Jules Rawlinson
jules.rawlinson@ed.ac.uk
**University of Edinburgh,
Edinburgh, UK**
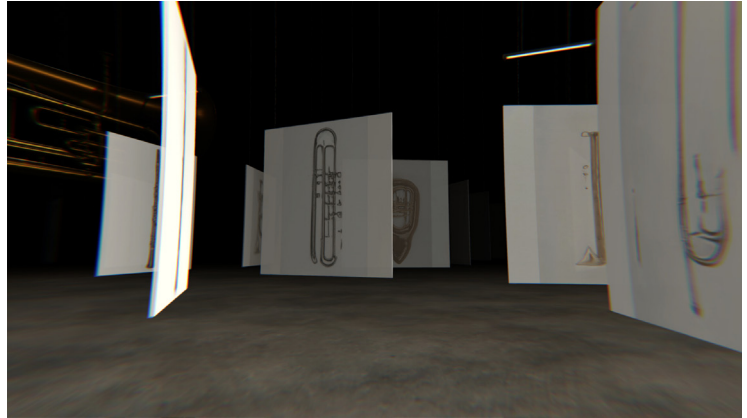
w[i]nd: An Interactive and Generative Audio-visual Composition/Installation in Procedural and Orchestrated Space

w[i]nd is an interactive and generative audio-visual composition / installation in procedural and orchestrated space. This project aims to be a rich and immersive visual and sound art experience in a first-person Virtual Environment (VE). The VE consists of an array of audio-visual exhibits of images of wind instruments created by generative image synthesis techniques coupled with samples of physically modelled synthesis and processed acoustic recordings that exhibit a range of characteristics of augmentation and abstraction. In addition to the displayed exhibits, there are a number of very large-scale three-dimensional models of conventional wind instruments emitting untreated sustained tones typical of these instruments that reciprocally contrast with and comment on the synthesised sounds and images. First person motion, underlying algorithmic parameters and triggers on exhibits are used as a mechanic to influence the sonic density and character of the experience. Among other things, the work explores open sonic form, timbre and character in non-linear interactive experiences and contributes to research into the use of sound and sounding objects

Description

w[i]nd is a virtual sound and image installation that brings sound, images and interactions together at scale to create a participatory and dynamic audiovisual work (Rawlinson 2020).

Procedural space is used as a term to describe the algorithmic and generative variety in sound and image as a function of space, and to account for responses to interaction with space through real-time, programmatic parameter control of sonic output. Orchestrated space is used as a term to describe the fixed spatial arrangement of musical register and timbre across the VE. The work is a walking simulator built with Unity and Wwise game audio technologies and it is informed and inspired by historical experimental spatial composition and open and mobile form works, and recent immersive sound and image installations.

Effort is required to navigate the work's fragmented audiovisual landscape as an act of "ergodic musicking" (Oliva 2019) in which the user plays a participative role in the construction of their experience of the work as they traverse space, blurring the role of composer and performer in the reconfiguration of soundscape. The simulation of a gallery space for the VE is a meta-place that combines characteristics of real space and computer-generated navigable space. The framing of the VE by the computer screen brings immersion at a cinematic level, through the composition of the exhibits in two-dimensional space. The virtual visitor has agency to explore and wind their own way through the VE which is rich with both perceptual and emotional affordances through unfolding audio-visual narrative.

In this VE users are involved and included in a mediated perceptual experience that aims to cause a subjective sense of presence as, and through, sensory immersion and transportation. Factors of presence and immersion are designed into the work as a methodology. Presence causing form variables in the work include: number, scale, dimensionality, resolution, colour and field of view of visual elements; number, spatial realism, frequency range and dynamic range of sonic elements (Lombard and Ditton 1997); natural real-time interaction, consistency of multimodal information, scene realism, environmental richness, movement perception and degrees of control (Witmer and Singer 1998). The virtual character of the work allows for many more spatial sound sources than would generally be possible in real life, and a controlled experience. Through considered visual staging, fixed and aleatoric combination of sound, and motion and proximity-based mechanics and processes the scope,

integration and resolution of the work is extended in terms of audiovisual impact and immersive effect.

