Listening in Extended Reality:

Recorded Music in and as Virtual, Augmented and Mixed Reality

NICOLA DIBBEN

Accepted by Journal of the Royal Musical Society, 26 May 2025. Forthcoming Autumn 2026. Supplementary material: N. Dibben (2025). 'A scoping review of auditory augmented reality and auditory mixed reality: for the journal article "Listening in extended reality: Recorded music in and as virtual, augmented and mixed reality"' doi:10.15131/shef.data.29862941.

Abstract This paper critically examines *musical extended reality* (musical XR) — a new medium for recorded music — arguing that it reconfigures the ontology of the recorded work through spatiality, interactivity and embodiment. Synthesising insights from musicology, media theory and sound studies, musical XR are situated within the history of sound reproduction and a new conceptual framework is developed, differentiating virtual, augmented and mixed reality experiences. Case studies including Björk's *Stonemilker* (360° video), Meredith's *Moon Moons* (AR), Eno and Chilver's *Bloom: Open Space* (MR) and Hot Sugar's *Melody of Dust* (VR) help identify underexplored creative possibilities, illuminating XR's significance for twenty-first-century musical practice.

Over the past decade, *musical extended reality* (musical XR) has emerged as a new medium through which music is composed, performed and experienced. Drawing on extended reality (XR) technologies, musical XR operates across computer-mediated and hybrid environments that destabilise distinctions between the virtual and the real. XR encompasses immersive and interactive virtual reality (VR), which situates users within simulated spatialities; mixed reality (MR) which integrates synthetic media with the physical environment; and augmented reality (AR) which overlays or removes digital information from the perceptual field.[1] These technologies have developed within a broader socio-technological assemblage that has reconfigured conditions under which music is produced, distributed and experienced, facilitating increasingly spatial, interactive and participatory forms of engagement with recorded music.

Far from a mere technological novelty, musical XR extends the long history of recording as a site of mediation between sound, space and listener. It contributes to what has been termed a *musical metaverse*[2] — a convergence of virtual and physical spaces that extends the social and material sites of musical practice and reconfigures the temporal, spatial and embodied dimensions of musical experience. Working within, and extending, the infrastructures of the established

[1] *CTA Standard Definitions and Characteristics of Augmented and Virtual Reality Technologies*, CTA-2069-A (Consumer Technology Association, 2020).
[2] Luca Turchet, 'Musical Metaverse: Vision, Opportunities, and Challenges', *Personal and Ubiquitous Computing*, published online 13 January 2023, doi:10.1007/s00779-023-01708-1.

recording industry, XR-based works introduce new modes of authorship and spectatorship, foregrounding the recorded artefact as an interactive and spatially dynamic experience.

This article investigates these developments through an analysis of recorded XR artefacts, referred to in industry discourse as XR 'musical content'. Examples such as *Björk Digital* (2015) 360° VR music videos, TOKiMONSTA's *Lune Rouge VR* (2017), and Sigur Rós' *Tónandi* (2018) for Magic Leap One mixed-reality headset illustrate how recorded music has served both as a testbed and a promotional vehicle for consumer XR technologies, continuing the practice of using musical content to promote hardware.[3] These 'audio-first' experiences, as music technologist Luca Turchet calls them,[4] feature spatially persistent and interactive virtual auditory elements,[5] ranging from fixed playback to procedural generation and sensor-based responsiveness. While the act of listening to recorded music has remained remarkably stable for the last 50 years — typically a linear experience mediated by stereo playback — musical XR presents something of a paradigm shift towards situated, embodied and participatory modes of engagement.[6]

By situating musical XR within the longer history of recording, mediation and listening, I ask how XR technologies challenge the conceptual boundaries of the recorded work. I argue that these practices compel a rethinking of what it means to record, to perform and, for the purposes of this article, to listen in an age where sound itself is spatially and interactively reconstituted.

Section 1: Conceptualising Listening to Recorded Music in Musical XR

1.1 Auditory virtual reality

Auditory virtual reality (AVR) comprises the simulation of a spatialised auditory environment and/or the movement of virtual objects, agents and sound sources[7] within it, with which the

[3] The idea that 'good content' is needed to sell hardware is the same strategy that underpinned smartphone and tablet marketing, either by the draw of a major artist or other kinds of compelling music experiences. The short duration of pop songs may also have been particularly compatible with early stage XR development since it avoided the physical discomfort associated with wearing/holding equipment for long durations.
[4] Luca Turchet, Rob Hamilton, and Anil Çamci, 'Music in Extended Realities', *IEEE Access*, 9 (2021), pp. 15810–32, doi:10.1109/ACCESS.2021.3052931; Anil Çamci and Rob Hamilton, 'Audio-First VR: New Perspectives on Musical Experiences in Virtual Environments', *Journal of New Music Research*, 49.1 (2020), pp. 1–7, doi:10.1080/09298215.2019.1707234.
[5] Turchet, Hamilton, and Çamci, 'Music in Extended Realities', p. 15815.
[6] Turchet, 'Musical Metaverse'.
[7] Eric Clarke, 'Music, Space and Subjectivity', in *Music, Sound and Space: Transformations of Public and Private Experience*, ed. by Georgina Born (Cambridge University Press, 2017), pp. 90–110; Ragnhild Brøvig-Hanssen and Anne Danielsen, *Digital Signatures: The Impact of Digitization on Popular Music Sound* (The MIT Press, 2016),

user/listener can interact. The definition of AVR as technologically-driven, interactive, spatial simulations is distinct from *virtuality* as used in musicological discourse[8] to mean something present in its effect but not its actuality, and encompassing a range of media.[9] Adopting a more precise formulation allows a critical assessment of the affordances of virtual reality for listening to recorded music.

