Continuum: a WebVR Experience Inspired by the Multiverse

Anon
Anon Uni
1800 Anon Drive
Anon, XY, ABC
anon@zzz.edu

ABSTRACT

Continuum is a WebVR experience inspired by the concept of the multiverse: the idea that there are an infinite number of possible universes all occurring simultaneously in other timelines. In this project, the manner in which this metaphor is conveyed is by the use of remixes, which are essentially permutations of an original musical idea. Continuum is as much a WebVR experience as it is a conceptual album, the experience adopts the formal structure of a binary tree forcing the listener to choose between two possible scenes at each layer. The project is part of an exploration into open-source tools which can be exploited for the dissemination of spatial audio.

Author Keywords

NIME, proceedings, LaTeX, template

CCS Concepts

•**Applied computing** → **Sound and music computing;** Performing arts; •**Information systems** → *Music retrieval;*

1. INTRODUCTION

WebXR is a powerful medium for the dissemination of spatial music as the experiences can be accessed not only on mobile devices but also on desktop computers or head-mounted displays (HMDs). In contrast to native mobile applications, which can be developed using an operating system (OS) specific software development kits (SDKs), WebXR experiences generally do not require one to own a smartphone device or HMD. In addition, these cross-platform applications do not need to be ported or re-compiled for multiple operating systems or devices but are instead designed to work seamlessly across hardware platforms.

Spatial audio has become an increasingly popular subject in new media in the last decade, with major investments by companies such as Meta and Google. Unfortunately, this compositional dimension of music-making is not equally accessible to everyone, and, to some extent, it appears that merit is attributed to artists for ownership of resources and the quality of their tools. Our research involves investigating how spatial-audio compositions can be created, captured, and disseminated using low-cost and open-source technologies.

Continuum is one of the projects we created which relies on open-source JavaScript (JS) libraries to accomplish this task. In the development of the work, we also learned about many of the possible drawbacks of working with WebXR frameworks over mobile spatial audio SDKs or game engines for developing on HMDs. This paper will outline the development of our composition and address some of the strengths and deficits of WebXR technologies. We will also discuss other WebXR projects which have made use of spatial sound, focusing on those featuring open-source and low-cost technologies. Lastly, we will address the ethical implications of working in this domain, and what criticisms one might level at a project of this nature.

2. RELATED WORK

McArthur et al. [6], in their 2021 paper, present a survey of 3D audio via the browser which covers the affordances and limitations of several libraries for spatializing audio on the web, and provide some useful guidance for newcomers to the domain. McArthur discusses several libraries that might be of relevance to the reader such as the: Web Audio API (WAA), JSAmbisonics library [7], Resonance Audio SDK[1] [3], HOAST (Higher-Order Ambisonics STreaming) platform [2], and many others. The purpose of the survey is not to categorize any one system as inherently superior, nor is there an evaluation of these, rather it hopes to delineate the strengths and weaknesses of each, allowing the reader to make informed decisions. In the concluding remarks, the authors note some general principles to follow when developing for the web[2]:

- Keep media files as small as possible.
- Develop for a primary browser.
- Consider the size of your server if live streaming.
- Invest in a test set-up to simulate actual use.

In our own projects concerning WebXR, we found these principles to be acutely pertinent. In Continuum, we opt for the Resonance Audio SDK for spatialization of point sources via an A-Frame plug-in[3]. In this process, each sound source was encoded as individual mp3 files in order to reduce the load time of the scene and we set a limit of eight, on the number of stems each scene could have. We also decided to confine our models to glTF files, specifically those designed to have a low-polygon count. Much like the authors of the aforementioned survey, we tested our experience on several

[1]`https://resonance-audio.github.io/resonance-audio/`
[2]The list was redacted and paraphrased.
[3]`https://github.com/mkungla/aframe-resonance-audio-component`

NIME'23, May 31st-June 2nd 2023, Mexico City.

different browsers and found that the behavior of the scene was not consistent across browsers[4]. While it is also important for us to test the behavior of the experience on various HMDs, it was more important that it worked on mobile devices, as we believe smartphones are more accessible to a global audience.

Another publication by Çakmak and Hamilton [1] describes the development of a VR experience pertinent to our discussion. In the abstract of the paper, the authors note one important deficit of WebXR technologies, which is the unfortunate reality that developing technologies can become deprecated or obsolete during the period of making the work. While we might be able to control for versions of libraries such as Resonance or A-Frame, browser changes cannot be predicted and are not always backward compatible. The only solution to this, we believe, is to not only specify in the documentation the browser that was used but the version as well. The paper provides a very reliable framework for creating multi-media WebXR experiences, however, many of the tools relied upon are not open-source, and some may be financially burdensome. The introduction to the paper also provides a quick, albeit important, glance at composers who made important contributions to the field of spatial music.

