Stowaway City
An Immersive Audio Experience for Multiple Tracked Listeners in a Hybrid Listening Environment

MICHAEL MCKNIGHT

Stowaway City is an immersive audio experience that combines electroacoustic composition and storytelling with extended reality. The piece was designed to accommodate multiple listeners in a shared auditory virtual environment. Each listener, based on their tracked position and rotation in space, wirelessly receives an individual binaurally decoded sonic perspective via open-back headphones. The sounds and unfolding narrative are mapped to physical locations in the performance area, which are only revealed through exploration and physical movement. Spatial audio is simultaneously presented to all listeners via a spherical loudspeaker array that supplements the headphone audio, thus forming a hybrid listening environment. The work is presented as a conceptual and technical design paradigm for creative sonic application of the technology in this medium. The author outlines a set of strategies that were used to realize the composition and technical affordances of the system.

Extended Reality (XR) is an umbrella term that incorporates all immersive experiences and their technologies. XR usually involves a wearable device to facilitate human-computer interactions, with varying amounts of sensory input and feedback. Real and virtual objects are combined or used solely to extend/change our perceived reality. The three predominant terms to categorize this relationship, at the time of this writing, are virtual reality (VR), augmented reality (AR) and mixed reality (MR). These exist on a spectrum from the completely real to the completely virtual, with interpolated areas and terminology around them; these terms are often used with some crossover. Recent rapid developments in XR technology have rendered these technologies much more affordable and accessible than ever before, which can therefore be suggested that XR is an ideal medium for electroacoustic/acousmatic composers, and sound designers more broadly, given the inherent spatial characteristics of sonic objects. The difference is that XR also affords the listener agency within the composition, which can impact how a composer intends their piece to be heard. Listeners are free to make choices that can change the relationship between sonic objects in an AVE, often in unexpected ways. This may seem undesirable, but conversely it also becomes a device available to the composer in their practice. Todd Winkler [3] highlights the influence of new technology on musical thought as a driver for innovation and its capacity for new sounds and performance capabilities. This article seeks to defend the view that, where this can be a positive influence, the technology should be kept artistically relevant.

Stowaway City is a narrated soundscape composition that seeks to ask these questions, exploring virtuality and immersive technology as sonic art. The work was designed with scalability in mind and can accommodate up to 10 listeners at any one time, all having unrestricted movement in the Sonic Lab at the Sonic Arts Research Centre (SARC) in Queen's University Belfast. Participants' tracked positions and rotations are utilized to afford six degrees of freedom (6DoF) in the binaural rendering of audio while the participants...
navigate the AVE. These individual sonic perspectives are received wirelessly by each listener through open-back headphones that are enhanced by a spatial loudspeaker array used to create a hybrid listening environment [5]. The intention is to blur the boundary between the two presented sound fields so there is an ambiguity in where one system ends and the other begins, resulting in an expanded sense of depth and space. Everyday listening conditions consist of complex relationships between multiple localized sources, spatial acoustic properties and psychoacoustic phenomena that are received by the whole body, not just the ears. A hybrid listening system aims to simulate this experience, encouraging believability and immersion in the work. The composition and unfolding narrative are then revealed to participants as they move through space. There are nine scenes in total plus various localized sound objects to discover, which are mapped to unique physical locations in the Sonic Lab. The work was included as part of the Cathedral Quarter Arts Festival, Belfast, in 2019 (Fig. 1).

The script was originally written as a soundwalk by Emily DeDakis, with each scene referencing a specific site in Belfast. The walk was timed with the narrative and set to play at corresponding locations along the route. A guide would set the pace to ensure that the participants would arrive at each site at the correct time. Bringing the walk into a virtual environment allowed greater scope for listener interactions, consistency and control over accompanying soundscapes. Translating a soundwalk to a virtual experience in this manner meant adapting the linear story to a nonlinear interactive narrative. Hand and Varan [6] state that, although a contentious area in research, interactive narratives have long been advocated as having the potential to create more immersive and transformative experiences for audiences. They suggest that audience empathy is significantly and consistently increased with the addition of interactivity; this resonated with the core themes of the Stowaway City script. Indeed, interacting with sound is fundamentally different than listening without interacting [7]; interacting changes how the listener engages with the work. We rely on the movement of our bodies to access and explore our environment [8], holding a close relationship between this movement and sound as either a product of an action or as an action in response to auditory stimuli. The virtual form that Stowaway City took is a natural extension of the script’s origin as a walk, reflected in the aforementioned agency.

The work is situated as part of my research, in which I seek to deepen understanding of the act of listening in immersive sonic experiences; the work also responds to the latest ubiquitous wave of XR technology. Despite XR being a relatively new medium, inherent design paradigms and aesthetics are often imposed by these development tools. Additionally, these new technologies are largely driven by their visual components, which can be restrictive. Compositional praxis from the electroacoustic tradition, which has a rich history in spatial audio, can be applied to current XR development in a new and engaging way. Likewise, the recent advances in immersive technology could inform electroacoustic composition.