The current technological foundation for AVR is spatial audio, which creates the impression of being surrounded by sound in 360° in a way which approximates the natural listening environment.[10] Spatial audio can be realised using a variety of different formats (e.g. objects, channels, or scene-based) for different audio rendering devices (e.g. speakers, headphones).[11] Two common means to create a realistic perception of this three-dimensional sound space are binaural recording and rendering[12] and soundfield recording and synthesis.[13] From a phenomenological perspective, 360° spatial audio gives rise to a sense of virtual sound objects and/or environments located in the world, and can integrate sensorimotor contingencies by rendering sound according to the listener's position and movement, contributing to a sense of presence. Head tracking and other kinds of interactivity contribute to this: a recent empirical study showed that instrumentalists moved more and preferred spatialised sound with head

doi:10.7551/mitpress/10192.001.0001; Allan F. Moore, *Rock, the Primary Text: Developing a Musicology of Rock*, Ashgate Popular and Folk Music Series, Third edition (Routledge/Taylor & Francis Group, 2019); Dennis Smalley, 'Space-Form and the Acousmatic Image', *Organised Sound*, 21.1 (2007), pp. 35–58.

[8] Shara Rambarran, *Virtual Music: Sound, Music, and Image in the Digital Era* (Bloomsbury Academic, 2021); Tim Summers and others, 'Music and Sound in Virtual/Augmented Realities—Questions, Challenges and Approaches', *Journal of Sound and Music in Games*, 2.2 (2021), pp. 63–83, doi:10.1525/jsmg.2021.2.2.63; *The Oxford Handbook of Music and Virtuality*, ed. by Shana Rambarran and Sheila Whiteley (Oxford University Press, 2016); Isabella van Elferen, '¡Un Forastero! Issues of Virtuality and Diegesis in Videogame Music', *Music and the Moving Image*, 4.2 (2011), p. 30, doi:10.5406/musimoviimag.4.2.0030; Mark Grimshaw and Tom Garner, *Sonic Virtuality: Sound as Emergent Perception* (Oxford University Press, 2015).

[9] Frances Dyson, *Sounding New Media: Immersion and Embodiment in the Arts and Culture* (University of California Press, 2009), p. 6. She makes the point that '"virtuality" often refers to both the set of technologies that constitute digital media and the ontological state or condition of virtuality'.

[10] *Immersive Sound: The Art and Science of Binaural and Multi-Channel Audio*, ed. by Agnieszka Roginska and Paul Geluso (Routledge, Taylor & Francis Group, 2018).

[11] *Immersive Sound*, p. 6; Gareth Llewellyn and Justin Paterson, 'Towards 6DoF: 3D Audio for Virtual, Augmented, and Mixed Realities', in *3D Audio* (Routledge, Taylor & Francis Group, 2022), pp. 43–63.

[12] The binaural system is commonly used in AR/VR products (e.g. the Oculus Rift, Playstation VR). It simulates the effect of the human two-eared auditory system by recording audio through separated headphones, either at the head of a listening subject or a dummy head, and using a range of localisation cues (including head related transfer functions (HRTFs) and dynamic cues from movement of the listener), to render the sound through headphones or a small number of loudspeakers (e.g. a two-channel stereo recording). Spatial audio effects are generated either by using the recordings themselves or by applying the HRTFs for the virtual source positions to a mono signal.

[13] Wen Zhang and others, 'Surround by Sound: A Review of Spatial Audio Recording and Reproduction', *Applied Sciences*, 7.5 (2017), p. 532, doi:10.3390/app7050532.

tracking to a stereo rendering, when playing with other musicians.[14] The authors attribute this to a mutually reinforcing sense of presence aided by an embodied action-perception loop. In this regard, AVR differs from mono or stereo recordings: in standard sound reproduction systems one cannot move within or between spatially located sound objects, nor play with associated proximity effects. This is partly because mono and stereo recorded sound is for the most part not registered in the world but is head-locked, i.e. they move with the user's head or device position no matter the user's orientation. An exemplar of this is headphone listening in which the location of the sounds (and their virtual sound sources) changes with the head movements of the headphone wearer: as soon as the listener moves their head, and perhaps even before, the virtual space of the recording is revealed to exist on an axis that is experienced as lying within, or running through the listener's own head rather than being located 'out there' in the world.

Precursors to AVR exist in earlier performance and recording which simulate and synthesise virtual space and movement, particularly in electroacoustic acousmatic music, popular music production, sound art installations, and also to some extent in acoustic and live performance.[15] These kinds of pre-computational techniques include the distribution of sound sources around a performance space to elicit spatial effects: for example, *cori spezzati* of Renaissance antiphony positioned groups of singers in ways which helped articulate musical structures; in opera, off-stage instruments and voices cue audiences' perceptions of diegetic distance; in chamber music, staggered entries of instruments simulate echos and their associated virtual spaces; and speaker placement and mix has been used to create navigable sound installations, such as in Janet Cardiff's 40-speaker re-spatialisation of Thomas Tallis' *Spem in Allium.*[16]

Virtual space is also a feature of stereo and mono sound. David Patmore and Eric Clarke suggest that recordings can act as 'gateways' into a virtual world of instruments and voices[17]; sound sources in that virtual world can be heard as occupying their own space in relation to each other

