Virtual Reality and Audiovisual Experience in the AudioVirtualizer

Adinda van 't Klooster[1*] & Nick Collins[2]

[1]Durham University (2019) and independent artist, UK
[2] Durham University Music Department, UK

Abstract

INTRODUCTION: Virtual Reality (VR) provides new possibilities for interaction, immersiveness, and audiovisual combination, potentially facilitating novel aesthetic experiences.

OBJECTIVES: In this project we created a VR AudioVirtualizer able to generate graphics in response to any sound input in a visual style similar to a body of drawings by the first author.

METHODS: In order to be able to make the system able to respond to any given musical input we developed a Unity plugin that employs real-time machine listening on low level and medium-level audio features. The VR deployment utilized SteamVR to allow the use of HTC Vive Pro and Oculus Rift headsets.

RESULTS: We presented our system to a small audience at PROTO in Gateshead in September 2019 and observed people's preferred ways of interacting with the system. Although our system can respond to any sound input, for ease of interaction we chose four previously created audio compositions by the authors of this paper and microphone input as a restricted set of sound input options for the user to explore.

CONCLUSION: We found that people's previous experience with VR or gaming influenced how much interaction they used in the system. Whilst it was possible to navigate within the scenes and jump to different scenes by selecting a 3D sculpture in the scene, people with no previous VR or gaming experience often preferred to just let the visuals surprise them. They used mainly head movement to change their point of view, whereas people with previous VR or gaming experience quickly learned inherent capabilities of the system. The AudioVirtualizer system is most suited to respond to electronic soundscapes and microphone input.

---

* Corresponding author. Email: adinda_18@hotmail.com

1. Introduction

Virtual Reality (VR) provides exciting new opportunities for highly immersive audiovisual experiences in alternative worlds. Virtual Reality artworks have been developed since the first wave of 1990s VR such as fully immersive pieces like *Osmose* and *Ephémère* by Char Davies. Although Davies used a Head Mounted Display (HMD) for showing the graphics instead of the more elaborate CAVE system available at the time, the three Silicon Graphics workstations used to run *Osmose* still cost more than a million U.S. dollars in 1995 [8] and all the software, which responds to biosignal input data from a vest the user wears, was written entirely bespoke by John Harrison. Another early example of VR art is *Memory Theater VR* (1997) by Agnes Hegedüs, which consists of a cylindrical space constructed out of wooden panels. At its center is a plexiglas model of the same circular construction in which a 3D mouse is moved to generate imagery that is projected back out onto the surroundings [9].

VR art installations have also been developed for a variety of different CAVE systems, which usually are five walls plus floor projected virtual reality rooms that visitors walk into. All participants wear active stereo glasses for their 3D experience and one person wears a tracker to allow the graphics to adapt according to his/her point of view [4]. The method of interaction can vary between different CAVE versions; for example, the StarCAVE provided interaction capabilities through multiple camera tracking and a wand [5]. Artworks developed for CAVE systems include *Face a Face* by Catherine Ikam (2000), and *World Skin* (1998) by Maurice Benayoun. In the latter the CAVE is used to show a world of war, which gradually disappears: as visitors take pictures of that part of the photographed image the projection is erased and replaced by a black silhouette [2]. A piece for networked CAVEs is *Traces* (1999) by Simon Penny, where the participants can see the users at the two other CAVE sites via blocky shadow images. Other networked CAVE pieces are *Beat Box* by Margaret Dolinski (2005) which facilitates collaborative music making and *PAAPAB* (2001) by Josephine Anstey, Dave Pape and Dan Neveu, encouraging participants to dance across different networked sites [6]. Although the potential of networking the different CAVE environments across the globe is an attractive attribute, downsides are lack of accessibility for most artists, the requirement to work with a team of people and the fact that you can still see the corners of the room, which breaks the sense of immersiveness.

A recent boom in graphics capability, hardware development and a merging with the game industry has made VR accessible to artists and musicians with much more moderate budgets. This technical revolution has given the current artist a range of HMDs to choose from, each providing full immersion and high-end graphics. Current headsets like the HTC Vive Pro and Oculus Rift have a higher resolution and frame rate than Android based devices, but do still depend on tethering and external base stations. Recent wireless headsets like the Oculus Quest feature inside out tracking from sensors on the headset itself, and function on their own without further playback equipment, making them suitable for art installations on a budget. Nonetheless, in our own exploration we found the HTC Vive Pro headset superior in terms of graphics and comfort, in particular for people wearing glasses.

