An audiovisual web-based VR platform for fully immersive storytelling

Regis Rossi Alves Faria, Adiginton J. King Ferreira, Charles K. Neimog, Almir Almas
University of São Paulo, São Paulo, Brazil
Eric B. White
Oxford Brookes University, Oxford, United Kingdom

https://doi.org/10.1145/3822516.3837792

This work presents results of an ongoing project on applied computing in arts and humanities and immersive audiovisual web application development for storytelling.

Abstract

Based on the integration of a virtual storytelling application and a spatial audio platform, we conceived an interactive VR platform for allowing an easy-to-use experience of immersive 360-degree visuals and 3D audio directly through the web. We present the major concepts of the project, with a description of the system architecture and component technologies used. Examples of virtual storytelling creation and users' experience demonstrate the unique features and processes under investigation, emphasizing the potential of spatial audio on platforms in the XR space on the web. The article discusses aspects related to current implementation, user agency and usage potential, reporting on ongoing and future work.

Introducing VSAT-AUDIENCE

Modern storytelling uses interactivity and spatial immersion to engage audiences, shifting them from passive readers to active participants with agency to shape the narrative.

If you want to tell a story of your family, a trip you took, or a person's life journey illustrated with text, images and sounds, composing a narrative within an immersive environment in which your audience can navigate, see, listen, and read text, you will love VSAT-AUDIENCE: a fully immersive 3D web-based platform designed for storytelling, combining two digital engines:

- the VSAT (Virtual Storytelling Application Toolkit) as the base web platform for visual VR, enabling the creation and experience of immersive 360° visuals, and
- the AUDIENCE spatial audio engine for creating and rendering 3D sound.

A solution was devised to integrate the two engines using a third component as a bridge, the Pd4Web framework, conceived to deploy Pure Data (Pd) patches as interactive web applications to run in standard web browsers.

Software components

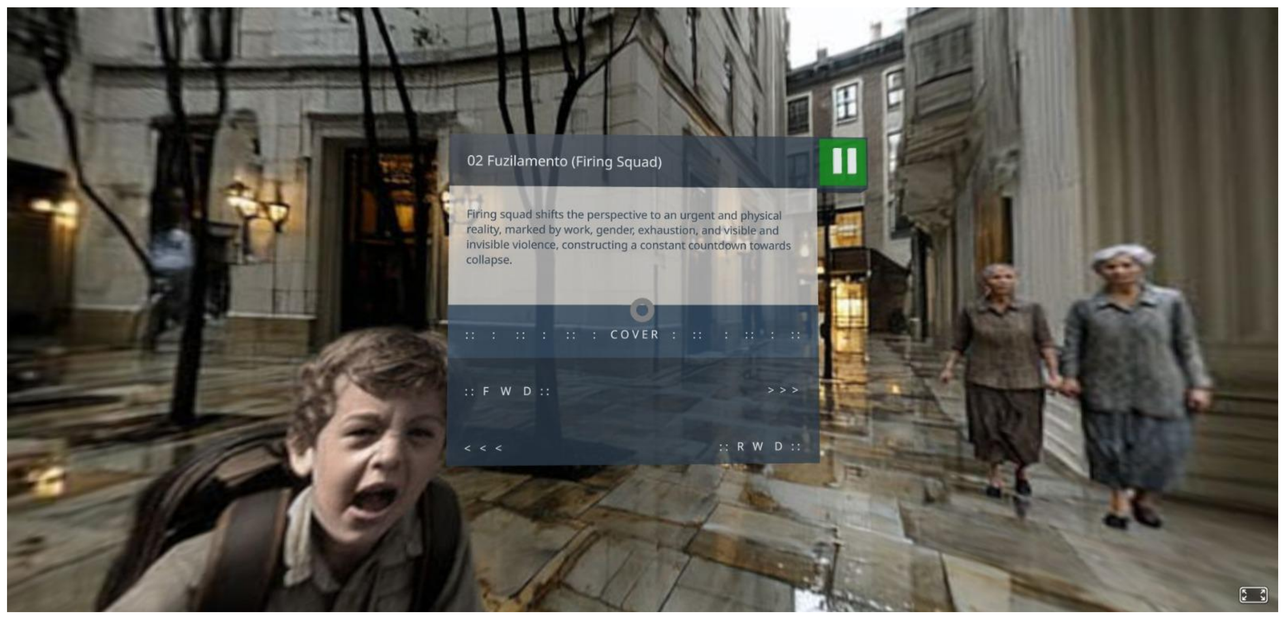
VSAT is a service-oriented solution created in the Oxford Brookes University using a web server to resolve clients' requests; a middleware solution for client page generation; a database and cloud-based solution for audio and image management. It has been developed in JavaScript/TypeScript and deployed using a cloud-based web service.

VSAT uses audio and 360° photographs to build a digital narrative, coding and production skills incidentally, as part of the learning process.