---

[14] Matteo Tomasetti and Luca Turchet, 'Playing With Others Using Headphones: Musicians Prefer Binaural Audio With Head Tracking Over Stereo', *IEEE Transactions on Human-Machine Systems*, 53.3 (2023), pp. 501–11, doi:10.1109/THMS.2023.3270703.
[15] Clarke, 'Music, Space and Subjectivity'; Pedro Rebelo, 'Sound and Space: Learning from Artistic Practice', in *3D Audio*, ed. by Justin Paterson and Hyunkook Lee (Routledge, Taylor & Francis Group, 2022), pp. 192–206; Barry Blesser and Linda-Ruth Salter, *Spaces Speak, Are You Listening? Experiencing Aural Architecture* (MIT Press, 2007); Matthew Wilson Smith, *The Total Work of Art: From Bayreuth to Cyberspace* (Routledge, 2007).
[16] Janet Cardiff, *Forty Part Motet* (2001).
[17] David N. C. Patmore and Eric F. Clarke, 'Making and Hearing Virtual Worlds: John Culshaw and the Art of Record Production', *Musicae Scientiae*, 11.2 (2007), pp. 269–93, doi:10.1177/102986490701100206.

and to the listener, due to acoustic cues of loudness and reverberation which form part of that recording:

> Listeners at home hearing the size of a recorded space or position of a voice are detecting the attributes of a virtual space – a space specified by the same perceptual attributes as a real space, but which is not physically present at the time.[18]

The technology and practices of 'detachable echo'[19]/'detachable ambience'[20] created the ability to take the characteristic ambience of one space and replay it in another,[21] simulate reality,[22] create performances in computationally modelled reconstructions of lost or inaccessible locations,[23] and present spatial characteristics that would not/could not happen in the real world.[24] From the first credited use of the term 'virtual acoustic space' by electroacoustic composer Trevor Wishart,[25] through to subsequent scholars' reference to the mono and stereo 'sound stage',[26] 'soundbox',[27] or *lydrom*[28] (soundroom/soundspace), music reproduction has been likened to a form of virtual reality,[29] despite the absence of a key distinguishing feature of AVR, namely interactivity.[30]

So, why has listening to recorded music so often been deemed to give rise to the experience of virtuality even when interactivity is absent? One factor is the influence of normative listening

---

[18] Clarke, 'Music, Space and Subjectivity', p. 95.
[19] Jonathan Sterne, 'Space within Space: Artificial Reverb and the Detachable Echo', *Grey Room*, 60 (2015), pp. 110–31, doi:10.1162/GREY_a_00177.
[20] Paul Roquet, 'Acoustics of the One Person Space: Headphone Listening, Detachable Ambience, and the Binaural Prehistory of VR', *Sound Studies*, 7.1 (2021), pp. 42–63, doi:10.1080/20551940.2020.1750270.
[21] R. Murray Schafer, *The Tuning of the World: Toward a Theory of Soundscape Design*, Paperback ed (University of Pennsylvania Press, 1977), p. 91.
[22] Allan F. Moore and Ruth Dockwray, 'The Establishment of the Virtual Performance Space in Rock', *Twentieth-Century Music*, 5.2 (2008), pp. 219–41, doi:10.1017/S1478572209990065; Peter Doyle, *Echo and Reverb: Fabricating Space in Popular Music Recording, 1900-1960*, Music/Culture, 1st ed (Wesleyan University Press, 2005); Patmore and Clarke, 'Making and Hearing Virtual Worlds'.
[23] Kenneth B. McAlpine, James Cook, and Rod Selfridge, 'Hearing History: A Virtual Perspective on Music Performance', in *3D Audio*, ed. by Justin Paterson and Hyunkook Lee (Routledge, Taylor & Francis Group, 2022), pp. 207–27.
[24] Clarke, 'Music, Space and Subjectivity'; Ragnhild Brøvig-Hanssen and Anne Danielsen, 'The Naturalised and the Surreal: Changes in the Perception of Popular Music Sound', *Organised Sound*, 18.1 (2013), pp. 71–80, doi:10.1017/S1355771812000258.
[25] Clarke, 'Music, Space and Subjectivity', p. 98.
[26] William Moylan, *The Art of Recording: The Creative Resources of Music Production and Audio.* (Van Nostrand Reinhold, 1992); Serge Lacasse, '"Listen to My Voice": The Evocative Power of Vocal Staging in Recorded Rock Music and Other Forms of Vocal Expression.' (unpublished, University of Liverpool, 2000).
[27] Moore, *Rock, the Primary Text*.
[28] Anne Danielsen, '"My Name Is Prince" – En Studie i Diamonds and Pearls' (unpublished Masters Thesis, University of Oslo, 1993); Anne Danielsen, 'His Name Was Prince: A Study of *Diamonds and Pearls*', *Popular Music*, 16.3 (1997), pp. 275–91, doi:10.1017/S0261143000008412.
[29] Brøvig-Hanssen and Danielsen, *Digital Signatures*, p. 154.
[30] Turchet, Hamilton, and Çamci, 'Music in Extended Realities'.

practices on the perception of virtuality. Modes of contemplative listening direct listener attention to music's acousmatic character: a darkened performance space and seated, silent concert listening to music at a volume which masks other sounds are practices which afford aesthetic contemplation of sound presented from performers/speakers rather than the (rest of the) real acoustic environment, and enables the music's spatial cues to dominate. These practices do not afford active exploration of the auditory environment with the consequence that awareness of the non-interactive character of the auditory environment is reduced.