Whilst virtual reality artworks remained fairly rare at the end of the twentieth century, recent hardware and software developments have caused something resembling a renaissance of VR art. Acute Art is (since 2019) a VR Art platform that makes VR artworks available for free through the STEAM VR online store. Curated by Daniel Birnbaum, previous Venice Biennale curator, Acute Art represents VR artworks and 360 degree videos by amongst others Marina Warner, Anish Kapoor, Louisa Clement, Christo, Nathalie Djurberg & Hans Berg and Jeff Koons. These highly immersive artworks can be experienced by any individual with access to a VR headset and workstation. Other examples of recent VR artworks include *Sentient Flux* (2013) by Nicola Plant and Alexander Adderly, a VR artwork that responds to users' body movement such as the waving of the arms,

MutatorVR (2017) led by artist William Latham and conceived of procedurally generated abstract sculptural shapes [12], and *Interference* (2019) by Julie Freeman where the user can locate the sounds of pulsars and interact with their magnetic fields to discover what the artist calls 'new dimensions'. Each dimension is inspired by the interaction of a pulsar with another celestial object, such as a black hole, a white dwarf, a planet or another pulsar, and consists of a graphic datascape and a navigable soundscape composed through the user's movement and gesture [7]. Other examples of artists and filmakers using VR include Michaela Pnacekova and Jamie Balliu (*A symphony of noise*, 2019), Kristina Buozyte (*A Trail of Angels*, 2018), Jakob Kudsk Steensen (*Re-animated*, 2018), Marc Zimmermann (*Conscious Existence*, 2018), Rachel Rossin and many more.

Software audiovisualizers, i.e., software that can generate graphics live in direct response to sound input, have existed for many decades, including the Atari Video Music console (1978) and early software prototypes in the 1980s demo scenes. Earlier artistic precedents include 1960s light shows, visual music, and colour organs dating back centuries [1]. The most famous software audiovisualizers include iTunes, MediaPlayer and WinAmp's in-built visualizers.

In VR audiovisualisers include *plane9* with 250 different scenes the user can mix, *FantaSynth*, *Chromestasia*, *Raybeam VR*, *CyberDreamVR* inspired by Rave music [15] and *Vision*, STEAM's own music visualizer. There are also musical rhythm games founded in pre-annotated music tracks or beat tracking technology such as *Beat Saber* and *Beat Blocks VR*. None however are responsive to live sound input as offered in the AudioVirtualizer.

2. Our design aims and considerations

VR is a medium being explored by filmmakers, artists and designers alike. The annual VRHAM international VR & Arts Festival in Hamburg is a good place to find out about the latest artistic creations in VR. Interestingly, the 2019 version of the festival showed mostly examples where interaction was kept at a minimum, often through head movement only, and only occasionally through controller or voice input. The VR experiences made available for free by Acute Art show a similar tendency to be 'observe and listen only' experiences. Their work by Christo for example is not much more than a 3D video of one of his large wrapped artworks. Navigation by the user isn't possible. In many ways it isn't a VR artwork, more 360 degree video documentation of an existing artwork. We wanted to use the new possibilities offered by VR such as its immersiveness *combined* with the ability to directly interact with the landscape/system. Virtual Reality artworks share the same art historical context as interactive artworks but the level of interaction varies strongly from one VR artwork to the next.

Using van 't Klooster's definition of interactive art: "An interactive art system uses technology to create a reciprocally active system between human(s) and machine(s)" [11] it

follows that more needs to happen than for the landscape, scene or film to become visible. Somehow further inputs such as sound/voice input or those provided by the controller should be utilised for a work to become interactive.