The starting point for the work was image synthesis of hybridized, augmented and imperfect simulacra of wind instruments using Style-Based Generative Adversarial Networks (Karras, Laine and Aila 2019) with a dataset of images of wind instruments held by the University of Edinburgh's musical instrument collection. The features in the images that are generated are at times smeared and simplified, at others glitched, complex and distorted, or liquid and abstract. Overall there's a painterly quality in the colour, stroke and texture to the generated images that reflects impressionist modes of music and sound. The visual work includes images from a number of training stages, where new models were created from the output of a previous model as a recursive feedback process, further amplifying the learned features, biases and errors in the neural representational latent space.

Fig. 1. Screenshot of *w[i]nd* Virtual Environment.

A sample library of physically modelled synthesis was created as a sonic response to the visual qualities of the generated images, both in terms of their likely acoustic character as simulacra of woodwinds or brass instruments, but also in respect of the aesthetic properties and features of the images so that pitch, amplitude and timbre are unstable, with rapid spectromorphological shifts between sustained material, sliding tones and fluttered and transient gestural articulations. The work also makes use of processed and untreated recordings of wind instruments to further extend the sonic character of the work.

The synthesized images are displayed as photorealistic exhibits that are procedurally hung in a large virtual exhibition space with an industrial character that

in real world measurements would be 500x 500m in dimension with a labyrinthine procedural distribution of irregular large open areas and smaller pockets of space that connect and separate dense clusters of exhibits. These open spaces offer some relief to the scale and density of the image exhibits, but can also be disorienting due to the variation in form and opacity of the surrounding images and the low light levels in the VE.

Each exhibit emits highly localized audio (best experienced via headphones) of varying stability of pitch, amplitude and timbre triggered by first-person proximity.

The audiovisual exhibits are clustered, dispersed and rotated such that the visitor is granted multiple viewpoints and points of audition, and the exhibits often prevent the user from moving in a linear, straightforward way forcing spatial interaction. The image exhibits are double sided allowing for multiple dialogues as exhibits are rotated to often face more than one other image, allowing for metamorphosis of experience, as the images and sounds are combined, compared and considered by the virtual visitor, and present themselves in new ways. There's a blurred audiovisual diegesis through the juxtaposition of various orders of reality, simulacra and abstraction together with audiovisual relationships that are observed through the "spatial magnetization of sound by image" (Chion 1994, 70) as users attempt to locate or relate sound sources to what's visible to them.

Fig. 2. Screenshot of w[i]nd Virtual Environment – Overhead instrument.
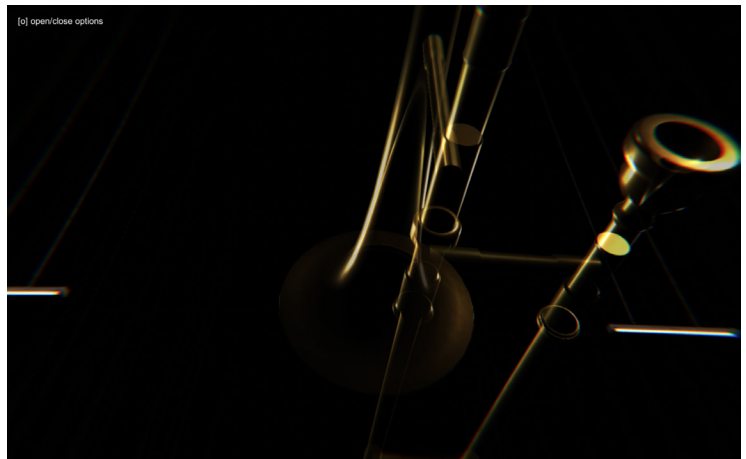
Three-dimensional models of wind instruments of partly-transparent, saturated character are hung overhead at significantly larger than life scale. These models regularly emit untreated sustained tones typical of these instruments. Density of audio events increases with proximity, and sounds travel across moderate distance in the VE, allowing users to orientate themselves and navigate by sound. The positioning of these models results in variation in timbre and register as a form of orchestration in space. Parameterized filtering and reverberation effects extend perceptual spatial characteristics. Stochastic and aleatoric generation of audio events leads to clusters of extended harmony and ambiguous tonality.

