

<CT-1>FOCUS

<CT>Into the Here and Now: Explorations within a New Acoustic Virtual Reality

<CAU>Laurence Cliffe

Laurence Cliffe (researcher, artist), Mixed Reality Laboratory, University of Nottingham, Computer Science Building, Wollaton Rd, Lenton, Nottingham NG8 1BB, U.K. Email: laurence.cliffe@nottingham.ac.uk. ORCID: 0000-0001-7111-5575.

See <https://direct.mit.edu/leon/issue/58/2> for supplemental files associated with this issue.

<AT>Abstract

<ATX>The author reflects upon listeners' experiences and the practice of developing a number of audio augmented reality sound installations deployed between 2019 and 2020. The installations realized audio augmented objects: physical real-world objects augmented with virtual audio content. Focusing on users' feedback from one such installation environment, the author suggests that audio-augmented objects are capable of realizing virtual audio experiences that are indistinguishable from reality, offering the potential for the development of new and exciting auditory experiences that embed virtual audio content within listeners' physical environments.

<TX-N>Recent developments in mobile augmented reality technologies have enabled the creation of experiences that are capable of robustly and persistently positioning virtual content at very specific locations within our physical reality. Along with real-time binaural rendering of virtual sound, it is possible to create virtual audio experiences with six degrees of freedom (6DoF); a binaural audio experience with both rotational and translational interactivity. Such an audio experience delivers virtual audio content that the listener receives in the same way as we experience sound in reality; the virtual sound source is perceived as being located at a fixed position in physical space even when the listener rotates their head, and the sound attenuates as the distance between the listener and the sound source's location grows. Within the sound installations described in this article, rather than residing at arbitrary locations in physical reality or attached to virtual 3D models, the virtual audio content resided at the locations of viable, recognizable and believable sources of sound within listeners' physical realities, such as a radio, a clock, or the space outside a window. This 6DoF binaural listening experience embedded within the physical reality constitutes a new type of listening

experience. Both active and mobile, it entices the listener away from the fixed position of static binaural listening and the fixed rotational experience of a three-degrees-of-freedom (3DoF) binaural audio experience (one that only contains rotational, rather than both rotational and translational, interactivity). Furthermore, the placing of the virtual audio sources at viable sound sources within the physical reality changes the experience of the acoustic virtual reality for the listener from one that has traditionally been *there*, to one that is now *here*.

<TX>If we consider the following point by Jonathan Sterne regarding concerns of audio engineers, we can consider the development of this new acoustic virtual reality as the latest step in an ongoing pursuit of acoustic realism within audio production technologies: “To understand the aural dimensions of virtual reality, we need to consider audio engineers’ century-long obsession with creating what we would now call virtual acoustic spaces in recordings” [1].

The point here is, when listening to a recording of a live concert in mono, one can feel transported to the concert hall or festival, and subsequently even more so within a stereo recording, and even more so within a binaural recording; within each subsequent evolution of audio formats and production technologies we are afforded a greater sense of acoustic virtual reality. What unites all these previous experiences is that they take the listener *into* the acoustic virtual reality, albeit with increasing sophistication. Within the new acoustic virtual reality realized through dynamic binaural audio and audio-augmented objects, the acoustic virtual reality is now *here* with the listener.

<h1>Sound Installations 2019--2020</h1>

<h2>*Alien Encounters*</h2>

During the deployment of *Alien Encounters* (2019) [2] as part of an exhibition celebrating the 50th anniversary of the moon landing, something strange happened. A person listening to the sound installation removed their headphones, rotated their head around, moved themselves nearer to the radio that was placed on the table in front of them, put their headphones back on, and then took them off again. The conversation that quickly ensued revolved around their inability to believe that the sound they were listening to through their headphones was not in

fact coming from the 1970s transistor radio on the table. The radio on the table was switched off, unplugged, and silent, and the sound they heard was a recording of David Bowie's *Starman*, spatially positioned at the location of the radio's speaker, binaurally rendered, and delivered through a set of stereo headphones with an iPhone mounted on the side.