Our aim with the AudioVirtualizer was to create an audio visualiser in VR that would create visuals akin to the style of a body of drawings by the first author of this article and which would allow the user to navigate fully within the 3D scenes through a controller and head movement. To keep a sense of physical connection we avoided teleporting with the exception that selecting a particular 3D sculpture visible in each of the scenes allows the user to jump to the next scene. Having looked at previous musical visualisers that respond in real-time to any music input, such as iTunes, WinAmp and MilkDrop and the more varied but not free Resolume, we knew we wanted a system that would be less generic. The aforementioned softwares are built on a limited understanding of the sound input, and often utilise bandwise energy from FFT frequency analysis. We wanted to make a system that would be more flexible in the potential ways it could analyse sound, and introduced mid level audio feature extraction. We also wanted our system be adaptable to different performance and exhibition contexts, be available in VR and sharable with other artists/musicians. This led to the design of a Unity plugin for machine listening as described in the next section.

In Unity we then built a system based on three different scenes described in more detail in section 4. It gave us the opportunity to create a system more suited to respond to electronic soundscapes. The sustained pitch feature, sensory dissonance and spectral entropy features were particularly useful in this regard.

To achieve a sense of the aesthetic language of the first author of this paper, sections of 2D drawings were cut into pieces and assembled as skyboxes in Unity to create a sense of them existing in a 3D landscape rather than the 2D medium. These sources were further manipulated in Photoshop to allow the illusion of infinity without breaking up at the seams. 3D objects and particles were then added which move and change appearance based on changes in the audio features. In the first scene the textures mapped onto the spheres are also adapted drawings by the artist, dynamically swapped with other drawings depending on how the features changed throughout the sound track.

The aesthetic experiential aims of the AudioVirtualizer are multiple. First, there is the challenge for the user to figure out how changes in the sound manipulate the image, which encourages the user to listen more closely to the sound. Then there is the visual sensual quality of the work; the user can simply enjoy navigating these landscapes and changing their point of view based on their interactions with the interface. The almost entirely abstract world of combined Op Art like hand-drawn shapes, is punctuated by slightly more ambiguous 3D shapes with vaguely sexual connotations, such as the red dildo shape that is the portal to the other scenes, the almost vagina like shapes in the background of scene 1 (see figure 3), and the black and white bullets that look similar to the head of a penis (see figure 8). These associations can easily be missed when not familiar with the first author's previous works, but will be fairly obvious for those who do know that body of work. Their purpose is to provide an extra layer of meaning and the genderisation of the abstract language which has historically been claimed as predominantly male.

3. A Unity plug-in for musical machine listening

As well as Unity assets to support VR work, Unity can call out to C++ plug-ins from its native C# scripting. It is even possible to avoid the sometimes tricky compilation process of making libraries for multiple target operating systems (such as Android for Oculus Quest, or Mac and Windows standalone building), by using Unity's IL2CPP technology.

It is perfectly possible to build inter-connected systems combining Unity and standard computer music environments such as Pd or SuperCollider through Open Sound Control messages (see Fredrik Olofsson's audiovisual programming tutorials [14], the extOSC Unity asset, and OSC-XR [10]). However, such an approach makes standalone building more difficult or impossible, with an application bottleneck in network communication, and possible licensing issues. We preferred to create our own code library which could be compiled in to an application for distribution, and substantially eased the issue of prototyping onto Oculus Quest in particular, and compiling for Steam or the Oculus Store.

Feature extraction code was adapted from SuperCollider machine listening facilities originally written by the second author, alongside a few additional custom C++ implementations, so unencumbered by GNU GPL licensing. The signal analysis runs with 512 sample hopsize, corresponding (for 44.1KHz sample rate) to around 86Hz and thus near a typical 90Hz VR visual frame rate. Some feature extractors are optimal for a sample rate of 44.1KHz due to development decisions (e.g. a test corpus of 44.1KHz audio for optimization), though they still operate at 48KHz. This requires setting up audio interfaces carefully in some cases.

The list of features comprises spectral centroid, power, spectral irregularity, spectral entropy, sensory dissonance, key clarity, pitch detection via constant Q pitch detector, density of onsets (onsets detected per second), mean Inter Onset Interval (IOI), standard deviation of IOIs, beat histogram entropy, beat histogram first to second entry ratio, beat histogram diversity, beat histogram metricity, alongside flags for a detected onset and predicted beat location, and a 'continuous held pitch' derived feature which is larger the longer a consistently held frequency is detected by the pitch tracker. All feature values were normalized to the range [0,1] according to max-min normalization values found across a large corpus of electronic music. Such normalization assisted mapping experimentation in controlling graphical objects in scenes.