Jot et al. [4] discuss in their publication an approach for six-degree-of-freedom (6Dof) object-based interactive audio. This framework considers optimization while addressing the: position and orientation of the source and listener, velocity vectors contributing to Doppler shifts, distance-based attenuation, source directivity models, and efficient reverberation, which can contribute to localization based on early reflections. 6DoF corresponds to experiences in which listeners can move laterally in addition to the 3DoF provided by yaw, pitch, and roll (e.g., head rotation). Ambisonics has become a simple way to create spatial sound in 3DoF as linear rotation matrices exist allowing the basis functions upon which this system relies to be non-destructively transformed. The resulting virtual loudspeaker signals, after decoding, are filtered using binaural impulse responses, which contribute spectral and timing cues allowing us to localize sounds.

In Continuum, the ability to move laterally is moot, as we target mobile devices which can experience the work via a browser and cardboard HMD. This method was chosen due to its low-cost and open-source nature. Cardboard experiences are generally only 3DoF, as the interactive components of the system are not ergonomically suited for 6DoF. Navigation in 6DoF is possible in the desktop version, however, the immersion level is lower in this configuration. Navigation is also possible from one scene to the next. Continuum counts currently with seven scenes which can be navigated using a fuse-based cursor, implemented in A-Frame. The cursor works by casting a ray from the camera look direction and calculating intersection points with other objects in the scene, looking at the red cube in the scene for a few seconds begins audio playback[5], and looking at one of the circles in the scene, allows us to navigate to the next song[6]. See Figure 1 for a visual representation of the scene.

Simiscuka et al. [8] also provide an important contribution to the topic of WebXR and spatial audio. The authors here conducted a Quality of Experience (QoE) evaluation using various audio quality levels to determine the merit of ambisonic audio in an artistic context. Specifically, the

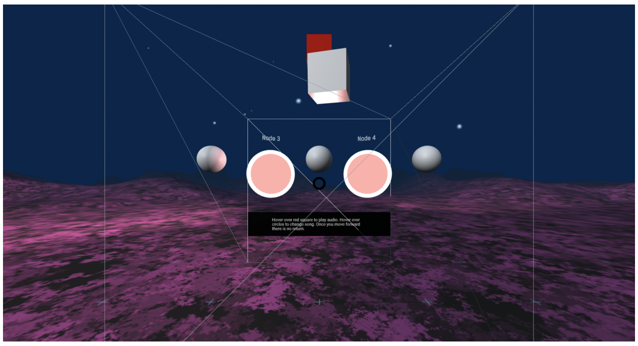

Figure 1: A sample image from Continuum depicting Scene 1.

authors developed a 360-degree video player which was evaluated by 24 participants from various countries. The study showed that: "more participants agreed or strongly agreed that the clip with ambisonic audio improved the immersiveness of the experience" - when compared to stereo or mono. Enjoyment and quality were also rated as higher with footage that featured stereo or ambisonics. In every question but one, ambisonic performed better than stereo[7]. This validates the use of ambisonic libraries in WebXR experiences, as quality, engagement, enjoyment, and, various other attributes, were all improved by the use of the technology.

3. CONTINUUM

The name Continuum comes from the ever-expansive nature of the universe, and the notion that time has no end or beginning in a world where infinite dimensions exist simultaneously, as has been proposed by numerous theorems. The piece is part of a larger initiative at our University called (Y) which defines itself as a creative community seeking to dissolve social barriers and create bridges between faculty, graduate and undergraduate students, staff, and alumni. During the COVID 19 pandemic, and in the aftermath of it, our academic institution remained partially shut down, which was part of the rationale to engage with a WebXR project.

The idea for the piece came about over multiple meetings with undergraduate students, who volunteered their time to this endeavor, and the author, who has been directing the group since its inception. Many of the artistic and technical decisions resulted from the facilities and competencies of the author, who had previous experience with A-Frame[8] (e.g., the WebXR library chosen for this project) and is an amateur guitarist. The decision to use Resonance as the method for spatialization came from the author's larger research direction, which investigates how free and open-source software (FOSS) and low-cost hardware can be leveraged to disseminate, record, and synthesize spatial music. In this process we had already developed artistic projects with Omnitone and the WAA, thus it seemed logical to explore Resonance as an alternative tool.