A second factor, implicit in the first, is the suppression of acoustic transparency. The design of listening spaces and speaker systems can suppress acoustic transparency, promoting the illusion of a virtual space.[31] For example, physical speaker layouts for domestic hi-fi listening delineate a 'sweet spot' for optimal listener experience of an acoustic virtuality; the design aesthetic of unobtrusive/hidden speakers which helps erase the source of speaker sound;[32] and ideologies of contemplative listening focus attention on sound rather than source.[33] These socio-material characteristics afford the experience of being immersed in sound or, indeed, being flooded with sound in the case of headphones.[34]

For Gascia Ouzounian, the very idea of spatial listening is a modern construct, which she traces through key moments in Northern Hemisphere, Western scientific and artistic thought and practice from the nineteenth century, within which AVR is a continuation of this logic and a normalisation of spatial hearing.[35] Ultimately, a recording's status as auditory virtual reality hinges on its spatialisation, interactivity and its ability to occlude the listener's real environment. But what about experiences that don't replace reality, but instead augment it by combining

---

[31] David Prior, 'Loudspeaker Listening: Tabula Rasa or Augmented Reality', *Leonardo Music Journal*, 26 (2016), pp. 3–6 (p. 3), doi:10.1162/LMJ_a_00957.

[32] Prior argues that the erasure of the physical objecthood of the loudspeaker is consistent with the affordances of stereo sound through which a virtual acoustic space exists between the speakers rather than at the source of a mono sound. Cinema sound functioned differently from hi-fi sound in this regard because it retained a central channel for voice, something still found today in sound design of audio-visual VR experiences.

[33] Notably, the affordances of loudspeakers for augmentation rather than replacement of the real acoustic environment have been neglected within scholarship as well as listening practices, no doubt for related reasons, namely the valorisation of the autonomous musical work and scholarly denigration of ubiquitous music and associated 'background listening' practices.

[34] Jacob Kingsbury Downs, 'Headphones, Auditory Violence and the Sonic Flooding of Corporeal Space', *Body & Society*, 27.3 (2021), pp. 58–86, doi:10.1177/1357034X211024352.

[35] Gascia Ouzounian, *Stereophonica: Sound and Space in Science, Technology, and the Arts* (The MIT Press, 2020), p. 16. Ouzounian has argued that spatial hearing 'dominates cultures of listening' (p.36). Divergent socio-technological trajectories, such as the return to mono sound with smartphones and smart speakers, warrant further investigation.

real-world and virtual sounds? It is to these experiences, of auditory augmented- and mixed-reality, that I turn next.

1.2 Auditory augmented- and mixed-reality

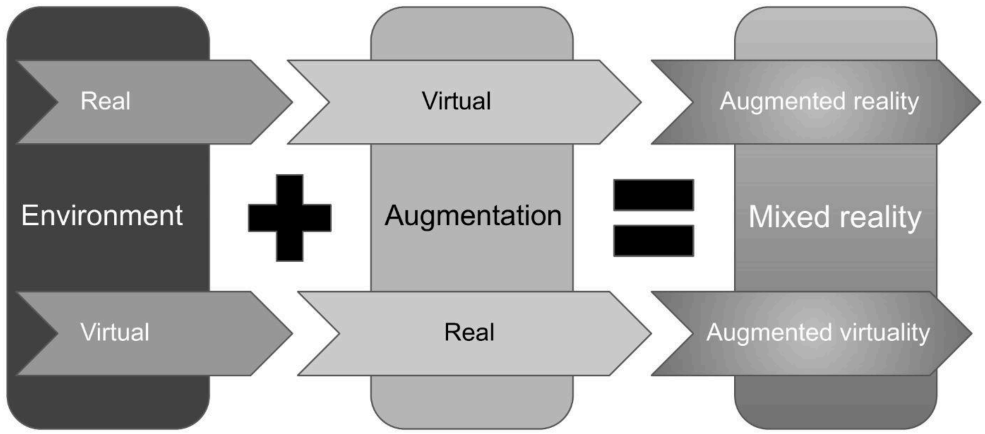
1.2.1 AR/MR and acoustic transparency

The mixed real-and-virtual percept of an AR/MR experience can be thought of in terms of layers of reality,[36] comprising two elements: first, the spatial surroundings a user is intended to be immersed in, which can be real or virtual (virtual in those situations where environmental stimuli in at least one sensory mode have been entirely replaced by virtual content); and second, the augmentation of those surroundings by computer-synthesised addition to- or deletion from the real environment ('augmented reality'[37]) or the addition of real-world sensory information to the virtual environment ('augmented virtuality') (Figure 1). This hybrid experience relies on the perception of acoustic transparency — the simultaneous hearing of real-world sound and augmented sound. This is often achieved technologically through active noise cancellation (e.g. hear-through headphones/hearables/wearables) or passive acoustic transparency (e.g. where the ear canal is open to real/virtual sound (e.g. bone conduction, non-contact wearable headsets using speakers),[38] with the mixing of real and computer-generated sound happening at different points in the pathway between the sound source and the perceiver. This mixing can happen at the point of environmental display (e.g. speakers), or sensory subsystem (e.g. acoustic hear-through, whereby computer generated sound is delivered by bone conduction and real-world sound through open ear canals) or computer (microphone-hear-through, whereby environmental sound is picked up by a microphone and relayed to the user along with computer-generated sounds).[39]

---

[36] Gheric Speiginer and Blair Maclntyre, 'Rethinking Reality: A Layered Model of Reality for Immersive Systems', *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, October 2018, pp. 328–32, doi:10.1109/ISMAR-Adjunct.2018.00097.
[37] Paul Milgram and Fumio Kishino, 'A Taxonomy of Mixed Reality Visual Displays', *IEICE Transactions on Information Systems*, E77-D.12 (1994), p. 15 (p. 15).
[38] Mark McGill and others, 'Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality', *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 21 April 2020, pp. 1–16, doi:10.1145/3313831.3376702.
[39] Robert W. Lindeman and Haruo Noma, 'A Classification Scheme for Multi-Sensory Augmented Reality', *Proceedings of the 2007 ACM Symposium on Virtual Reality Software and Technology - VRST '07*, 2007, p. 175, doi:10.1145/1315184.1315216.