<Fig 1>

The app running on the iPhone was using Apple's augmented reality (AR) framework ARKit to track the phone's position and movement through physical space---and thus the position and movement of the listener's head. The app was also employing Simultaneous Location and Mapping (SLAM) technology, then a somewhat still emergent feature of AR that enabled the continued and persistent tracking and positioning of virtual content within a location in physical space, even when the location in question was beyond the iPhone camera's field of view.

From an artistic standpoint, the impact of this visitor's experience seemed twofold. First, an experience containing virtual content had been realized that was indistinguishable from reality. Second, the experience of consuming pre-recorded audio content appeared to have changed from one of being *there and then* to one of being *here and now*. This second point hinged on the realization that the source of the recorded music had been projected into the physical space of the listener, rather than the listener being projected into the virtual acoustic space of the music. This perceptual change in the spatiotemporal experience of listening to recorded audio content mirrors the serendipitous connections between the visual physicality and the auditory virtuality that are capable of realizing auditory experiences that have been described as *cinematic* or *filmic* [3,4]. Such experiences, like those first enabled by the

mobile listening experience associated with the advent of the Sony Walkman, rely on chance associations between the visual and auditory. But, through the directly authored association between the visually perceived physical environment and the delivered virtual audio content, such experiences of heightened meaning no longer appear to have to rely on chance association. This echoes what Frauke Behrendt specifically refers to as *Placed Sound* [5]; sounds placed by someone other than the listener or sounds that have been placed by someone to be found and listened to by someone else. We therefore begin to see, with the move from a standard linear mobile experience to one of what Behrendt defines as “placed sound,” a way in which meaningful encounters between sound and the listener’s physicality can be purposely constructed.

<h2>*The McMichael Experiment*</h2>

The previously described listener experience related to *Alien Encounters*, one of four audio-augmented reality installations I developed and deployed between 2019 and 2020.

Another, *The McMichael Experiment* (2019) [6,7], was similar although comprising a single 1920s radio from the collection of the UK’s National Science and Media Museum. It was augmented with multiple period appropriate archival radio broadcast recordings that listeners could tune in to amid the sound of radio static by moving around the radio. Four archival radio broadcasts were situated around the four sides of the physical radio object in virtual space; as listeners moved around the object they could use the position of their bodies much like the tuning dial on an analogue radio receiver.



