out on a 2D display. Designing 3D spaces with such interactions requires graphical user interfaces that map 3D elements to 2D controls. A common approach to such mappings involves axis handles for the dimensions on which an object can be moved. For instance, in Unity, game objects placed into the scene display arrows for the X, Y, Z axes which can be click-and-dragged to move the object in 3D space. Another approach to the 2D representation of a 3D environment is using separate viewports for different combinations of spatial dimensions. For instance, while one viewport can give the user control over the X and Z axes with an overhead view, another viewport can show a lateral view with Y and Z controls.

Previous research has shown that separate viewports can improve performance with tasks that involve 3D manipulation [10]. Axis handles, on the other hand, enable an easier contextualization of objects within a single viewport. With spatial audio design tools, the characteristics of the auditory system can inform the approach to the mapping of 3D objects to 2D interfaces. For instance, our auditory systems can localize sound sources more accurately on the horizontal plane than it does on the vertical plane [8]. In *Inviso*, this property was exploited with a singular viewport that grants control over different axis combinations based on the angle at which the user is viewing the 3D scene. The UI is initiated with an overhead view that allows the user to move the sound objects on the X and Z axes. If, however, the user tilts the 3D scene beyond a certain degree, the UI switches to a mode where the user can move objects exclusively on the Y axis. This approach prioritizes the horizontal plane, similar to many 2D sound spatialization tools, but at the same time offers a state in which the vertical axis can be manipulated within the same viewport.

Over the last decade, new technologies such as the Microsoft Kinect and the Leap Motion have facilitated the incorporation of depth tracking into interaction design. With increasing use of depth cameras in consumer products, it may soon be more commonplace for computers to have 3D input systems. Authoring tools for immersive audio can leverage such systems by mapping the immediate physical surroundings of a designer to the virtual space where they are designing an audio scene.

3.4 Authoring VR Audio within VR

In our studies, users who designed immersive audio environments with Unity reported that having to switch between editing the scene on their computers and testing their design in VR has been disruptive to their workflows. To address such disruptions, an increasing number of developers today are designing tools for content creation in VR. Popular examples are Google's Tilt Brush and Blocks applications, which facilitate the design of 3D visuals within VR. In the audio domain, DearVR has recently introduced the Spatial Connect application, which can be used to control the spatial audio features of a DAW from within VR. [8]

In another example, Atherton and Wang recently introduced an audio programming language for VR, where 3D representations of signal processing blocks can be combined to generate procedural audio within VR [1].

Similarly, we are currently working on a VR implementation of *Inviso* using Unity and Google Resonance Audio. This version will afford a first-person view of the soundfield, which the user can navigate virtually. Using hand-held controllers, such as the Oculus Touch and the HTC Vive Controllers, the user will be able to directly manipulate the sound objects within VR and save their designs for later use in the web browser or in VR.

[8] https://dearvr.com/products/dearvr-spatial-connect

3.5 Head-tracking Systems

The head-tracking systems that are bundled with modern VR platforms offer notable benefits for immersive audio design. Previous research has shown that head-tracking can improve sound localization performance in spatial audio applications. Particularly, front-back confusions encountered in binaural audio can be significantly remedied with head tracking [2]. Therefore, outside-in tracking technologies, such as the Oculus Constellation and the Vive Lighthouse systems, afford increased spatial fidelity in immersive audio applications. With upcoming VR systems that employ inside-out tracking, such as the Oculus Quest and the HTC Vive Cosmos, these applications can be untethered from a base station, allowing immersive audio authoring in large-scale environments.

4 Conclusion

In this article, we discussed some of the considerations that can inform the design and implementation of new creativity support tools for immersive audio. We argue that these tools should enable a range of users to make creative use of spatial sound, in the same way that graphics editing applications help users materialize visual ideas. We believe that such an endeavor would necessitate the formulation of new user interaction schemes that combine common UI elements found in existing spatial audio design tools with new ones that are informed by the inherent properties of spatial sound and the affordances of new VR platforms.

References

[1] J. Atherton and G. Wang. Chunity: Integrated audiovisual programming in unity. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 102–107, June 2018.

[2] D. R. Begault, E. M. Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *Journal of the Audio Engineering Society*, 49(10):904–916, 2001.

[3] A. Çamcı, K. Lee, C. J. Roberts, and A. G. Forbes. Inviso: A cross-platform user interface for creating virtual sonic environments. In *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp. 507–518. ACM, 2017.

[4] T. Carpentier, M. Noisternig, and O. Warusfel. Twenty years of ircam spat: looking back, looking forward. In *Proceedings of the International Computer Music Conference*, pp. 270–277, 2015.

[5] B. Fox. Early stereo recording. *Studio Sound*, 24(5):36–42, 1982.

[6] M. Geier and S. Spors. Spatial audio with the soundscape renderer. In *Tonmeistertagung–VDT International Convention*, 2012.

[7] M. L. Heilig. Stereoscopic-television apparatus for individual use, October 1960. US Patent #2,955,156.

[8] E. A. Lopez-Poveda. Development of fundamental aspects of human auditory perception. In R. Romand and I. Varela-Nieto, eds., *Development of Auditory and Vestibular Systems*, pp. 287–314. Academic Press, San Diego, 2014.

[9] C. Miyama and G. Dipper. Zirkonium 3.1-a toolkit for spatial composition and performance. In *Proceedings of the International Computer Music Conference*, pp. 312–216, 2016.

[10] J.-Y. Oh and W. Stuerzlinger. Moving objects with 2d input devices in CAD systems and desktop virtual environments. In *Proceedings of Graphics Interface*, pp. 195–202, 2005.

[11] J. C. Schacher. Seven years of icst ambisonics tools for maxmsp–a brief report. In *Proceedings of the International Symposium on Ambisonics and Spherical Acoustics*, 2010.

[12] A. Sèdes, P. Guillot, and E. Paris. The HOA library, review and prospect. In *Proceedings of the ICMC-SMC Conference*, pp. 855–860, 2014.

[13] B. Shneiderman. Creativity support tools: Accelerating discovery and innovation. *Communications of the ACM*, 50(12):20–32, 2007.

[14] I. E. Sutherland. The ultimate display. In *Proceedings of the IFIP Congress*, pp. 506–508, 1965.