Figure 1. Diagram showing how layering of the real and the virtual produces varieties of augmentation.



The rise of consumer technologies for acoustic transparency has been accompanied by considerable terminological and conceptual inconsistency in public and academic discourse. This paper addresses this confusion by drawing on a scoping review of audio-focused augmented, mixed and virtual reality (see Supplementary Materials[40]). As this review highlights, there is a proliferation of terms and a lack of adherence to a single conceptual framework when referring to AR and MR in the auditory domain. A key instance of the resultant confusion is that what are commonly called 'music AR' (or even VR) experiences in popular discourse, are most often *visual* augmentations rather than sonic ones.[41] For example, Popins webAR,[42] a QR code-triggered smartphone app, renders an audio-visual recorded performance within the user's surroundings such that the performer's visual image is anchored to a surface in the user's

[40] 'A scoping review of auditory augmented reality and auditory mixed reality: Supplementary material for the journal article "Listening in extended reality: Recorded music in and as virtual, augmented and mixed reality"' doi:10.15131/shef.data.29862941. The scoping review was too substantial to be included in this article but it forms part of the underpinning research for it.
[41] Visually augmented reality occupies a variety of product niches in the music industry, although inconsistent labelling often belies their characteristics: AR packaging, in which a QR code or similar triggers visual and sometimes audio content rendered through a smartphone; heritage tours and sound/audio walks which guide listeners through a route; AR training apps, which superimpose visual information onto a musical instrument or sheet music; 'VR music videos', which most often comprise 360° ('spherical') linear videos viewed via a headset or the web; and 'AR music videos', comprising a linear video of a performance superimposed through a smartphone camera onto the viewer's immediate visual environment. Holographic performers, events in online spaces and interactive performances which add content to live performance ('augmented performance', 'augmented listening', and the 'augmented stage') lie beyond the scope of this paper which focuses on sound rather than XR visuals.
[42] *Popins* (n.d.) <https://www.popins.io> [accessed 28 October 2022].

environment as viewed through the smartphone camera, and is therefore spatially persistent congruent with the listener. However, the audio is rendered as mono/stereo, moves with the user's head, and masks the sound of the environment. So, while the app can be described as presenting visual AR, the audio is rendered in standard fashion.

The scoping review also highlights inconsistent differentiation between augmented and mixed reality audio experiences in the literature. While the term 'audio [or auditory] augmented reality' (AAR) is the most frequently used (mirroring its visual equivalent, 'visual AR'[43]) the precise differences between AAR and mixed reality experiences (AMR) are often unclear. To provide greater terminological consistency I propose the umbrella term 'auditory AR/MR'[44] for situations involving acoustic transparency that merge real and virtual auditory information. The next section considers auditory AAR and AMR in more detail.

[43] Where the intention is to refer to the content of the experience, rather than the medium, then the term 'augmented reality audio' is grammatically appropriate.

[44] Use of the terms AR and MR is inconsistent therefore I use the term 'AR/MR' to refer to the range of experiences and technologies combining real and virtual environments within the same modality (as described by Milgram and Kishino's notion of 'mixed reality') and distinct from 'virtual reality' (VR) in which perceptual information in one modality is replaced by computer synthesised perceptual information.

Table 1. Definitions of AR, MR, VR adapted from *CTA Standard Definitions and Characteristics of Augmented and Virtual Reality Technologies*, CTA-2069-A (Consumer Technology Association, 2020), pp. 1–2.

Term	Definition
Augmented Reality (AR)	Overlays digitally-created content into the user's real-world environment. AR experiences can range from informational text overlaid on objects or locations to interactive photorealistic virtual objects. AR differs from Mixed Reality in that AR objects (e.g., graphics, sounds) are superimposed on, and not integrated into, the user's environment.
Mixed Reality (MR)	Seamlessly blends a user's real-world environment with digitally-created content, where both environments coexist to create a hybrid experience. In MR, the virtual objects behave in all aspects as if they are present in the real-world – e.g., they are occluded by physical objects, their lighting is consistent with the actual light sources in the environment, they sound as though they are in the same space as the user. As the user interacts with the real and virtual objects, the virtual objects will reflect the changes in the environment as would any real object in the same space.
Virtual Reality (VR)	Is a fully immersive user environment affecting or altering the sensory input(s) (e.g., sight, sound, touch, and smell) and can allow interaction with those sensory inputs by the user's engagement with the virtual world. Typically, but not exclusively, the interaction is via a head-mounted display, use of spatial or other audio, and/or motion controllers (with or without tactile input or feedback).

1.2.2 Auditory augmented reality

Auditory augmented reality (AAR) overlays digitally-created musical content onto objects or locations in the user's real-world acoustic environment. For example, the combination of geo-location and computer sensors means that soundart can be superimposed onto the listener's environment and it can change dynamically with data input from that environment.[45]

A historical precursor to this is the medium of sound walks which emerged in the late twentieth century and whose formats include headphone-transmitted audio, walking pieces which require the listener to manually playback recordings and/or to make music/sound in specific locations.[46] For example, Janet Cardiff's soundwalks and audio tours combine spoken narration with pre-recorded sounds placed in site-specific locations they originate in.[47] This design uses the

---

[45] Ignacio Pecino and Ricardo Climent, 'Sonicmaps: Connecting the Ritual of the Concert Hall with a Locative Audio Urban Experience', *International Computer Music Conference Proceedings*, 2013, pp. 315–20.
[46] Frauke Behrendt, 'The Sound of Locative Media', *Convergence: The International Journal of Research into New Media Technologies*, 18.3 (2012), pp. 283–95, doi:10.1177/1354856512441150.
[47] Juliana Hodkinson, 'Creating Headspace: Digital Listening Spaces and Evolving Subjectivities', *Musicology Research*, 3 (2017), pp. 163–77.

affordances of headphone listening for virtual reality (the replacement of the real auditory environment by a virtual one from sound recordings) to create a kind of augmented reality; the pre-recorded sounds in this case augment/add to those of the real auditory environment, albeit those sounds have been pre-recorded. Another example, this time from listening rather than compositional practices, is the practice of playing back recorded music which incorporates found sound in contexts where those sounds might naturally occur, thereby creating a hybrid sonic experience.

