

Cross-platform and Cross-reality Design of Immersive Sonic Environments

Aaron Willette, Nachiketa Gargi, Eugene Kim, Julia Xu, Tanya Lai and Anil Çamcı

University of Michigan
Ann Arbor, USA

{aawill, ngargi, keugene, juliawxu, tanyalai, acamci}@umich.edu

ABSTRACT

The continued growth of modern VR (virtual reality) platforms into mass adoption is fundamentally driven by the work of content creators who offer engaging experiences. It is therefore essential to design accessible creativity support tools that can facilitate the work of a broad range of practitioners in this domain. In this paper, we focus on one facet of VR content creation, namely immersive audio design. We discuss a suite of design tools that enable both novice and expert users to rapidly prototype immersive sonic environments across desktop, virtual reality and augmented reality platforms. We discuss the design considerations adopted for each implementation, and how the individual systems informed one another in terms of interaction design. We then offer a preliminary evaluation of these systems with reports from first-time users. Finally, we discuss our road-map for improving individual and collaborative creative experiences across platforms and realities in the context of immersive audio.

Author Keywords

Virtual Reality, Augmented Reality, Immersive Audio, Creativity Support

CCS Concepts

•Human-centered computing → Virtual reality; *Human computer interaction (HCI)*; •Applied computing → Sound and music computing;

1. INTRODUCTION

Creativity support tools play an important role in expanding the user base of media platforms. With emerging technologies in particular, making such tools more accessible can support the sustainability of a platform by way of facilitating the work of content creators who will design engaging experiences that will draw in more users. The current selection of design tools for VR content creators are arguably geared towards experts and can be prohibitive for inexperienced designers. Furthermore, the existing VR design workflows are less intuitive than that of many modern media design tools due to the separation between the design and experience stages involved in VR development.

In this paper, we discuss a suite of immersive audio design tools, named Inviso XR, that can support workflows across desktop, virtual reality and augmented reality platforms. Designed to cater for both novice and expert users, these tools facilitate diverse creative experiences and within-XR content creation. We detail the overall design considerations behind each component of this suite. We then identify some of the challenges we encountered while implementing functionality across different platforms. Finally, we offer the results of a preliminary study designed to reveal usability considerations that will inform the next phase of our work on this suite.

2. RELATED WORK

Many existing tools enable the design of spatial audio scenes on desktop computers [2, 7, 3]. In these tools, the user is often represented with a stationary listener node at the center of the virtual space as they create complex sonic environments with such components as point sources, directional cones, and reverberant zones.

A growing number of projects implement similar functionality in VR. In their exploration of audio engineering in VR, Jordan et al. created a system where predefined sound sources can be attached to objects in the scene and manipulated with the Oculus Touch controllers [4]. In a user study about this system, the researchers found that the users were able to easily utilize grab-based controls, but had trouble using a laser pointer mechanism to control objects from a distance.

Other VR tools function as an immersive controller layer over more traditional digital audio workstations (DAW). For instance, in Dear VR Spatial connect, ¹ the tracks of a DAW session is visualized in VR as spherical sound objects, which the user can reposition in 3D space to change how the corresponding sounds are spatialized.

In their sound art installation, which was designed to promote collaborative experiences in audio augmented reality, Kiefer and Chevalier designed a system where they augmented the physical components of their installation with audio elements in AR [5]. In an evaluation of this system, the users reported a high degree of immersion to the extent of confusing whether sounds originated from the physical environment or AR.

3. INVISO

Inviso is a browser-based application for designing immersive sonic environments. As a creativity support tool, it is designed to enable non-experts to engage with immersive audio creatively while providing functionality that can satisfy the needs of expert users. Using open-source APIs such

¹<https://www.dearvr.com/products/dearvr-spatial-connect>



Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). Copyright remains with the author(s).

