# Spatial Audio Designer

Nicholas F. Polys
Virginia Tech
Blacksburg, USA
npolys@vt.edu

Noam Bendelac
Virginia Tech
Blacksburg, USA
noamb@vt.edu

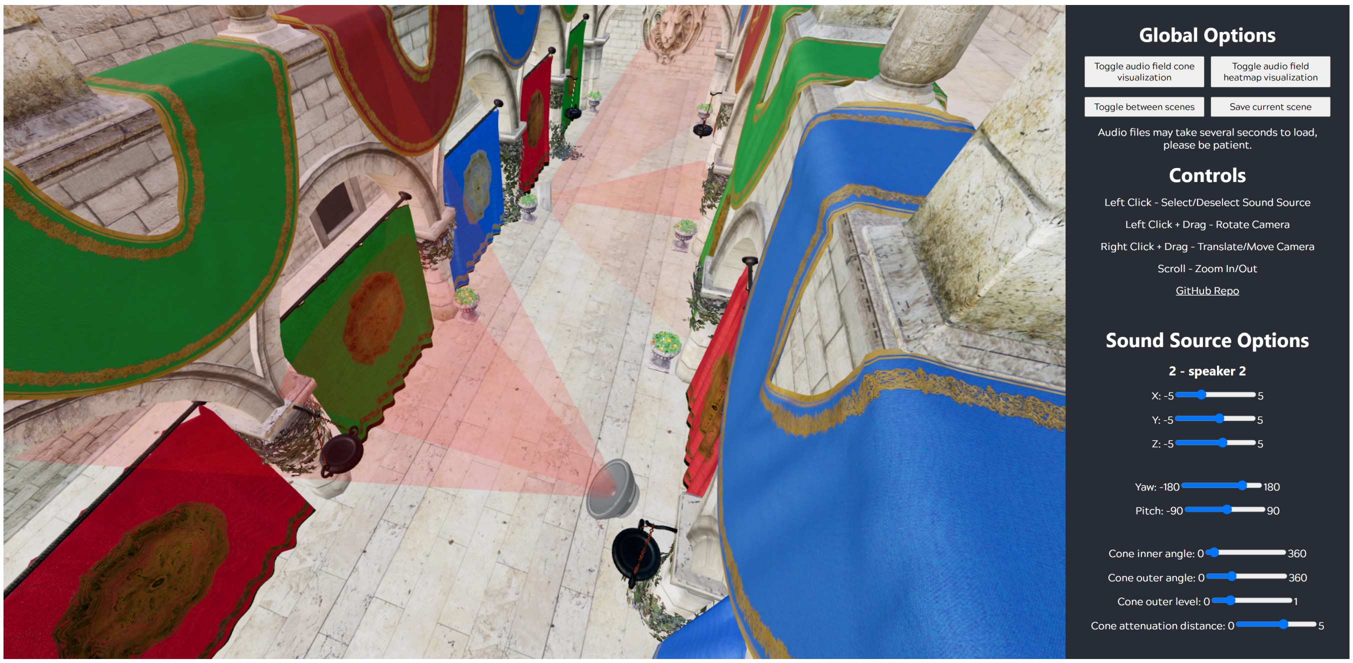


**Figure 1: Web3D prototype with sound cone representation.**

## ABSTRACT

The Web Audio API is an underutilized technology that provides a potential for rich interactive control over sound generation and rendering. Our team made use of the API in combination with Web3D technologies to create a spatial audio design tool for digital audiovisual creators. Our primary design challenge was creating an interface for visualizing and manipulating sound design in 3D space. We wanted our interface to be learnable and usable for our target user groups: digital music creators, digital audiovisual 3D artists, and physical audiovisual installation artists who wish to develop ideas in a virtual space. From user interviews, we learned that users needed a detailed visual 3D space as a starting point to populate with sound, as well as fine control over positioning of sound sources. The prototype web app can be used by digital and physical artists to create novel virtual audiovisual experiences, or to model a physical audiovisual installation to share and test with others. More work needs to be done to add direct spatial controls for sound fields and to make the app more natural to use. We asked artists of varying technical skill to use the app and re-create a reference scene, and measured how accurate their re-creation is.

## CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools*; • **Applied computing** → *Media arts*.

## KEYWORDS

sound design, 3D graphics, HTML5, Web Audio

**ACM Reference Format:**
Nicholas F. Polys and Noam Bendelac. 2022. Spatial Audio Designer. In *The 27th International Conference on 3D Web Technology (Web3D '22), November 2–4, 2022, Evry-Courcouronnes, France.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3564533.3564563

## 1 INTRODUCTION

This Senior HCI Capstone project examines how multimedia information can be spatialized in Web3D, and how user interaction can effectively manipulate and change properties of the environment or the information (in our case W3C WebAudio sources). Our goal is to create a spatialized audio design tool for digital audiovisual


*Web3D '22, November 2–4, 2022, Evry-Courcouronnes, France*

ACM ISBN 978-1-4503-9914-2/22/11.
https://doi.org/10.1145/3564533.3564563

3D artists and physical audiovisual installation artists to develop their ideas in virtual space and to visualize and manipulate areas where multiple sound sources are audible. We interviewed digital and physical artists at Virginia Tech to better understand how to support user creation of detailed 3D spaces populated with sounds and provide users with fine-grained control over sound source placement and sound properties.

## 2 BACKGROUND

### 2.1 Sound Design

Sound experience designers face new challenges and opportunities when faced with the medium of virtual environments. As simulacra, such environments can only provide proxy models for reality as we can describe it. Applying design methods for traditional physical spaces can be productive, although designers must grapple with new digital tools as well as the facts of delivery, including audio spatialization and headphone clients [Candusso 2015]. As digital platforms have matured, designers need interactive user interfaces to manage sounds in space, including multichannel [Austin-Stewart and Johnson 2022] and ambisonic parameters [Melchior et al. 2009]. While original spatial sound design software was 2D, clearly new tools will need to represent the 3D spatial configuration of loudspeaker arrays [Ledoux and Normandeau 2018]. Future sound simulations platforms will have impact not only in our artistic environments, but also our every-day urban environments [Hong et al. 2017].

### 2.2 Web Audio API

This prototype is a web application that utilizes WebAudio and Web3D technologies. The Web Audio API, which is provided by W3C, contains a multitude of features, the primary one we utilize being spatialized audio. The core elements used to spatialize audio are the "PannerNode" and "AudioListener." The PannerNode represents a sound source with a position, a direction, an attenuation distance factor, and a level of directionality controlled by an "inner angle" and an "outer angle" that define cones of influence. The AudioListener represents where sound is 'measured', like a microphone, to then play to the user. The browser utilizes a "Panning Algorithm" to calculate the sound transformations needed to achieve this effect [X3D 2022].

### 2.3 Sound Parameter Manipulation and Rendering

3D User Interfaces are concerned with the support and mapping of features, such as navigation, selection, manipulation, and system control [LaViola Jr et al. 2017]. In the case of a desktop virtual environment, there are several design challenges to map user agency from 2D desktop screens to 3D virtual environments [Polys and Bowman 2004]. Indeed, human-subjects studies show that the screen size changes the value of information and interactions in Object Space or Display Space [Polys et al. 2011], [Polys et al. 2005]. In this project, we render the sound effects in Object Space (with two different techniques) and the parameter control in Display Space.

For this prototype, we considered how the 6 Degree-Of-Freedom (DOF) spatial and audio parameters would be mapped for both control and rendering. 3D manipulation widgets [Houde 1992] [Conner et al. 1992] were considered as a design target (Figure 2); however in our case, we separated the DOF as 1D slider widgets [Chen et al. 1988]. This is a choice for accuracy over speed, but also has implications for immersive interaction, where DOFs maybe coupled [Apostolellis et al. 2014].