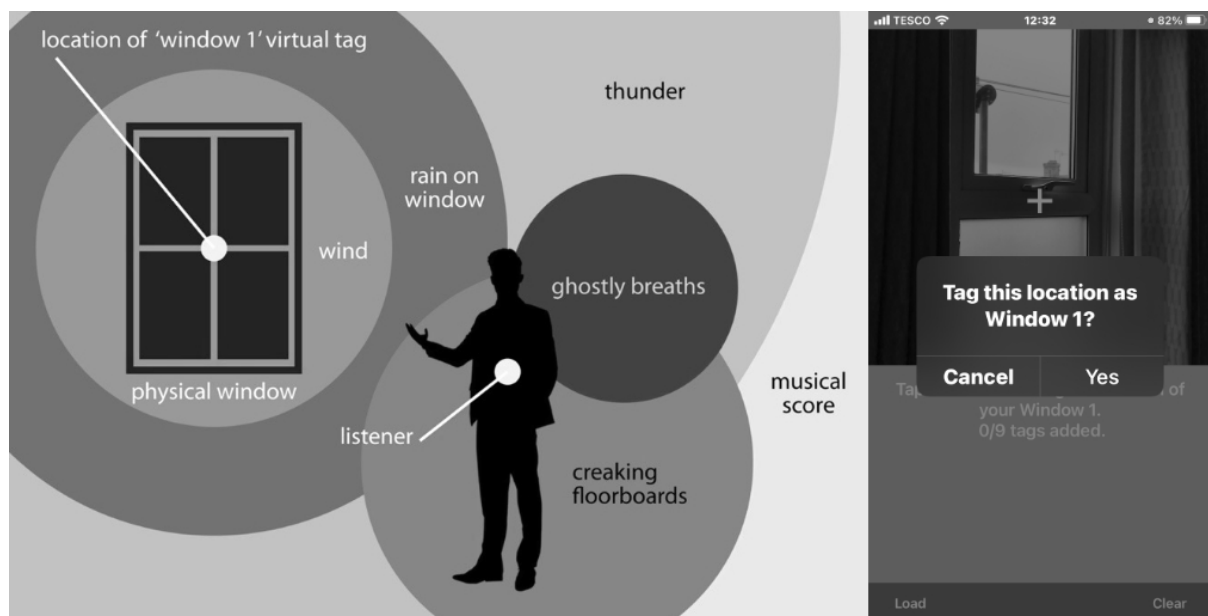
<Fig 2>

As Behrendt points out, the interactional modality of placed sound is walking [8], and within an indoor environment the direction and space within which one can walk is determined by the shape and confines of architectural space and the obstacles and objects situated within it. Therefore, when considering the interactional design of indoor placed sound experiences, the architectural setting of the experience needs to be carefully considered and built into the interactional model. Furthermore, by associating virtual and spatialized audio sources to objects or features that may not be directly related to the experience, one can consider how objects within an experience can advertise their presence beyond the traditional confines of line of sight. For example, a unique architectural feature within a gallery could be tracked to position and trigger an audio source attached to an object behind a partition wall.

Within the National Science and Media Museum, *The McMichael Experiment* demonstrated how museum artifacts can act as interfaces to digital audio archival content, at once contextualizing archival audio and recontextualizing silent museum objects [9].

<h2>Horror-Fi Me</h2>

Also deployed over this period was *Horror-Fi Me* (2020) [10], an audio-augmented reality mobile app-based experience that allowed users to place virtual sound sources onto objects and locations around their homes and construct an explorable mixed-reality horror movie soundscape. Within the AR mobile app, a list of objects that users could tag within their home was presented, such as windows, a clock, a faucet, and a door. Once tagged with a virtual marker, each of these objects then had virtual sound sources attached to them; windows had rain and wind samples attached to them, the door had an intermittent knocking sound attached to it, the clock had the sound of a grandfather clock persistently chiming midnight, and the dripping faucet echoed within a bathroom. These samples were accompanied by a specially commissioned contemporary classical score that placed the experience firmly within the cinematic horror genre.



<Fig 3>

While Apple's AVFoundation framework with which *Horror-Fi Me* was authored allowed for custom values to be set for a distance attenuation model across the whole audio environment, at the time of development it did not provide a method through which distance attenuation characteristics could be determined for individual sound sources. In essence the framework assumed, quite rightly, that the perceived acoustic characteristics of the sound within the virtual environment should be determined by the acoustic characteristics of the environment that it is within. This assumption makes perfect sense when we want sound sources to be perceived within the virtual environment just as they would be within a real

physical environment, such as sound sources situated within virtual spaces with differing acoustic properties of a video game. But this model creates problems for the construction and deployment of audio-augmented reality experiences, specifically if we consider the idea that *believability* is different from *realism*.

First, *Horror-Fi Me* was designed for use within any indoor domestic environment, and the dimensions and acoustic properties thereof are variable. For this reason, rather than adopting a specific audio environment with specific characteristics, a more generic approach was required that would render the sources as *possible* or *probable* within a variety of real-world spaces. A default audio environment for the space was therefore adopted, and the sound sources were attributed their own characteristics (range, reverb, rendering, and type) so that they could be individually adjusted for the best possible perception of their reality in general space. This approach also proved useful for the rendering of sources that were intended to be perceived as existing outside of the *acoustic space*, such as the sound of crows outside the window or the sound of thunder. This meant that the range, reverb, and attenuation of these virtual sound sources could be individually adjusted to better reflect their behavior within their intended reality rather than being determined by the audio environment of the audio engine.