The process we followed was rather simple; the first step was to create three original compositions, one for each layer of our binary tree. In total there would end up being seven total scenes. Each of the nodes to the right of the one preceding it modulates to a musical key a fifth above, creating musical flow, while the nodes to the left would modulate

[4]We recommend Firefox for viewing Continuum.
[5]This interaction is required for the AudioContext() in the WAA to allow audio playback.
[6]Using the `link-controls` A-Frame component.
[7]The question being: "I believe that the immersive experience is comparable to a live opera".
[8]`https://aframe.io/`

to a fifth below. In order to define the color palette of each scene, we used the visible light spectrum as our template and decided to use longer wavelengths for nodes to the right, and shorter ones for nodes on the left[9]. The nodes at the top of the tree are also supposed to be darker while those lower are brighter. An example of a binary tree is depicted in Figure 2.

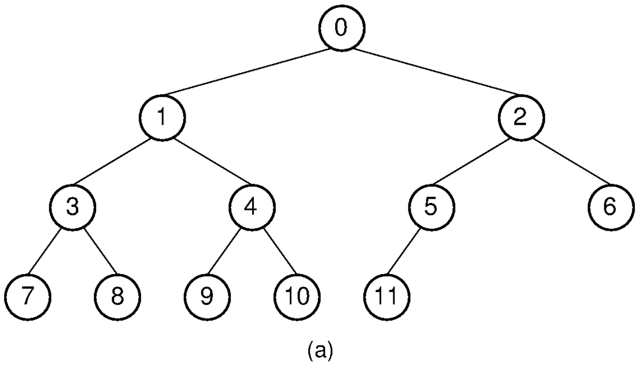

Figure 2: An example of a binary tree (data structure).

Each of the three compositions used for the remixes which populated each layer was composed in Ardour and then stems were exported for each of our collaborators to manipulate. These were transposed according to the requirements of the project by each team member, most of whom selected Ableton[10] as their tool for the remixing process. Each of these sessions subsequently was condensed into at most eight mp3 files[11] which were assigned to eight sound objects in the A-Frame scene. Special care was taken to make sure each sound object was equidistant from the origin in order to preserve the balance of each song. We decided to use a small number of possible sonic elements to reduce loading times, as this is a pervasive problem with WebXR works.

In addition to the spatial audio present in each scene, we also wished to create some visual material to maintain the listener engaged. To this end, we adopted another A-Frame component capable of creating visualizations from the sound spectra of musical elements[12]. A secondary mix, called a full mix, was piped to these visualization objects which react to a rendering of the musical material containing all the elements of the track. Furthermore, we chose to include glTF models and simple animations, based on A-Frame primitives, in the scene, to further accentuate and stimulate the audience.

4. AFFORDANCES AND LIMITATIONS OF WEBXR

In contrast to native mobile XR applications built using iOS or Android XR SDKs, web applications require a stable and resilient network to operate smoothly. In contrast, standalone mobile applications, once packaged and downloaded, can be executed without having to rely on an internet connection or concerns about dropouts in connectivity. Another related format for XR experiences is the standalone application built using a game engine such as Unity or Unreal Engine. While these formats are generally more powerful, they also require specialized equipment and are considerably less accessible than mobile devices, which with the birth of the iPhone have become a ubiquitous technology around the world.

WebXR spatial audio technologies themselves, as noted by [6] each have their affordances and limitations. McArthur et al. note three important criteria for the adoption of these tools: usefulness, ease of use, and documentation. Usefulness is defined by the authors as the number of barriers that the tool overcomes, and the range of use cases. WebXR technologies intend to overcome a number of barriers which it seems in their current state they cannot fully satisfy. While the goal of these libraries is to create a system that can perform equally on mobile, desktop, and HMDs, the reality is that all these platforms yield vastly different results.

Using Continuum as a test case, we found that Safari, both on mobile and desktop did not behave as desired. Furthermore, in previous experiments with Omnitone[13], we found that iOS as a whole was unsuitable for distributing ambisonic audio via the browser. Likewise, with desktop browsers, our experience has been that security measures and API (e.g., application programming interface) implementations vary from browser to browser, making only some browsers compatible with the project. More so, these projects often require a large amount of Random Access Memory (RAM) and have distinct buffering mechanisms from project to project[14], making the performance heavily hardware dependent.