Jonathan Sterne has identified listening over loudspeakers, as another precursor to AAR because it can allow one to hear space within space:

> Recordings that bear a mix of sonic signatures that would be impossible to hear in a live performance resound out of speakers in a room. The audio then becomes part of the sonic space of the room alongside other noises [...] This sonic scenario is an almost textbook case of the effect that visual augmented- and mixed-reality practitioners are attempting to accomplish.[48]

However, the ability to spatially locate the synthetic relative to the real, characteristic of AAR, is more readily realisable in some circumstances than others. In the case of background music broadcast in a supermarket, for instance, the small dynamic range of the compressed sound, its few noticeable transitions, and homogeneous and flattened ambient quality, all mitigate against perception of its spatial location. By contrast, listening to an unaccompanied Bach Partita, recorded with the ambience of the church it was performed in, over stereo speakers in a domestic kitchen could be experienced as the intrusion of a separate acoustic environment superimposed onto, and heard along with, the playback space. Were it not so normalised, so readily assimilated as part of the local soundscape, and resistant to exploration (one can't actively explore the spatial relationships between sound sources within the mix through head or bodily movement) it could be regarded as the auditory equivalent to the AR (visual) portal — a virtual doorway/gateway through which one can step into a different, virtual world. Nonetheless, these historical precursors are sufficient to lead Jonathan Sterne to conclude that 'The overlay of physical and mediatic space in digital media *has already happened* in the sonic domain' and that therefore 'In their apprehension of multiple spaces within a single space, with their multi perspectival

[48] Sterne, 'Space within Space', p. 120.

perceptions, the subjects of augmented and mixed reality will not be radically different from the media subjects we already know, such as the audiences who have attended to radio and popular music for close to a century'.[49]

1.2.3 Auditory mixed reality

Influenced by Azuma's seminal papers,[50] many scholars regard MR as an umbrella term for any VR or AR experience. However, there is some agreement that AR/MR experiences are differentiated by the extent of world knowledge and interactivity that the system manifests[51] and I use the term in this more precise formulation, drawing on contemporary definitions.[52] According to this perspective, sound in AMR behaves as if it is present in the real world, i.e. it can be masked by other sounds, its reverberation is consistent with the actual environment, they sound as if they are in the same space as the listener, and as the listener interacts with the virtual sound it reflects those changes in the same way as a real sound would. For example, an AMR experience might involve exciting an object in the real-world which then produces a virtual sound, which might be an addition, modification or subtraction from its original sound which changes according to changes in the environment as would a real object. Hence, MR is premised on a system that has knowledge of the environment it is operating in, can adapt to changes in it, and supports the user's actions and sense of being present in that world.[53] The world knowledge that is implied here encompasses both place (whether that be specific to a geographical location and has modelled that environment, or whether its spatialisation is translocal) and space (its size, materials, the listeners' location within it and how these change dynamically). Moreover, the extent to which the system supports the user's sense of presence in that environment is partly

[49] Sterne, 'Space within Space', p. 120.
[50] Ronald T Azuma, 'A Survey of Augmented Reality', *Presence*, 6.4 (1997), pp. 355–85; R. Azuma and others, 'Recent Advances in Augmented Reality', *IEEE Computer Graphics and Applications*, 21.6 (2001), pp. 34–47, doi:10.1109/38.963459.
[51] Richard Skarbez, Missie Smith, and Mary C. Whitton, 'Revisiting Milgram and Kishino's Reality-Virtuality Continuum', *Frontiers in Virtual Reality*, 2 (2021), doi:10.3389/frvir.2021.647997; Hanna Schraffenberger and Edwin Van der Heide, 'Everything Augmented: On the Real in Augmented Reality', *Journal of Science and Technology of the Arts*, 6.1 (2014), p. 17, doi:10.7559/citarj.v6i1.125.
[52] 'Definitions and Characteristics of Augmented and Virtual Reality Technologies (CTA-2069-A)', Consumer Technology Association®, n.d., p. 1 <https://shop.cta.tech/products/definitions-and-characteristics-of-augmented-and-virtual-reality-technologies> [accessed 10 January 2023].
[53] Azuma, 'A Survey of Augmented Reality'; Milgram and Kishino, 'A Taxonomy of Mixed Reality Visual Displays'; Skarbez, Smith, and Whitton, 'Revisiting Milgram and Kishino's Reality-Virtuality Continuum'.

determined by its ability to allow the user to act within and on it as they would with real world sound. Existing composition and production practices of place, space and interactivity in music highlight continuities and breaks between mixed reality in the auditory domain and recorded musical practices.