The plug-in actually makes available three independent feature extractors operating on mono audio. One is intended for a mono averaged mix from a stereo source, and two are for the left and the right channels. For the purposes of smooth animation, the audio frame rate is often too high. We introduced a basic IIR filter on all feature signals, with variable smoothing controlled by parameter $\alpha \in [0,1]$, to control this:

$$y = (\alpha * x) + ((1 - \alpha) * y)$$

x is raw feature output, and y is smoothed (previous y is used to update the next smoothed value, i.e., y[n-1] in digital filter parlance). Both raw feature output, and smoothed, were available for mapping.

In order to test out the Unity plug-in with stereo machine listening, a sample project was built utilizing various 3D graphics objects in Unity, controlled via audio analysis. An arbitrary stereo input audio file had left and right channels separately analysed, to control left and right placed objects in the game scene, as well as a further analysis of the mono mixture to control some central objects (Figure 1).
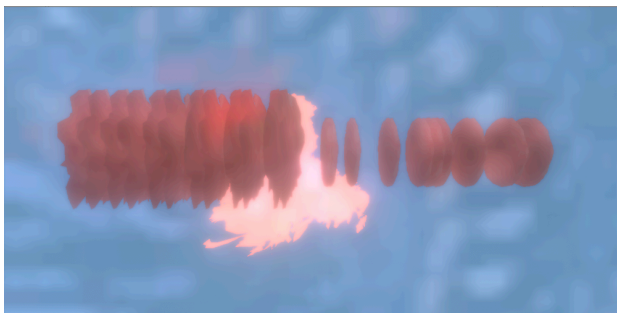

Figure 1: Stereo virtualizer demo, objects in the row on the left are controlled by analysis of the left audio channel, those on the right from the right channel, the central background shape from analysis of the mix

The Unity plug-in code is available at <https://github.com/sicklincoln/MusicalMachineListeningUnityPlugin> under a permissive BSD-3-Clause license.

4. The AudioVirtualizer Installation

Our AudioVirtualizer system was created for SteamVR and currently runs as a standalone application with the Vive Pro and Oculus Rift interfaces. These are tethered interfaces, so a VR ready PC is still a requirement for smooth running; we worked with an NVIDIA GForceGTX1070 graphics card and Intel Core i5 system.

Graphics are generated in response to changes in low and mid level sound features as well as already existing in the 'landscape' via skyboxes that contain drawings by the first author of this paper. Certain elements from the drawings are then repeated in objects that are generated in response to the sound and move through the landscape in different ways. The drawings have a sense of minimalism and are manipulated digitally to continue their endlessly

repeating patterns in the 3D space as seamlessly as possible. The user is able to see how the graphics respond to different sound sources through a built-in menu option that allows the user to choose one of four soundtracks by the authors of this paper or live microphone input.

Pressing the menu button on the controller overlays the scene with five sculptural objects, each of these lead to a different sound source when selected. Pressing the menu button again will remove the options and play the current selection. All graphics and system design are by Adinda van 't Klooster, machine listening tech and some Unity programming was by Nick Collins, with additional assistance with Unity VR programming provided by Nathan Flounders.

There are essentially three different scenes and an entrance space where navigation can be practiced. Using the large circular touch pad on the VIVE controller one can move forwards (up), to the right (right), the left (left) and backwards (down). Going up and down in the 3D space is achieved by pressing the trigger as well as keeping the finger on the up or down position of the circular dial. For non-gamers this takes some practice so this is done in an antechamber space without animated graphics (Figure 2). One can be teleported into the first proper scene by navigating to a red dildo-like sculpture and selecting it.

Figure 2. Entrance portal in AudioVirtualizer

The three main scenes each use a different drawing as a starting point that is built into a 3D skybox background in Unity. The first scene is constructed from two circles (one horizontal and one vertical) with planet-like spheres with drawings by the first author texture mapped onto them. Changes in audio signal power push the spheres around the loop and the intensity of this effect changes over the spheres and over time.
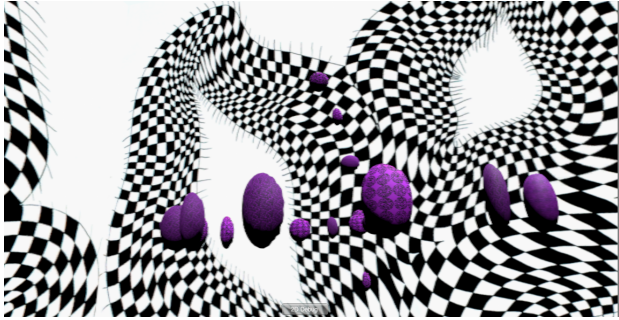
Figure 3. AudioVirtualizer scene 1a