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

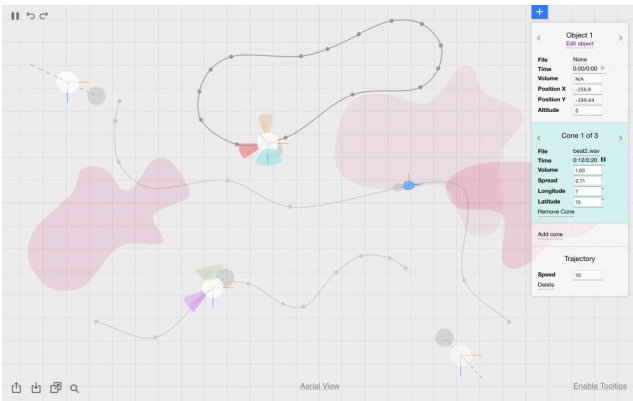


Figure 1: A screenshot from Inviso.

as WebGL and Web Audio, it implements fundamental components of spatial audio such as sound objects, sound cones, and sound zones. The user can navigate through the scene, as seen in 1, using the W, A, S, D keys on the keyboard. The binaural output of the system is based on the design of the environment and the virtual listener node's position in it. The sound objects can be populated with an arbitrary number of cones that can have variable base radius and height properties. The user can also define arbitrary motion trajectories for each object. A more detailed description of Inviso's functionality is offered in a previous publication [1]. The tool itself is available online at <http://inviso.cc>.

4. INVISO VR

VR presents an opportunity to overcome two common limitations in spatial audio design: first, in spatial audio tools that implement surround audio rendering, the 3D virtual space is often controlled with a 2D interface on a 2D display. Second, in these tools, the visuals are not spatially mapped to the acoustic output, offering users what is essentially a 3rd-person control over the design. In VR however, the dimensional disparity between the interface and the content, as well as the displacement between the virtual listener node and the user's head can be eliminated by giving the user a first-person view into a binaural audio scene.

In the VR implementation of Inviso, seen in Fig. 2, we offer a room-scale representation of a virtual audio environment with visual representations of the sound elements akin to those in the desktop version. The user can generate an arbitrary number of sound objects, populate them with sound cones, and give them motion trajectories. Once an object or a cone is created, these can be manipulated with intuitive grab-and-move actions. When trajectory drawing is activated, moving an object creates a spline curve, which the object will follow once released. All sounds are rendered to binaural audio using Google Resonance Audio.

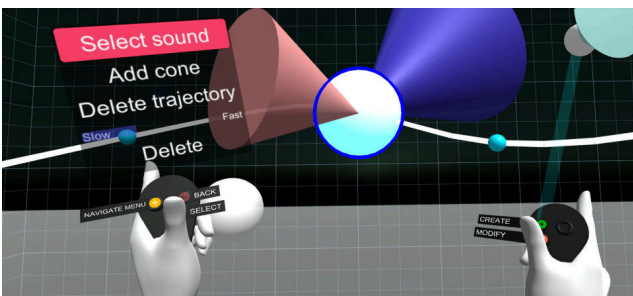


Figure 2: A screenshot from Inviso VR.

5. INVISO AR

The AR version of Inviso ports the core functionality of the browser system to handheld AR platforms. The plane-detecting systems used to save generated objects in world space allow AR environments to be created on devices such as tablets and phones. Inviso AR utilizes the ARKit and ARCore frameworks for device pose estimation. Furthermore, plane detection available in these platforms are used to implement acoustic and visual occlusion of objects. Inside-out motion tracking based on plane detection is used to allow users to clearly see how the objects in world space interact with their environment.

As seen in Fig. 3, the viewport primarily consists of the 3D augmented camera overlay with a context-dependent 2D interface on the right. Inviso AR's UI implements a variety of input methods including AR interactions as well as those achieved with standard buttons and sliders. Placing and moving objects are room-scale interactions that depend on the position of the device in 3D space, while the manipulation of cones is achieved by touch controls such as the pinch gesture used for resizing cones.

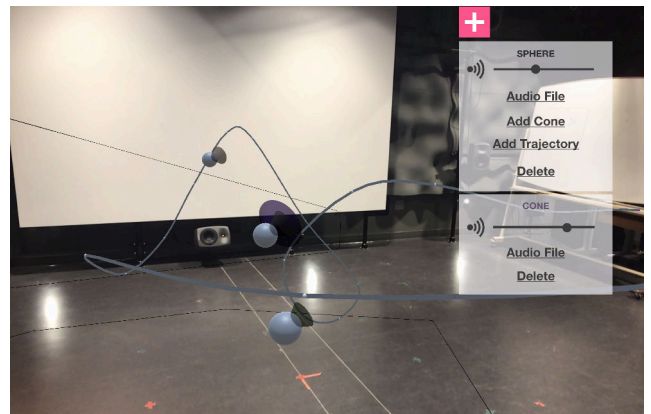


Figure 3: A screenshot from Inviso AR.