Place-based, (located/site-specific) music is often associated with ritual, historical performance practice and late twentieth- and twenty-first century soundart.[54] By contrast, performance and listening conventions for a large proportion of music assume that musicking can happen in different geographical locations without detriment, albeit in performance spaces with relevant kinds of acoustic characteristics. The advent of mobile computing and Global and Visual Positioning Systems since the second decade of the 21st century have enabled development of geo-located XR, and these now constitute an established medium for site-specific music and soundart. Many of these are realised as apps enabling geo-located renderings of mono/stereo sound recordings for headphone playback from a mobile device as the listener moves through a location (e.g. Bluebrain *Central Park*[55] and *Music for Trees*[56]). An underexplored opportunity is acoustically transparent and location aware music/sound art which augments and/or removes and/or interacts with elements of environmental sound. This may be a key potential of AMR since it exploits the affordances of acoustic transparency, using the sounds of that real environment and adding to, removing, and/or altering them.

The spatial component of mixed-reality music manifests by sounding as though it originates in the same space as the listener, and by changing congruently with that environment (e.g. being masked by other intervening objects or sounds) and with the position and actions of the listener. This interactivity connects with existing musical practices in which the listener can move around or through the music,[57] rather than the music moving around the (static) listener (e.g. diffusion concerts of electroacoustic music). Composer Nye Parry argues that such translocal, locative (i.e. mobile) experiences enable listeners to experience the open work in other ways than the

---

[54] Ouzounian, *Stereophonica*.
[55] James C. McKinley Jr, 'Central Park, the Soundtrack', Arts, *The New York Times*, 7 December 2011.
[56] Matt Steinmann, *Music for Trees Mobile App: A Collaboration between The Royal Parks and the Royal Academy of Music* (The Royal Parks, 2019) <https://www.royalparks.org.uk/parks/the-regents-park/things-to-see-and-do/music-for-trees-mobile-app>.
[57] Frauke Behrendt, 'The Sound of Locative Media'.

normative linear, one-time realisation.[58] For him, the real environment in locative experiences can provide reference points ('landmarks') by which the listener can build a cognitive representation of the virtual sound space, orient themselves and navigate within it.[59] Similarly, most translocal experimental and pop music apps use visual AR to provide virtual markers for sound sources. There remains untapped potential for AMR experiences which use the listeners' movement through space as a means to experience real and virtual sound in situated experiences.

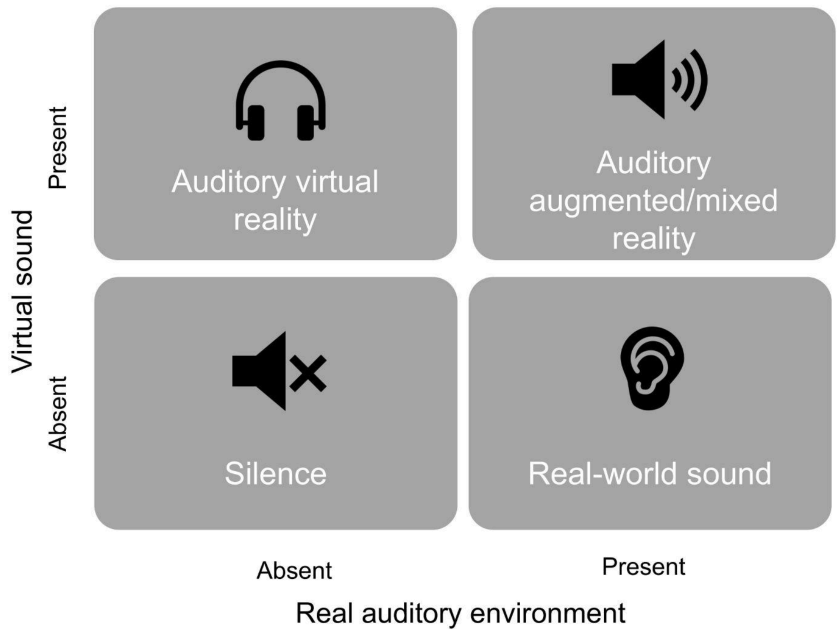
1.2.4 A framework for musical XR

I have identified principles useful to differentiating and characterising audio-first recorded music XR experiences from a listener perspective. XR comprises some addition to, mixing with, or replacement of sensory information, which may be specific to a particular site or translocal. Two dimensions can serve as the foundation of a typology for musical XR: the presence/absence of real-world sound, and the presence/absence of virtual (computer synthesised) sound. Applying these two dimensions produces a matrix of types of auditory experience (Figure 2): the absence of real-world and virtual sound manifests as *silence*; the presence of sound from the real world without augmentation constitutes *real-world sound*; the presence of both real-world sound and spatialised addition or modification to that sound comprises *auditory augmented or mixed reality*; the occlusion of real-world sound by spatialized, interactive and virtual sound is *auditory virtual reality*.

---

58 Nye Parry, 'Navigating Sound', in *The Oxford Handbook of Interactive Audio*, ed. by Karen Collins, Bill Kapralos, and Holly Tessler (Oxford University Press, 2014), pp. 31–44 (pp. 31–32).
59 Parry, 'Navigating Sound', p.34.

Figure 2. Types of auditory XR according to presence/absence of sound and presence/absence of augmentation.



Whereas 'auditory virtual reality' (AVR) denotes the occlusion and replacement of real-world sound by audio, AAR/AMR describes acoustic transparency in which real-world sound and audio are heard together; and while some scholars use the term 'MR' as a larger category encompassing AR and VR,[60] to achieve more fine grained differentiation, and in line with industry practice I propose the term 'Auditory mixed reality' (AMR) for those situations involving computer perception/knowledge of the world and human input in order to blend real and virtual auditory content.[61] In some cases this will be site-specific, mapping and augmenting sound to and from the real world, and registered in three-dimensions, while in others it will be translocal – music which retains spatial relationships between sounds but in which the actual place of performance may be redundant. I have also highlighted the need to distinguish in which sense modalities any virtual content occurs: in the case of consumer XR involving music, the virtual content is currently most often visual, or visual and auditory.