Second, not having the ability to control distance attenuation at an individual sound level resulted in incoherent and indiscernible sound sources or rather created a soundscape that became part of an overarching ambience rather than being individually discernible and easily perceptible as coming from a specific location. By individually controlling the sounds' characteristics, more control could be leveraged over the authorship of the audio experience; specific sounds could be designed to travel further than others and attract a listener's attention; controlled zones of ambience could be created; and, perhaps most importantly, limiting the ambient range of a source made it easier to localize and isolate the source (when in range) than when within a soundscape composed of sound sources with largely equal volume levels. Therefore, it appeared that this approach facilitated the perception of a more precise augmentation and the perception that individual physical objects were augmented, not just an approximate location in three-dimensional space.

<h3>The User Experience</h3>

Horror-Fi Me was launched and presented as part of *Live Cinema III: Festival of Research and Innovation 2020*. A download link for the application was circulated as part of the festival's online marketing material, and attendees were encouraged to download and try out the experience prior to its presentation at an online version of the festival.

The analytical data provided within the mobile application's TestFlight account logged a total of 102 installations of the *Horror-Fi Me* mobile application and a total of 97 sessions. The TestFlight application defines an installation as the number of times the app has been installed on a compatible device, and it defines a session as "the number of times the app has been used for at least two seconds" [11]. Both these totals are based on users who have agreed to share this data with me, the developer.

After the launch of *Horror-Fi Me*, TestFlight reported that a total of 29 copies of the application had been downloaded and installed on to a compatible device with at least one session logged against it. Of these 29 installations, 17 had more than one session logged, with 10 installations logging 5 sessions or more. All 29 participants associated with these installations were approached for an interview, with 10 participants responding and agreeing to this request. The 10 participants interviewed were aged between 27 and 65 years old, of mixed gender, all had successfully downloaded and installed the app, authored an AAR experience, and undertaken that AAR experience within their own homes with their own iPhones. These 10 participants used a variety of stereo headphone types to listen to the resulting audio content, including on-ear, over-ear and earbud headphones.

The following are excerpts from post-participatory interviews with the 10 participants:

<EX>I think all of the sounds felt realistic, yes definitely, I think I was a bit surprised .
..

it made me want to move away from the door, it was a bit scary, I guess. So, it's like, well it's horror, so you know not to open the front door when it's knocking.

. . . all of a sudden you've got this heavy breathing somewhere which did make me look around . . .

It didn't make me think, oh yeah, well my clock doesn't make that sound. You know what I mean? It's like that virtual reality stuff . . . it's so real.

I, er,... tagged is quite a contemporary clock --- it doesn't tick, you know. And suddenly it becomes an old clock...

It's sort of like it's not your house... while you're doing it, it does feel like it's not your . . . space. It wasn't my house anymore; I was a stranger. . . .

. . . it felt a little bit like playing a game, like a *LA Noire*-type game on PlayStation or something.

. . . specifically [it called to mind] *Silent Hill* . . . I played it on a GameCube, probably about 15 years ago . . . it was a ghost adventurer type of thing . . . the atmosphere took me straight back to that . . . the atmosphere and being able to kind of take myself to a completely different time very quickly. . . .”

<h1>Discussion</h1>

<h2>An *Unwilling* Suspension of Disbelief</h2>

The inability of the listener to accept that the silent radios within the installation *Alien Encounters* were not the actual source of the audio content they were hearing, along with the users' feedback from *Horror-Fi Me*, suggest that these experiences created situations where virtuality was, at times, indistinguishable from reality. This experience highlights the distinction between a *willing* and an *unwilling* suspension of disbelief; the difference between a conscious suspension (in the literary sense), to pursue the author's intent, and an unconscious suspension, reflecting an immersion in virtuality as reality. Here I reference the English poet Samuel Coleridge's concept of the suspension of disbelief, which implies a *will* on the part of the audience to believe in what they know to be the fantastical within a fictional narrative experience.