**IMPLEMENTATION**

I designed the piece to utilize the spherical speaker array in the Sonic Lab (SARC), which has 48 channels including a subwoofer. Compositional elements diffused through this system were mixed in third-order ambisonics (3OA). This complements the spatial audio in the listener’s headphones, which was binaurally decoded from first-order ambisonics (1OA). I preferred working with higher order ambisonics, due to the greater perceptual accuracy in source localization and a larger “sweet spot,” once decoded. 3OA ambisonics was chosen as a compromise between these benefits and the computational system resources, which can be impacted dramatically with the exponentially increasing number of audio tracks required for higher orders. There is also only a marginal improvement of perceived localization accuracy, approximately 2°, between third- and fifth-order ambisonics [9]. The use of 1OA for headphone audio was imposed by limited channel numbers in Unity, the game engine (software environment) in which the system was implemented.

The choice of headphone was important, as minimal coloration or attenuation of the loudspeakers was essential,
as was sufficient audio quality and comfort for the listener. Sennheiser HD600 headphones satisfied these needs. Careful attention was given to ensure that the level relationship between the headphones and the loudspeakers was balanced, with listeners encouraged not to adjust this balance during the performance. The balance was decided partly as an aesthetic choice by myself, as the composer, and through a consensus of opinions from a group of trained listeners at SARC. Although this balance was subjective, formal listening tests are planned as part of a hybrid listening study [10] to further investigate this relationship.

The tracking system is based around HTC Vive Trackers, which allow full room-scale 6DoF tracking for areas up to approximately 100 m², with a high degree of accuracy and ability to function in low light. In order to achieve this maximum volume, four second-generation base stations are required; however, for smaller areas two are sufficient. The band of each headphone is modified with a tracker that communicates via Bluetooth to a computer running SteamVR software that manages the pairing of the device. The trackers’ data is then passed into Unity, which contains all logic and audio for the piece. Room calibration is also managed in Unity and adjusted for fine tuning the performance space. The tracker’s normal and intended use is in a VR environment in conjunction with the manufacturer’s related hardware but it was repurposed for this work. The output for each listener’s binaurally decoded audio is then connected via a Dante network and routed to the corresponding stereo transmitter. This transmission was achieved using an in-ear monitoring system by Sennheiser over radio frequency in the UHF spectrum. It was essential that listeners’ movements were not restricted, as this could impact their potential immersion. This method reduces the chance of errors from a predetermined fade duration, due to a listener exiting and reentering a scene too quickly. To aid navigation from one scene into the next, a small number of sonic beacons (localized mono sources) were placed in between scenes. These comprised

**COMPOSITION**

The narrative is pieced together through a series of fictitious voicemail messages left on an answering machine by the story’s protagonist. Every scene contains a single voicemail accompanied by a soundscape, each with a different sonic character and location to echo the events being portrayed. With a few exceptions, all sounds were composed of location field recordings using a Rode NT-SF1 ambisonic microphone. These were used not only for the “natural” soundscapes but also to derive the majority of other sounds heard in each scene—for example, drones. During a sound’s capture and design, its perspective was considered broadly as either from a source or an observer, which helped me decide where in the sound field it should be placed, its proximity and which playback system it was intended for. The aim was to support the narrative by considering sound as a device for storytelling, to draw focus on the protagonist and thereby encourage the listener into “her reality.”

Each scene is attached to a game object in Unity and placed on a plane that represents the floor of the lab. The transitions between scenes fade in and out based on the listener’s position relative to the center of the scene and follow the participant’s speed of movement. As the listener moves toward a scene, they are guided toward its center as result of this behavior. This method reduces the chance of errors from a predetermined fade duration, due to a listener exiting and reentering a scene too quickly. To aid navigation from one scene into the next, a small number of sonic beacons (localized mono sources) were placed in between scenes. These comprised

---

**Fig. 2.** System setup including audio and data flow. (© Michael McKnight)
phone-related sounds, including an engaged tone, ringer tone, etc., to reinforce the idea of the dialogue as voicemail. Figure 3 shows an overview of the lab and the position of sounds within the space.

The composition was formed around a hierarchical structure that examined relationships between the sounds themselves, their position in space and the system as a whole. On a low level, each scene was composed within itself, as an arrangement of sonic elements that were broadly considered either dialogue, effects or soundscape. There is a deliberate interplay among these elements; each sonic layer seemingly impacts another in a way that contradicts reality. I therefore treated the dialogue as a spatial object, which is perceptively external to the listener, as opposed to using head-locked stereo that has an “in-head” effect. This kept the narrative as part of the soundscape. At a mid-level, consideration was given to how scenes relate to each other for continuity, navigation and transitions between them. A high-level view of the composition concerned the position of the scenes in space, the loudspeaker system that formed the hybrid listening environment, and how this material related to each scene. The strategies I employed can be categorized by three emergent themes: spatial, spectral and temporal.