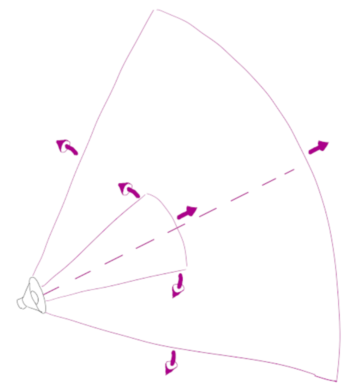

**Figure 2: Design sketch of situated gizmos in object space controlling audio parameters.**

Finally, rendering the influence of sound sources into an environment is a manifestation of a common problem in visualization (making the invisible visible): that is, representing multiple sources' contributions to a field. This is a common vis pattern: not only for sound, but also for other sorts of waves and pressure fronts that may interact with each other on the geometry of the environment. Other applications include radio-frequency propagation, line-of-sight calculation, solar exposure, and blast effects. In the case of sound fields, particle systems have been used with good results to portray sound propagation [fonseca 2015]. Here we compare a representation of the spatial sound sources using geometric shapes (cones) to one of a heatmap metaphor that paints the existing surfaces of the environment.

## 3 PROTOTYPE REQUIREMENTS AND DESIGN

We researched existing systems and then interviewed six practitioners and educators who do spatial audio design. We derived and prioritized a set of features that we felt our system should support:

- The user must be able to control the movement of their virtual camera and audio listener in the 3D space
- The user must be able to hear the spatialized audio change relative to their position in real-time.
- The user must be able to place and remove sound sources in the scene
- The user must be able to translate, rotate, and change the shape of the spatial sound field of sound sources
- The application should make it easy for non-technical artists to learn the tool and share their results

The prototype was built with React, three.js, and TypeScript. Our design had to provide visual feedback as to the directionality (omnidirectional vs. unidirectional) and reach of spatial audio sources in the 3D scene, and to that end we implemented two alternate views: one that shows the sound field cones as their own geometry, and