Detected percussive onsets swap the visible texture on the outside of the sphere and the surface shader is perturbed by power, sensory dissonance (roughness of sound) and irregularity in the spectrum when the system detects an onset. Three black particles are also emitted when the system detects an onset. The size of these particles increase with power and the paths of the particles are determined by power and spectral irregularity.
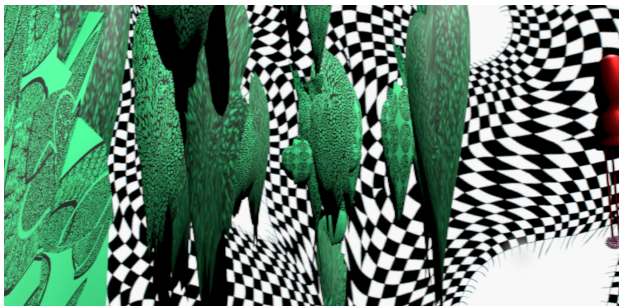

Figure 4. AudioVirtualizer scene 1b

When the music gets more frantic the spheres overlap and cause interesting patterns. At all times the user can navigate through the scene and get close-up to any particular area of interest. Teleportation to the next scene again happens by clicking the four-legged red dildo sculpture.

The second scene is built up from a skybox containing parts of the artwork *Each Egg a World* by the first author of this paper and a mountainous landscape.
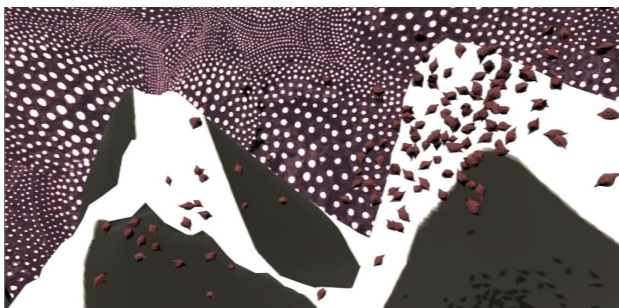

Figure 5. AudioVirtualizer scene 2a

At each detected beat, the system creates four new sculptural objects that look like something in between a beech nut and a surreal flower. These shapes are moved across the scene depending on power and brightness of the sound. If power is sufficient, the shapes will move towards a central point and a wobble is applied to their path depending on brightness of the sound and the density of onsets. This scene works particularly well in terms of responsiveness to sound when using microphone input, which can be selected by pressing the menu button, as vocal sounds have a lot of brightness in them and so this effect is particularly noticeable when used with the voice.
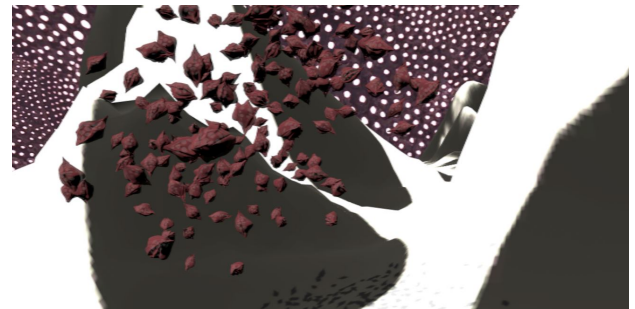

Figure 6. AudioVirtualizer scene 2b

The third scene is built up of a third Op Art like artwork. White smoke particles are emitted when the system detects sustained pitch, these particles increase with size as pitch is maintained and sensory dissonance increases, sometimes causing the whole scene to go white.

Figure 7. AudioVirtualizer scene 3a

Furthermore, when an onset is detected, ten sculptural shapes are generated starting as a circle and gradually falling apart depending on the brightness of the sound, sustained pitch and spectral irregularity (movement in X is controlled by brightness of the sound, movement in Y by sustained pitch and movement in Z by irregularity).