6. INTERACTION CHALLENGES

Desktop, VR and AR platforms offer vastly different user experiences. Adapting a desktop tool to VR and AR applications therefore require unique solutions to various interaction design challenges. Here we discuss some of our findings on how to address such challenges.

6.1 Desktop Implementation

The desktop version of Inviso adopts common UI elements such as buttons, sliders, and number boxes. Furthermore, it uses contextual menus that appear based on object selection to give the user parametric control over object attributes. One of the main challenges in designing a desktop interface for a 3D audio tool is the handling of the spatial dimensions with a lower-dimensional input device (i.e. a 2D trackpad or mouse), and the display of this content on a 2D screen. Among the methods to alleviate these challenges are the use of axis-handles or multiple views that display the same object from different reference points. Inviso adopts a two-tiered approach to the manipulation of sound objects and trajectories in 3D. An aerial view allows the user to move the sounds only on the horizontal plane. If, however, the user tilts the camera view beyond a certain degree, the system switches into an altitude mode, where the lateral position of the objects become fixed and they can only be moved vertically. This limited approach exploits the human auditory system's ability to localize sounds more accurately

on the horizontal plane [6], while at the same time enabling the manipulation of object elevation in the same viewport without causing perspective ambiguities.

6.2 VR Implementation

While the VR implementation of Inviso simplifies some of the desktop interactions with intuitive room-scale grab-and-move actions, it also brings up design challenges with regards to the parametric and menu-based UI elements that are easier to control with a 2D input devices, such as a mouse or a trackpad. Using modern UI standards for VR, we implemented virtual representations of the Oculus Touch controllers with persistent labels for each button, as seen in Fig. 2, to reduce the need for memorizing button functionality. Furthermore, global interactions like selecting and deselecting objects are mirrored on both hands to minimize the variety of interactions between the two controllers. Other UI elements, such as volume controls and sound selection menus are navigated with the thumbsticks on the Touch Controller.

6.3 AR Implementation

The AR version of Inviso combines standard 2D UI elements such as buttons and sliders from the desktop version with room-scale interactions. Since AR depends on filling the screen with a viewport into the augmented world, one major challenge was determining a balance between 2D UI elements that offer direct control but clutter the screen and direct interactions with the objects in AR space itself, such as moving sound objects and pinching to change the base radius of cones.

Another design challenge for the AR version originates from having multiple points of spatial reference when executing interactions with 3D objects overlaid onto a 3D space that is viewed and controlled with a 2D input mechanism. Specifically, when rotating a cone around a sphere, the virtual object, the hand-held device and the user's touch input operate on different spatial layers, causing potential conflicts between user intent and system response. Drawing inspiration from the desktop version that enables cone placements on the visible surface of a sphere but requires the sphere itself to be rotated to place cones elsewhere, we limited cone placement to the surface of the sphere that the AR device's camera is facing. This yielded natural interactions where the user can hold onto a cone and move around the sphere to place it in an opposing direction.

7. EVALUATION

We conducted a preliminary evaluation of these three systems to better understand the usability challenges and opportunities in each platform. We recruited 4 users to complete the same design task with each system and respond to questions about their experience in written format.

7.1 Task

The users were given a task to prototype a simple sonic environment with three elements: a stationary sound object, a stationary sound object with two cones, and a sound object with two cones on a trajectory. They were first asked to sketch a plan for their design on paper. They then implemented this plan in all three environments. As a follow-up, they were asked to respond to a survey including the following questions:

- Please rank the three systems in terms of ease of use when completing the given task.
- Were there features in either platform that were problematic to deal with?

- Please rank the three systems in terms of how convincingly they managed to map the sounds to the visuals?
- Which system offered the most compelling design experience overall?

7.2 Results and Discussion

All users were able to recreate their sketches in each platform. While none of the users described themselves as expert VR users, they all reported having tried VR at least once. Regardless of their experience level, they were able to use the thumbsticks on the Oculus Touch controller to navigate through menus. None of the users reported feeling discomfort with the VR experience.