Described by composer Winifred Phillips [12] in relation to the musical content of video games, environmental authenticity, along with cultural authenticity (identifiable and

generally accepted environmental and cultural associations between visual and auditory content), are seen as contributing factors in the creation of an atmosphere that helps realize a suspension of disbelief. This willing suspension is pressed upon the audience, and their complicity is manipulated by such literary and cinematic trappings as grounding the fantastical within the believable, or as Phillips puts it: environmental authenticity. Or, indeed, the creation of a narrative compelling enough that, to immerse oneself within it, the reader or listener may forgo any misgivings regarding its probability. Either way, ultimately the virtual is still distinguishable from reality, and disbelief's suspension, as Phillips points out, is a result of a conscious contract between viewer and author.

In her book *Gothic Music*, musicologist Isabella van Elferen [13] cites Schopenhauer's conviction that "music is an unconscious exercise in metaphysics in which the mind does not know it is philosophizing" to make a distinction between a *willing* and an *unwilling* suspension of disbelief. Though interesting with regard to the inclusion of musical content within an audio AR experience, it is van Elferen's unveiling of the second way that belief can be suspended that seems of specific interest here: "Musical experience is an unwilling suspension of disbelief that can mould listeners into believing in Gothic ghosts: the haunting voices of film music, the domestic spectres of television, the interactive supernatural of computer games" [14].

Within the audio experiences presented here we see, through environmental authenticity, cultural authenticity and the "*unconscious philosophising of the musical listening experience*," the creation of both *willing* and *unwilling* suspensions of disbelief, the latter manifesting itself with participants seemingly unable to distinguish reality from virtuality.

The creation of an *unwilling* suspension of disbelief seems to be, at least for the time being, specific to audio AR, rather than visual AR, and we can attribute this to two main factors. First, the invisibility of sound means that the varying amounts of drift that leads to virtual content wondering away from its assigned physical location within AR experiences is not as perceptible within an audio AR experience as it is within a visually oriented AR experience; this therefore contributes to the realization of a more robust and perceptibly realistic virtuality [15]. Second, with modern smartphones capable of delivering real-time, high-fidelity, and responsive binaurally rendered audio, we have the means to deliver virtual, high-resolution audio content that is perceptibly real, a factor that is far less easy to achieve, in a fully mobile and accessible way, with virtual graphical content.

<2>The Reembodiment of the Listening Experience

Audio-augmented objects signal an experiential change in the common consumption of acoustic virtual realities; they have the potential to transform the acousmatic acoustic virtual reality into a more direct, or live listening, experience [16], one within which the sound sources of the acoustic virtual reality can be perceived as being present. As such, they cease to be audio-only experiences, or indeed augmented reality or augmented virtuality experiences, and become true mixed-reality experiences composed of virtual audio content and components of the listener's physical reality that have the potential to be indistinguishable from each other.

While Schaeffer's "acousmatic field" deals with the removal of the sound's source from the listening experience, with the audio-augmented object, we are dealing with its reintroduction. This is perhaps most evident when we are audio-augmenting sound-making objects; objects that are directly responsible for the creation of their own sound, or to use Schaeffer's Pythagorean distinction, those objects with which we have a visible, touchable, or measurable relationship with the direct source of the sound.

An audio recording of rain pattering against the glass of a window is a good example. If heard through loudspeakers or headphones with no visual or experiential association to its original physical source in reality then we are involved in an acousmatic listening experience; the Pythagorean veil remains intact. On the other hand, by directly associating this audio recording with a plausible and probable physical source, we are bypassing the acousmatic experience; the sound source has visibly and physically returned to the listening experience, and it ceases to be perceived as acousmatic.

Placing sound in three-dimensional space does not make it non-acousmatic (at least from Schaeffer's Pythagorean perspective), as the source of the sound is still not visible. This, on the other hand, changes as soon as we place that sound at the same location as a plausible, physical sound-making object, as the source can be perceived as present and visible. Therefore, this change only takes effect if the physical object is indeed perceived to be the source of the virtual content, but this is entirely possible, and in these cases a perceivable transformation has taken place from what Schaeffer terms a "non-live listening" experience into a "live listening" experience.