Figure 8. AudioVirtualizer scene 3b

A video of the in-the-system perspective of person navigating through the different levels can be found here:
https://www.youtube.com/watch?v=ITpmXXP_Ij8

5. Discussion

Evaluation of the developing system was carried out informally in a rapid design cycle between the authors. The art installation had an initial public testing at the new tech hub PROTO in Gateshead in October 2019, demonstrating versions for HTC Vive Pro and Oculus Rift. The primary finding was that those with previous gaming and VR experience were active navigators, used to and expectant of interactive control, also appreciating the painterly nature of the skybox backdrops as an artistic deviation from video gaming tropes. Those new to VR preferred a more hands-off immersion where they could dwell on looking around them, since this already provided a novel experience for them. As VR novices they preferred to fully take in the landscape rather than get side-tracked by learning how to navigate the system using the controls. Such expectations are therefore likely to change and adapt as VR becomes a more familiar and an available platform in the home.

In our own usage of the HTC Vive Pro versus the consumer Oculus Quest we found the Vive Pro device more comfortable, and less vertiginous, with higher graphics quality, than the Quest. Nonetheless, the inside out tracking of the Quest works very well, the controllers' use of a joystick rather than a 2-D pad felt more natural in navigation (giving some boon to the Oculus Rift version of the AudioVirtualizer), and there is no denying a greater market exists for the much cheaper Quest device.

As a further evaluative strand, Thor Magnusson's epistemic dimension space provides a useful tool to interrogate an interactive system as to its experiential affordances [13]. The eight spokes of a plot correspond to eight dimensions intended to tease out interactive capability. The tool is qualitative, and judgement of where exactly to place a given system in the space is a matter of comparison to the examples in Magnusson's paper and a relatively subjective interpretation of his descriptions of the various factors. Whilst intended for digital musical instruments, the audiovisual system here may also be examined with this model, particularly when considered as live microphone input through audio feature analysis to visual output.

Figure 9 presents an epistemic dimension space plot for the AudioVirtualizer; the authors themselves selected the position with respect to the dimensions. The *Expressive Constraints* dimension refers to the limitations of a given system restricting musical (audiovisual) possibility; since the generativity of the AudioVirtualizer is not highly varied in output, but locked to certain core trajectories and behaviours, constraints are higher. In an associated dimension, *Generality* is low, since the AudioVirtualizer is not a highly reconfigurable system for the user, and does not admit many possible styles of use. *Autonomy* is low since there is little independence in the work, its generation being dependent on user action and reactive to audio input. On the *Music Theory* axis there are some Western music theory assumptions underlying aspects of the audio analysis, such as the 4/4 time signature assumed by the beat tracker, or the key clarity feature presuming 12 tone equal temperament is meaningful. The AudioVirtualizer is meant to be explored, so some *Explorability* is present, even if there is not a long learning curve and years to achieve mastery. Use of the system assumes little in the way of *Required Foreknowledge. Improvisation* is possible in terms of real-time reactivity, especially when using microphone as the sound input, but the depth of improvisation is contingent on the more constrained set-up over the scenes. Finally, there is a balance on the *Creative-Simulation* axis between the precedents given by visualizers, VR video games, and other VR artworks, and some novel machine listening and graphical decisions. The work is not in the mould of highly customisable domain specific languages, nor a standard piece of production software, but an experience somewhere in between an art installation and VR game, meant to be explored and enjoyed.
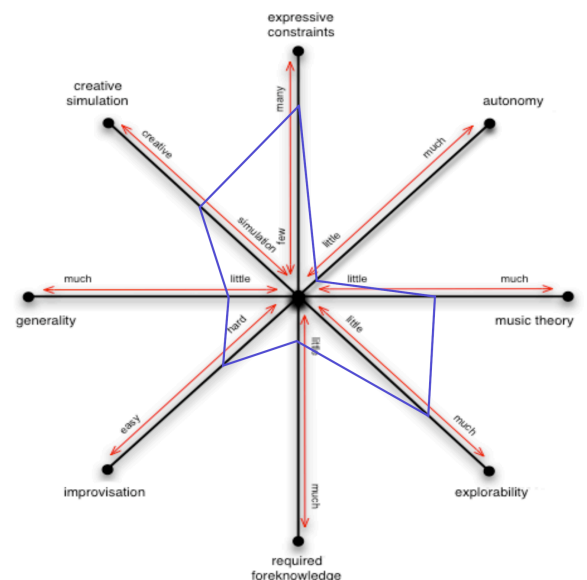