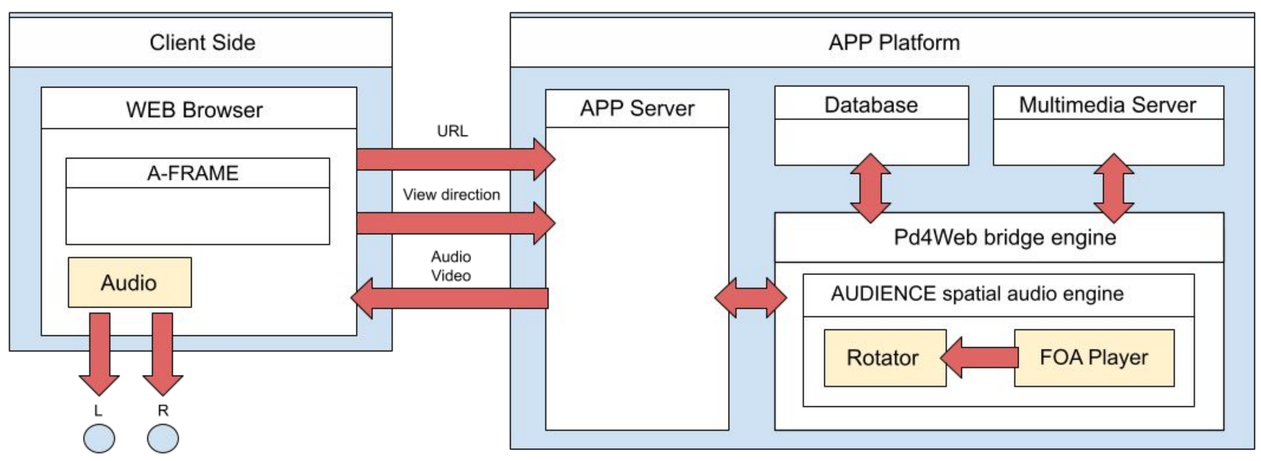
A VSAT experience is made up of a number of interconnected "scenes". A scene consists of:

- a 360° image and an audio clip,
- a "Passage" built in the form of a billboard-like text box with a "header",
- "Exit" buttons, used to navigate between scenes.

The "billboards" allow the user to move through the scenes by "exiting" the current space and proceeding to a new one.







The **AUDIENCE** spatial audio engine is built using patch-oriented abstractions and objects to run on the Pd platform, consisting of a library of components for building diverse spatial sound solutions. Since 2005 it has been the base for several applications from music and sonic immersive VR to 3D film soundtrack production.

The platform provides a full experience for describing 3D immersive sound scenes using spatial audio formats (e.g. Ambisonics) and rendering them to a given listening system (e.g. 2-channel earphones or a multichannel layout).

Pd4Web is a framework for deploying Pd patches as interactive web applications that run directly in standard browsers.

Given a Pd patch, it analyzes its dependencies, downloads and compiles the required external libraries, generates the initialization code for Pd external objects, and builds a JavaScript module that can be loaded and executed in a web browser.

Use cases

As an example of an interactive immersive story created using VSAT-AUDIENCE, **the figures show two neighboring screenshots of a 3D music scene** from the EP "*View from the Point*", conceived as a composition within a 360° photosphere.

Lyrics and poetic motives are exhibited in the billboard, which also provide the buttons to Play/Pause, Forward and Rewind to other song scenes or to the album cover scene.

The spatial audio provides a way to spatially connect sounds and images based on the movement of the spectator's gaze, creating motion trajectories for musical elements around the listener.

To create a story, users access an authoring module on the web. A number of VSAT stories authored by a variety of creators are already available on the web at https://vsatool-d2dd85389e42.herokuapp.com/story/.

Platform implementation and behavior

VSAT-AUDIENCE platform uses AUDIENCE functions to allow the reproduction of Ambisonics audio and the rotation of the sound field relative to the user's gaze.

The **block diagram shows its architecture**. To validate the integration of the spatial audio engine and the correctness of the auditory display, we compiled the AUDIENCE patch containing a spatial audio player and a scene rotator, and then performed a series of tests by running a test scene (encoded in an html file) on a local server.

The test procedure aimed to locally reproduce a story experience on a browser, which starts from a client call to a URL. The HTML file contains the initialization of the visual scene, rendered by A-Frame in the client side, and of the audio scene, rendered by the AUDIENCE player-rotator engine previously compiled by Pd4Web engine into an audioworklet.

The worklet also encapsulates the Pd4Web itself, that handles communication between A-Frame and the AUDIENCE patch, receiving the start signal to load the scene's Ambisonics sound file and the viewer's gaze vector.

As it plays, the AUDIENCE renders the story's Ambisonics (FOA) audio file and rotates the virtual sound field with 3 degrees of freedom (3DoF), according to the user's gaze direction (indicated at the center of the screenshot).

The auditory space of the sound scene is rotated with each update of the gaze vector. The rotated sound is then decoded for two-channel output on headphones.

The experience is best enjoyed when loaded onto a VR headset.

Conclusions and future work

The innovative contribution of this work lies in the integration of the aural and visual engines into a web platform, enabling users to create fully immersive audiovisual VR storytelling experiences with a low learning curve.

Its interactive features extends the possibilities with current 360° VR experience and represents a new generation of storytelling platforms targeting a further level of immersive quality and accuracy of depth.

At this stage, validation focused on verifying the feasibility of integrating the software components, and the correctness of the sound field rotation, ensuring the correspondence between the visual and auditory positions of the sources.

Next phase will deploy the VSAT-AUDIENCE repository and platform on the web and conduct user studies with story creation, an evaluation of system effectiveness, and performance analysis.

VSAT-AUDIENCE offers a quick and easy innovative way to tell stories in 3D, opening up the "black box" of coding in VR. Consistent with the user interest observed in published stories, we have reason to believe that, with it, the narrative can sustain affective agency, boosting the user's level of engagement and interest.

Acknowledgments

We thank partial financial support from the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) and the São Paulo Research Foundation (FAPESP).

Web3D 2026
13-15 October 2026 – Doha, Qatar

University of São Paulo
Laboratory of Audio and Music Technologies
LabArteMídia
OXFORD BROOKES UNIVERSITY

to know more