This transformation relies not only on the effective positioning and rendering of realistic virtual sound sources but also on the plausibility of the augmented object's emission of the virtual sound with which it has been associated. It seems that the nature of this transformation is also dependent on the type of object that is being audibly augmented. For example, to reunite a radio broadcast from a sound archive with a radio set would be to reconstruct a lost acousmatic experience. To attach a sound recording of rain against a window to a physical window replaces a potentially acousmatic experience with a perceptively direct listening experience. Therefore, the effect of either breaking or remaking the acousmatic field appears to be a consequence of the *type* of object and its relationship with the audio content with which it has been augmented.

In either case, whether recreating an acousmatic listening experience or creating a direct listening experience (where the augmented object is perceived as the direct source of the sound itself), a more connected and intimate relationship between object and listener is created. While the audio-augmentation of a radio with archival radio broadcast material is not a direct listening experience, it is more familiar and truer to its original context and therefore a *more* direct listening experience than, for example, listening to the same content via a web-based audio player.

This familiar, contextualised, and more direct experience with the source of the sound (whether perceived as direct or contextually acousmatic) constitutes a more *haptic* type of listening experience, one that communicates the physicality of the sound, its source, and their relationship to the listener.

The embodied and powerful listening experience that this new acoustic virtual reality enables can be thought of as a return to the embodied listening experience obtained from listening to a live musical performance, though realized through recorded sound. This reembodiment evokes the psychoacoustic qualities that Sterne [17] attributes to the mp3. This new acoustic virtual reality is, as Sterne suggests regarding the mp3, embodied, mobile and a celebration of the capabilities of human auditory perception. Perhaps also of interest here, and what most clearly distinguishes the consumption of this acoustic virtual reality from the psychoacoustic qualities that Sterne attributes to the mp3, is its overt embodiment of virtuality. As Sterne suggests, with the mp3 and other recorded and transmitted audio formats, the medium plays the user in an unconscious embodied interaction; here the user consciously plays the medium.

<1>Conclusions

While this vivid perception of reality within an acoustic virtual reality signifies a new type of listening experience enabled by AR technologies, the association between virtual audio and physical location is nothing new. The difference here is that this association is explicitly authored. The connection between the virtual audio content and the physical reality is predetermined, finely tuned, and designed, without a reliance on serendipitous connections between the virtual and the real as outlined by Michael Bull [18]. The audio-augmented object reality, like the locative audio experience, can place sound within a specific reality, the difference being between it having the potential to happen and it being designed to happen.

Although, while the locative audio experience has the potential to effectively reframe a geographic location to a practitioner's specifications, the audio-augmented object reality has the potential to effectively reframe both location and object to a practitioner's specifications. As with the locative audio experience, one of the key aspects is the specification of function through authorship that, in the case of the audio-augmented object reality, can be realized at the level of individual objects.

Ultimately, and more generally, the character of this new acoustic virtual reality is a product of its physicality and its perceived reality. This combination enables, across contexts, a virtual acoustic environment that can be perceived as being within the listener's reality rather than as a virtual acoustic environment that the listener perceives as being somewhere else. This creates an acoustic virtual reality that is *here and now*, rather than *there and then*. As such, the new acoustic virtual reality that is realized through the creation of audio-augmented objects has potential for creating new and highly immersive mixed reality experiences with virtual audio, including dramatical, musical, and gaming experiences within which the listener's physical reality plays a key role.

<ACKH>Acknowledgments

<ACK>Many thanks to Steve Benford, Chris Greenhalgh, James Mansell, Adrian Hazzard, and Annie Jameson. The author also thanks the two anonymous reviewers for their comments on earlier versions of this article. The author would also like to acknowledge his current research funding: XTREME---Mixed Reality Environment for Immersive Experience of Art and Culture, HORIZON Project ID: 101136006.