Three out of four users identified the desktop version as the easiest to use. On the other hand, this version was found to offer the least convincing mapping between sounds and visuals, while the VR version was ranked the highest in that regard.

One of the users who ranked the desktop version as the easiest to use added that the environment in the desktop version did not feel as real as in other platforms because it did not offer a first-person perspective, ranking the mapping between the sounds and visuals in the desktop version as the least convincing. This user also ranked the VR version as the one that offered the most compelling experience due to its affordance of a first-person view and the fact that their view was blocked from visual distractions in the physical world. Conversely, another user found the AR experience to be most compelling because they were able to contextualize objects alongside the elements of the physical world.

One user described the mapping between the sounds and visuals to be the most convincing in the desktop version and attributed this to being able to see an overview of the relationship between the sound sources and the listener node. While they found the desktop experience to be the most compelling for the same reason, they added that if they had more experience with the VR version, it would likely be their preferred platform for design. Similarly, two other users reported that, while the UI in the VR version felt like a barrier for design, this mainly had to do with their lack of experience with the interface.

Three out of four users ranked the VR version as the one that offered the most compelling experience. The fourth user, who ranked the AR version as such, indicated a preference towards being able to mix the virtual audio elements with real ones.

With regards to the problematic aspects in either platform, one user reported that they had wanted to move the cones on an object that was moving on a trajectory and that they were not able to do so. The same user reported some of the AR objects would momentarily disappear from their view; this is likely due to the tablet device losing tracking of the planar surfaces, which the placement of the objects are based on. Another user mentioned that modifying objects with one hand while holding the tablet with the other felt cumbersome at times.

8. CONCLUSIONS

In this paper, we discussed our ongoing work on a suite of tools for immersive audio design called Inviso XR. We believe that this suite can support creative workflows in many situations including artistic performances, content creation, VR and audio education, and architectural prototyping. Our preliminary findings with users indicate that the UI elements in the VR version require a learning process. Despite this, the VR experience was found to offer the most compelling experience overall with a user preference towards

the first-person perspective into the sonic environment. On the AR and desktop versions, the users' familiarity with desktop and touch interactions facilitated their design processes. The mapping between the sounds and visuals were generally found to be convincing with the desktop version being ranked as the easiest to use although there was a clear preference towards the spatial mapping afforded by the AR and VR versions.

Based on our findings, we plan to continue exploring intuitive interaction methods for the VR version that mitigates to need for experience with VR UIs. Furthermore, we plan to explore distant modification techniques so as to allow the user to design more expansive environments. In the near future, we also hope to implement real-time networked communications between the different components of the Inviso XR Suite. This will enable multi-user creative and instructional experiences: for instance, a user working with a VR headset will collaborate with a remotely-located user who is working on the same design with a tablet in room-scale AR while a third user contributes to the design on a desktop computer.

9. REFERENCES

- [1] 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*, pages 507–518. ACM, 2017.
- [2] T. Carpentier. Binaural synthesis with the Web Audio API. In *Proceedings of the 1st Web Audio Conference*, January 2015.
- [3] M. Geier and S. Spors. Spatial audio with the soundscape renderer. In *27th Tonmeistertagung-VDT International Convention*, 2012.
- [4] D. Jordan, F. Müller, C. Drude, S. Reinhold, V. Schomakers, and M. Teistler. Spatial audio engineering in a virtual reality environment. In W. Prinz, J. Borchers, and M. Jarke, editors, *Mensch und Computer 2016 - Tagungsband*, Aachen, 2016. Gesellschaft für Informatik e.V.
- [5] C. Kiefer and C. Chevalier. Towards new modes of collective musical expression through audio augmented reality. In *Proceedings of the International Conference on New Interfaces for Musical Expression*. Association for Computing Machinery, 2018.
- [6] E. A. Lopez-Poveda. Development of fundamental aspects of human auditory perception. In R. Romand and I. Varela-Nieto, editors, *Development of Auditory and Vestibular Systems*, pages 287–314. Academic Press, San Diego, 2014.
- [7] R. Penha and J. Oliveira. Spatium, tools for sound spatialization. In *Proceedings of the Sound and Music Computing Conference*, pages 660–667, 2013.