Figure 9. The AudioVirtualizer plotted in Magnusson's epistemic dimension space

6. Conclusion and Future Work

In this article we presented the AudioVirtualizer, an audiovisual interactive VR system and artwork, with generative visuals responsive to features derived from the audio input. A Unity machine listening plugin has been publicly released under a permissive license and the AudioVirtualizer system will be made available for free on the VIVE port to coincide with publication of this article.

The use of Skyboxes in Unity to make 3D environments out of essentially 2D drawings has been explored creatively with 3D elements added into the landscape to respond directly to audio features. The results were tested on a mixed audience, some with and some without previous VR experience and those without previous VR experience enjoyed a slightly more passive approach whereby they used only head movement to change perspective within a given scene and asked for help when they wanted to change scene, whereas people with previous VR experience quickly learned the full interactive capabilities of the system.

In the future we would like to increase the use of immersive sound in such installation work. We would also like to make the system more generative, including creating related skybox graphics on the fly rather than through prepared drawings.

It would also be interesting to transport the system to work within the browser so that the choice of any music input would be easily facilitated without having to go outside of the main scene to find alternative sound inputs. In such an interface machine learning would need to be employed on the audio input to be able to distinguish between different musical styles and the patterns of visual response then should be changed dependent on musical style. The AudioVirtualizer system is most suited to respond to electronic soundscapes and microphone input.

Acknowledgements

References

[1] Alexander, A., Collins, N. Live audiovisuals. In: N. Collins, J. d'Escrivan, editors. *The Cambridge Companion to Electronic Music* (2nd edition). Cambridge University Press. 2017

[2] Bennayoun, M. World Skin, *Siggraph 98, Electronic Art and Animation Catalog* [online available at: https://doi.org/10.1145/281388.281408]

[3] Blackwell, A.F., Green, T.R. and Nunn, D.J., Cognitive dimensions and musical notation systems. In: *Proceedings of International Computer Music Conference*; 2000; Berlin.

[4] DeFanti, T.A., Acevedo, D., Ainsworth, R.A. et al., The Future of the CAVE. *Central European Journal of Engineering*. 2011; 1(1): 16-37.

[5] DeFanti, T. A., Dawe, G., Sandin, D.J. et al., The StarCAVE, a third-generation CAVE and virtual reality OptIPortal. *Future Generation Computer Systems.* 2009; 25 (2): 169-178

[6] Dolinsky, M., Anstey, J., Pape, D.E. et al., Collaborative virtual environments art exhibition. *SPIE Digital Library*. 2005. [online available at: https://doi.org/10.1117/12.589457]

[7] Freeman, J. Personal communication, 2020

[8] Grau, O. *Virtual Art: From Illusion to Immersion*. The MIT Press, 2003

[9] Hegedüs, A. Memory Theater VR, ADA Archive of Digital Art. 1999-2014 [online available at: https://www.digitalartarchive.at/database/general/work/memory-theatre-vr.html]

[10] Johnson, D. OSC-XR: A Toolkit for Extended Reality Immersive Music Interfaces. In: *Proceedings of Sound and Music Computing*;. 2019; Malaga. 2019

[11] van 't Klooster, A. Balancing Art and Technology: The Aesthetic Experience in Body and Technology Mediated Artworks, PhD Thesis ,University of Sunderland. 2011. 52-54.

[12] Latham, W., Todd, S., Todd, P. Putnam, L., Exhibiting Mutator VR: Procedural Art Evolves to Virtual Reality. *Leonardo*, 2019

[13] Magnusson, T. An Epistemic Dimension Space for Musical Devices. In: *New Interfaces for Musical Expression*; 2010; Sydney.

[14] Olofsson, F. Audiovisual Programming workshops. Checked December 2019 [online available at: https://redfrik.github.io/udk00-Audiovisual_Programming/]

[15] Weinel, J. Cyberdream VR: Visualizing Rave Music and Vaporwave in Virtual Reality. In: *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound*; September 2019; Nottingham.