Binaural audio developments in the last decades have made it possible for people to experience spatial music with considerably less hardware in a personalized setting, however, the audio quality when compared to speaker arrays is substantially reduced. Many of the SDKs created for spatialization in Unity, JS, or Virtual Studio Technologies (VSTs) rely on generic head-related transfer functions (HRTFs) which may take the listener time to adapt to. While HRTFs are becoming easier to measure, there still does not seem to be wider agreement regarding how formats such as SOFA [5] can be incorporated into WebXR technologies.

Unfortunately in the last few years, support for mobile HMDs such as the Google Cardboard or Daydream has been waning, as there is a greater financial incentive for companies to sell specialized equipment than to sell peripherals that can turn smartphones into capable virtual reality hardware. For most consumers, VR is still a luxury rather than a tool for work, which means a very small percentage of the population has adopted this technology. While most WebXR libraries support HMDs such as Vive or Oculus, the majority of our audience we suspect will explore the work on a desktop computer, while a small number will experience it with a Cardboard viewer.

5. CONCLUSIONS

This paper has discussed the WebXR experience Continuum, a conceptual album based on the principle of the multiverse. The piece was created in A-Frame in collaboration with various undergraduate students at Y whom remixed various instrumental pieces to develop permutations of each universe or scene. The structure of the project is that of a binary tree, which effectively means different viewers will have different experiences based on their chosen path. This

[9]Using the application `https://colorswall.com/` to facilitate this process.
[10]`https://www.ableton.com/`
[11]From mix busses.
[12]`https://www.npmjs.com/package/aframe-audio-analyser`
[13]`https://github.com/GoogleChrome/omnitone`
[14]Some use adaptive streaming, while others download all assets prior to playback.

paper has also addressed some of the affordances and limitations of WebXR, as a tool for the dissemination of spatial audio. In Section 7 this paper will conclude by articulating certain ethical and philosophical implications of the work.

6. ACKNOWLEDGMENTS

I want to thank X, Y and Z. Left out for anonymity.

7. ETHICS STATEMENT

While this project makes a great effort to address some of the ethical concerns regarding XR development from a socio-economic perspective, the piece is not without its faults. One of the primary concerns with this type of art-making is that it helps us further dissociate from our environment, which makes it easier to ignore the catastrophic changes which are being felt globally. The only defense to this claim is that the beauty of the virtual environment, albeit synthetic, attempts to inspire a sense of wonder through music, which we hope can be put to positive use in the real world. Furthermore, the elements comprising this work are multi-functional and have a relatively small footprint (e.g., the materials are relatively environmentally friendly).

While we seek to create methods for developing spatial music with low-cost hardware, the reality is that many of the systems used to create this piece are still prohibitively expensive to large swaths of the global population. Graphical processing units (GPUs) are heavily relied upon to process all the animations and models which make these works captivating, and while these have been getting cheaper, they remain out of reach for many people still. Furthermore, while smartphones are fairly accessible devices, often these experiences require state-of-the-art models with fast processors and advanced connectivity hardware.

8. REFERENCES

[1] C. Çakmak and R. Hamilton. od: Composing spatial multimedia for the web. *Journal of the Audio Engineering Society*, 68(10):747–755, 2020.

[2] T. Deppisch, N. Meyer-Kahlen, B. Hofer, T. Latka, and T. Zernicki. Hoast: A higher-order ambisonics streaming platform. In *Audio Engineering Society Convention 148*. Audio Engineering Society, 2020.

[3] M. Gorzel, A. Allen, I. Kelly, J. Kammerl, A. Gungormusler, H. Yeh, and F. Boland. Efficient encoding and decoding of binaural sound with resonance audio. In *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*. Audio Engineering Society, 2019.

[4] J.-M. Jot, R. Audfray, M. Hertensteiner, and B. Schmidt. Rendering spatial sound for interoperable experiences in the audio metaverse. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pages 1–15. IEEE, 2021.

[5] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig. Spatially oriented format for acoustics 2.1: Introduction and recent advances. *Journal of the Audio Engineering Society*, 70(7/8):565–584, 2022.

[6] A. McArthur, C. van Tonder, L. Gaston-Bird, and A. Knight-Hill. A survey of 3d audio through the browser: practitioner perspectives. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pages 1–10. IEEE, 2021.

[7] A. Politis and D. Poirier-Quinot. Jsambisonics: A web audio library for interactive spatial sound processing on the web. In *Interactive Audio Systems Symposium*, 2016.

[8] A. A. Simiscuka, M. A. Togou, R. Verma, M. Zorrilla, N. E. O'Connor, and G.-M. Muntean. An evaluation of 360° video and audio quality in an artistic-oriented platform. In *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pages 1–5. IEEE, 2022.