A smaller number of exhibits that are different in character to the main images, instruments and each other, are present in pockets of space. These exhibits are obscured and revealed by virtual visitor movement, and draw attention through contained active motion, staging and sound. Proximity triggers on these exhibits switch layers of audio on and off, extending indeterminate compositional scale and structure.

A final group of audio emitters playing constantly looping material of longer duration are randomly distributed throughout the space. The spatial experience of these emitters sits between the highly localized experience of the image exhibits and the more distant models, and the character is often highly unstable in terms of pitch, amplitude and timbre throughout the loop. In addition to the emitter outputs, audio is fed into a low-level processing chain of doppler and granular effects. The depth and value of these effects are affected by real-time parameter control dependent on first-person location and motion.
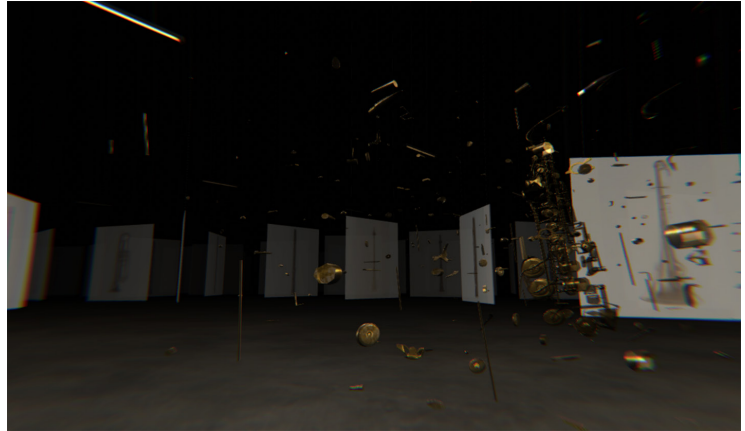
Anecdotal feedback from users has so far been positive including comments such as "a convincing and coherent physical space with dynamic, fluid content that is responsive to gazing and micromovement". One user said "there's a clear connection between space, motion and sonic output… a good range of musical events and gestures makes the environment come to life… it responds in surprising ways… there's always something new that makes you want to keep exploring and looking for surprises". Another noted "the spatial arrangement of the visual and also auditory components was compelling, motivating me to spend more time within the installation that I might have done."

The work presented here combines natural and exploratory modes of interaction and degrees of control in a multimodal presentation offering movement perception, environmental richness and scene realism. The experience is dynamically configured in space and time. Open spaces and stable sounds act as connectors giving shape to the experience, while clustered and variable elements demand attention and focus or invite curiosity and may be seen as attractors and retainers through "perceptual opportunities" (Fencott 2005). Overall, the approaches taken in developing the work have resulted in an immersive and engaging experience with characteristics of agency and presence.

Link: Standalone applications for Mac and PC are available at
https://pixelmechanics.itch.io/wind.

References

Chion, Michel.
1994. *Audio-Vision: Sound on Screen. New York*, NY: Columbia University Press.

Fencott, Clive.
2005. "A Methodology of Design for Virtual Environments." In *Developing Future Interactive Systems*, edited by Maria-Isabel Sanchez-Segura, 66-91. London: Idea Group Publishing

Karras, Tero, Samuli Laine, and Timo Aila.
2019. "A Style-Based Generator Architecture for Generative Adversarial Networks." Paper presented at 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), Long Beach, CA, 15-20 June 2019. https://doi.org/10.1109/CVPR.2019.00453

Lombard, Matthew, and Theresa Ditton.
1997. At the Heart of It All: The Concept of Telepresence. *Journal of Computer Mediated Communication*, Vol. 3, No. 2, September 1997. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Oliva, Costantino.
2019. Musicking with Digital Games. Paper presented at DiGRA 2019 *Game, Play and the Emerging Ludo-Mix*, Kyoto, Japan, 6-10 August 2019.

Rawlinson, Jules.
2020. *w[i]nd* [Unity Standalone Application]. https://pixelmechanics.itch.io/wind

Witmer, Bob G., and Michael J. Singer.
1998. "Measuring presence in virtual environments: A presence questionnaire". *Presence: Teleoperators and Virtual Environments*, 7(3), 225–240. https://doi.org/10.1162/105474698565686