Within a sonic composition involving recorded or synthesized media, it is possible to consider “space” as both inherent in the work as well as the physical space in which it is projected. There is also a connection between these that can affect a listener’s perception. Natasha Barrett [11] explains how space can be more than just a setting in which sonic events unfold and suggests a number of aesthetic strategies for the composer to utilize “space” as an important carrier for musical structure. At the low-level compositional process in this work, much of Barrett’s aesthetic strategies apply, such as spatial illusion, spatial allusion, proximity and motion. However, introducing a variable listening position for each listener blurs the distinction between physical and compositional space, as well as how that listener engages with their auditory environment. What is interesting to me in Stowaway City is how the space is shared by the audience both physically and perceptually.

The hybrid listening system of Stowaway City brings a physicality to the experience, in addition to an expanded sense of depth in the AVE. Placing sonic objects “physically” in space via loudspeakers, beyond the immediately present sound field of the headphones, makes these sounds felt as much as heard and helps contribute to the immersion of the listener. These events are experienced by everyone simultaneously, thus promoting a “shared perceptual space” [12] in these moments. It was important to ensure that these objects remained relevant for all listeners at all times regardless of scene or position. The material was composed of sonic motifs with the idea that they would both echo previously heard sounds, subtly allude to what was next or coincide with their currently occupied scene. They were also slightly abstracted and kept to short localized bursts with varying density. These were timed and spaced so as to allow for at least half of the experience to be on headphones only. A spectral split was adopted, prioritizing lower and higher frequencies in the loudspeaker system. The motifs were based on the sonic themes of each scene that are separated temporally not only through the experience but in the narrative as well, which facilitated the nonlinear structure. The “spaces” between scenes were intended to suggest this separation of time in the narrative and to give each one a distinct spatiotemporal location. During the compositional process, this temporality also manifested itself through the development of a rhythm for each scene, which centered around the dialogue while leaving “space” for other sonic elements. This linguistic character provided a level of continuity and structure for the listener. Upon discovery of a scene, an environmental soundscape is presented to the listener, recorded at various sites around the city. These are heard before, during and after each scene but are augmented by other sonic elements for the scene’s duration. The beginning and end of each scene is communicated with a “marker” by way of a typical voicemail beep. The same sound, in the same relative point in space, is used to bookend the scene and reinforce the concept of phone messages.

The spatial, spectral and temporal themes are interconnected and do not exist as separate entities. We come to know space through the body and movement experienced through time. It takes time to move through space, to understand the relationships of and within space; it is through a myriad of complex sensual relationships that we are provided a notion of place [13].
RESULTS AND FURTHER WORK

Qualitative results extracted from a thematic analysis of questionnaires from participants' feedback showed 74 percent positive responses toward the experience. From approximately 50 questionnaires, themes emerged around temporality, ambiguity, curiosity towards others' perceptions, responsibility and gamification. The localized mono sources between the scenes, intended for navigation, brought mixed reactions, but most participants mentioned "playing" with sound and searching for it. This was particularly evident in relation to the voice. This was spatialized as part of the ambisonic scene, and listeners would therefore be unable to get perceptively closer to it. Having the voice as a spatialized mono object would have allowed more opportunity to "play" with it. More research is required to discern participants' patterns of movement, but empirical evidence showed a trend of rapid exploration of the space at the beginning, before the user settled down to listen to each scene, followed by a final rapid sweep of the space before the listener leaves.

Future work will include integrating the tracking system to affect loudspeaker audio as well as in headphones. This will include strategies for audio materials around this interaction and how collective listening positions could affect properties or behaviors of sound objects.

CONCLUSION

This paper contributes to the growing field of immersive audio by demonstrating an approach to composition that combines storytelling, electroacoustic composition and XR technology. It also presents the technical aspects of the system used in the work Stowaway City as a concept for creative sonic application of immersive technologies and highlights the importance of sound design in creating plausible AVEs. There is a need in the current climate of immersive experiences to understand the unique set of relationships that coexist in the medium and how to compose for them in an engaging and artistically relevant way.

Acknowledgments

Stowaway City was a collaboration with writer Emily DeDakis, as well as Luke McNeice and Jake Young from Kainos to develop the tracking system software in Unity.

References and Notes

4 "Degrees of Freedom" is a term used primarily in the context of VR and refers to the number of axes on which a body can move in 3D space. 6DoF encapsulates pitch, roll and yaw, the three axes for head rotation and three axes to translate position.
10 See Rebelo and McKnight [5].

Manuscript received 30 December 2019.

Michael McKnight is an artist and Ph.D. researcher based in the Sonic Arts Research Centre at Queen's University Belfast. He works as a location field recordist and sound designer specializing in immersive audio.