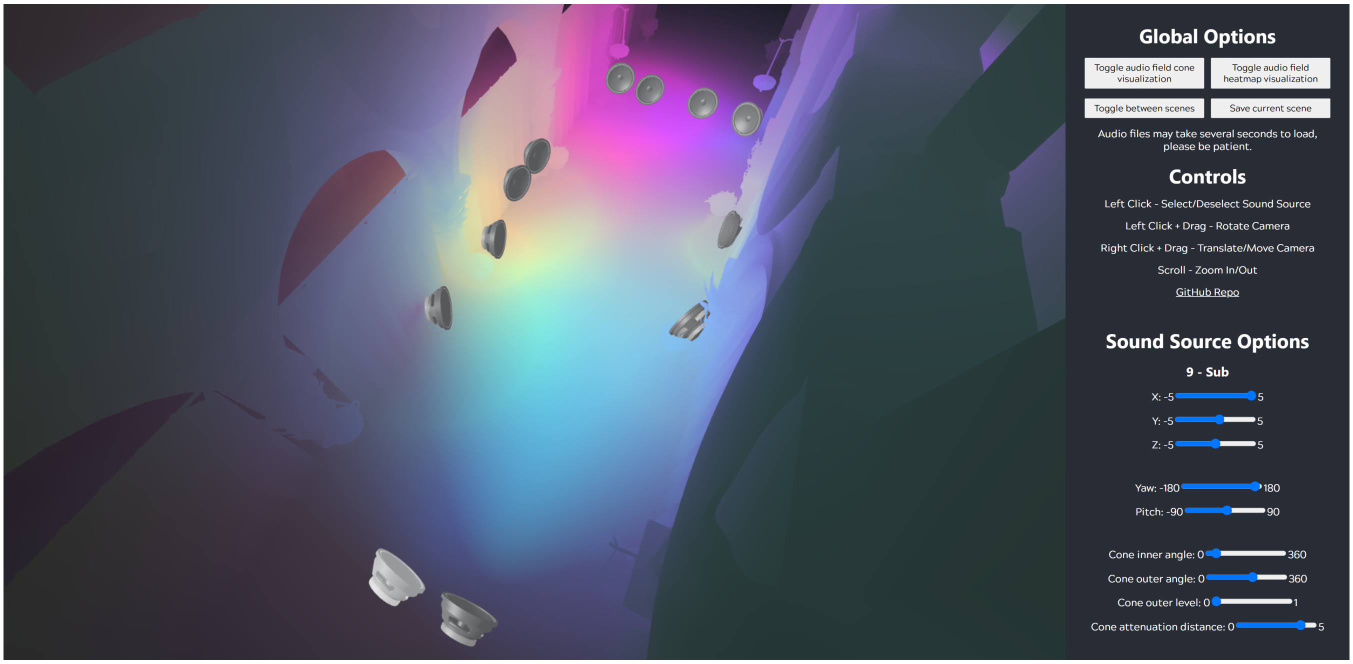


**Figure 3: Interactive 3D scene populated by sound sources and painted with a heatmap of each sound's amplitude.**

one that shows the sounds' cumulative projection onto geometry in the scene, mimicking a "Heatmap" (Figures 1 and 3, respectively).

The Heatmap visualization was implemented using a custom shader that was set as a material on all existing scene geometry. For every point in space (on the surface of the geometry), the shader calculates the color-coded contribution of each sound source if the AudioListener were at that point, using the same math that the WebAudio API uses to compute the audio output. In the below pseudocode, `theta` is the angle between the sound source's central axis and the point in space. `angularCoeff` is set to 1 when inside the inner angle, `coneOuterGain` when outside the outer angle, and linearly interpolated in between.

```
distanceCoeff = refDist / (refDist + rolloffFactor *
 (max(distToSource, refDist) - refDist));
angularCoeff = mix(1.0, coneOuterGain,
clamp( ilerp(innerAngle, outerAngle, theta), 0.0, 1.0 ));
volume = distanceCoeff * angularCoeff;
```

# 4 EVALUATION

## 4.1 Quantitative

We asked experienced 3D digital audio artists to work with our prototype. In order to assess the interface quantitatively, we created a reference scene and asked the users match the audio properties of the design with their configuration (similar to [Apostolellis et al. 2014]. We presented two audio alignment tasks and recorded the resulting spatialization properties the artists entered. Accuracy was calculated as a percentage of the range of possible values for each property: translation along the 3 axes from -5 to 5; rotation in yaw from -180°to 180°and pitch from -90°to 90°; and cone angles from 0°to 360°. The measured results are listed in Table 1 below. While sparse,

| | User 1 | User 2 | User 3 |
|---|---|---|---|
| Test Accuracy (percentage): | 91.07 | 97.27 | 91.1 |
| Time Taken (sec): | 7:28 | 5:34 | 6:07 |

**Table 1: Initial Evaluation Results**

these results are a good indication that the interface supports the task consistently across expert users.

## 4.2 Qualitative

These are our domain experts' thoughts on aspects of the prototype:

- Cones visualization: The Cones visualization is clear for 2 sound sources, but gets crowded for more, e.g. 13. It shows angles clearly, but does not show outer cone level/gain very well. There can be artefacts with the cones visualization when it comes to transparency sorting of the cones and the cone angles.
- Heatmap visualization: The Heatmap visualization is helpful for differentiating sound sources by color, seeing relative loudness between sources and varying with distance, and seeing the outer level slider setting. It needs an explanation to users of what it's showing. It is aesthetically pleasing, but only proved useful to some users. It could benefit from showing a gauge of sound sources' levels at the camera's position, not just at every point like the heatmap.
- Control sliders: The sliders are effective but may be overly sensitive, or it's too easy to move something too far off screen. Being able to see slider values without having to hover would help. Gizmos could be nicer than/in addition to sliders, especially for users used to other 3D software.

- Environment control: Camera controls are intuitive for some but need explanation for others. It would help if the X, Y, and Z directions were shown with an indicator. Having a scene/building is helpful, but the chosen scene does not reflect how the sound isn't attenuated by walls in the WebAudio API.

# 5 FUTURE WORK

One feature that would help audiovisual artists specifically is "Listener Mode," where a user can walk through a scene from the context of the listener without editing capabilities, mimicking a real life scenario. The inclusion of this option would break the tool into two major portions: the aforementioned Listener Mode and the Editor Mode. Editor Mode would use a WASD and mouse system for user movement along with the space bar and shift key for flying around the scene. Listener Mode only use the WASD and mouse system to limit the user to stay at eye-level.