<REFH>References and Notes

- <REF>1. Jonathan Sterne, *The Audible Past: Cultural Origins of Sound Reproduction* (Duke University Press, 2003), 338.
2. Laurence Cliffe, *Alien Encounters* (2019), <https://lozcliffe.com/alien-encounters/> (accessed 8 May 2024).
3. Lanfranco Aceti. “Not Here, Not There: An Analysis of An International Collaboration To Survey Augmented Reality Art,” *Leonardo Electronic Almanac* 19, no. 1 (2013): 1--16.
4. Michael Bull, *Sounding Out the City: Personal Stereos and the Management of Everyday Life* (Berg, 2000).
5. Frauke Behrendt, “Locative Media as Sonic Interaction Design: Walking through Placed Sounds,” *Wi: Journal of Mobile Media* 9 (2015).
6. Laurence Cliffe, *The McMichael Experiment* (2020), <https://lozcliffe.com/the-mcmichael-experiment/> (accessed 8 May 2024).
7. Laurence Cliffe et al., “Materialising Contexts: Virtual Soundscapes for Real-world Exploration,” *Personal and Ubiquitous Computing* 20 (2020): 1--16.
8. Behrendt, “Locative Media as Sonic Interaction Design.”
9. Laurence Cliffe et al., *The Audible Artefact: Promoting Cultural Exploration and Engagement with Audio Augmented Reality*, in *AM’19* (ACM Press, 2019): 1--7.
10. Laurence Cliffe, *Horror-Fi Me* (2020), <https://lozcliffe.com/horror-fi-me/> (accessed 8 May 2024).
11. “App metrics” Apple Developer, accessed 21 December 2020, <https://help.apple.com/app-store-connect/#/itc21781223f>.
12. Winifred Phillips, *A Composer’s Guide to Game Music* (MIT Press, 2014), 55.
13. Isabella van Elferen, *Gothic Music: The Sounds of the Uncanny* (University of Wales Press, 2012), 180.
14. van Elferen, *Gothic Music*, 180.
15. Cliffe, *Horror-Fi Me*.
16. Pierre Schaeffer, *Treatise on Musical Objects: An Essay Across Disciplines* (University of California Press, 1966).
17. Jonathan Sterne, “The mp3 as Cultural Artifact,” *New Media & Society* 8, no. 5 (2006): 825--842.
18. Bull, *Sounding Out the City*.

Glossary

binaural audio---Binaurally rendered audio uses HRTF (Head Related Transfer Function) profiles that can be standardized or specifically modeled for the listener. These emulate both the interaural time difference (ITD) and interaural level difference (ILD) responses of spatialized sound sources (the difference in time and frequency a sound source is perceived by the listener's two ears) to create a realistic perception of a sound positioned in three-dimensional space.

locative audio---Typically, an interactive audio experience that triggers and delivers to the listener specific segments of prerecorded audio content based on their current GPS coordinates.

simultaneous location and mapping (SLAM)---a specific feature of AR technology that allows the persistent positioning of virtual content within physical space beyond the field of view of the camera, and beyond the length of the AR session if needed. This is primarily achieved by obtaining and collecting an array of feature points of the physical environment that can be used to localize and re-localize the position of both the user and the virtual content as required.

six degrees of freedom (6DoF)---Refers to the freedom of movement in three-dimensional space and includes the three degrees of rotational movement (pitch, yaw, and roll) plus the three degrees of translational movement (surge, heave, and sway).

three degrees of freedom (3DoF): Refers to the freedom of movement in three-dimensional space and includes the three degrees of rotational movement (pitch, yaw, and roll).

<REC>Manuscript received 13 December 2023.

<BIO>LAURENCE CLIFFE is a researcher and artist based at the University of Nottingham's Mixed Reality Laboratory and Virtual and Immersive Production Studio. He received his PhD from the University of Nottingham in 2021.

<CAP>Fig 1. Hardware interface for *Alien Encounters*. (<c> Laurence Cliffe)

Fig 2. Listener's '*tuning-into*' *The McMichael Experiment* at the National Science and Media Museum, Bradford, UK. (<c> Laurence Cliffe)

Fig 3. Left: a representative spatial audio diagram of an authored *Horror-Fi Me* soundscape. Right: a screenshot showing the mobile application's "tagging" or authoring process. (<c> Laurence Cliffe)