Another major feature would be the ability to both import audiovisual scenes and audio files as well as export the created scene. We currently have the ability to export the spatialization settings of the sounds in the scene as a JSON file, but we want to eventually be able to export the entire scene and sounds, and import such a JSON file back into the app. Having this would allow users to save an entire scene to use later or to share with another person. This improved save functionality would provide increased creative flexibility and encourage cooperation for the user base. It may also be useful to have GLTF and/or X3D export options.

The last future feature to highlight is the use of "gizmos" situated in 3D object space around sound sources to control their positioning and spatial sound field shape, as sketched in Figure 2. There exist gizmos like this for translation and rotation in existing 3D applications, but gizmos for controlling cone angles and attenuation distance factor would be novel and allow more intuitive control over the sound design.

# ACKNOWLEDGMENTS

Thanks to Senior HCI Capstone team members, Sarah Jane Lynch and Bob Monroe, who contributed to the design and development of the system.

# REFERENCES

2022. X3D design criteria and specification publication. (2022). https://www.web3d.org/x3d4

Panagiotis Apostolellis, Brennon Bortz, Mi Peng, Nicholas Polys, and Andy Hoegh. 2014. Poster: Exploring the integrality and separability of the Leap Motion Controller for direct manipulation 3D interaction. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 153–154.

Jesse Austin-Stewart and Bridget Johnson. 2022. The Multichannel Monophonic Simulation Tool. *Chroma: Journal of the Australasian Computer Music Association* 38, 1 (2022).

Damian Candusso. 2015. Designing Spatial Sound: Adapting Contemporary Screen Sound Design Practices for Virtual Reality. In *SMPTE17: Embracing Connective Media*. 1–10. https://doi.org/10.5594/M001749

Michael Chen, S Joy Mountford, and Abigail Sellen. 1988. A study in interactive 3-D rotation using 2-D control devices. In *Proceedings of the 15th annual conference on Computer graphics and interactive techniques*. 121–129.

Brookshire D Conner, Scott S Snibbe, Kenneth P Herndon, Daniel C Robbins, Robert C Zeleznik, and Andries Van Dam. 1992. Three-dimensional widgets. In *Proceedings of the 1992 symposium on Interactive 3D graphics*. 183–188.

nuno fonseca. 2015. immersive sound design using particle systems. *journal of the audio engineering society* (may 2015).

Joo Young Hong, Jianjun He, Bhan Lam, Rishabh Gupta, and Woon-Seng Gan. 2017. Spatial Audio for Soundscape Design: Recording and Reproduction. *Applied Sciences* 7, 6 (2017). https://doi.org/10.3390/app7060627

Stephanie Houde. 1992. Iterative design of an interface for easy 3-D direct manipulation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 135–142.

Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

D. Ledoux and R. Normandeau. 2018. An Immersive Approach to 3D-Spatialized Music Composition: Tools and Pilot Survey. In *Proceedings of the Audio Mostly 2018 on Sound in Immersion and Emotion (AM'18)*. Association for Computing Machinery, New York, NY, USA, Article 10, 4 pages. https://doi.org/10.1145/3243274.3243300

Frank Melchior, Andreas Gräfe, and Andreas Partzsch. 2009. Spatial audio authoring for Ambisonics reproduction. In *Proc. of the Ambisonics Symposium*.

Nicholas F Polys and Doug A Bowman. 2004. Design and display of enhancing information in desktop information-rich virtual environments: challenges and techniques. *Virtual Reality* 8, 1 (2004), 41–54.

Nicholas F Polys, Doug A Bowman, and Chris North. 2011. The role of depth and gestalt cues in information-rich virtual environments. *International journal of human-computer studies* 69, 1-2 (2011), 30–51.

Nicholas F Polys, Seonho Kim, and Doug A Bowman. 2005. Effects of information layout, screen size, and field of view on user performance in information-rich virtual environments. In *Proceedings of the ACM symposium on Virtual reality software and technology*. 46–55.