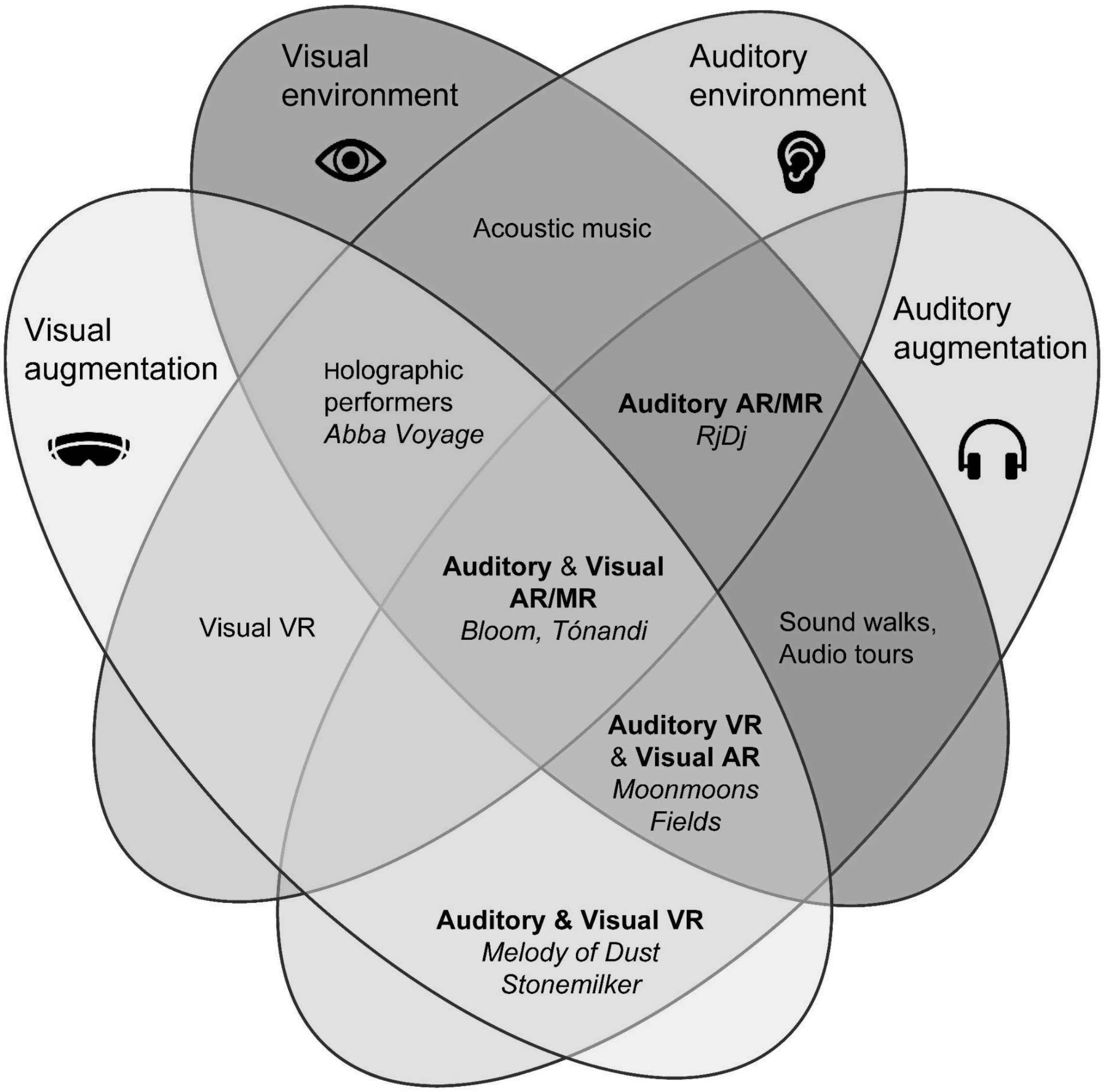
---

60 McGill and others, 'Acoustic Transparency and the Changing Soundscape of Auditory Mixed Reality'.
61 'Definitions and Characteristics of Augmented and Virtual Reality Technologies (CTA-2069-A)'.

Figure 3 presents a listener-centric typology of XR afforded by these combinations of visual and auditory real and simulated content, populated with illustrative examples. Live music performance is conceptualised within this framework as a form of simultaneously experienced (unaugmented) visual and auditory sensory information from the real environment. At the other extreme, a listener whose real-world visual and auditory content is replaced by computer synthesised content would be experiencing virtual reality, such as in the music VR experience *Melody of Dust*, discussed later.[62] Superimposition of speaker sound can be regarded as affording the perception of an augmented auditory reality for the reasons given above: presentation of 'space within space', where the lack of interactivity is masked by socio-technological constraints on perceptual information that would 'give the game away'. While recognising that this is a more inclusive notion of auditory augmented reality than common technocentric definitions, it captures the continuities of perceptual experience with other forms of (stereo) recorded music listening. As well as enabling particular musical media to be contextualised within XR as a whole, the framework helps identify gaps in the kinds of experiences so far explored by creators: for example, auditory AR/MR making use of acoustic transparency is relatively underpopulated as a musical type.

[62] *The Melody of Dust*, dir. by Hot Sugar (Viacom NEXT, 2017).

Figure 3. Framework for recorded music in musical XR populated with types (bold) and examples (italic) of XR experiences. 'Environment' refers to the sensory information from the real-world; 'augmentation' refers to the virtual information/content rendered through a display device. For the purpose of this diagram superimposed (AR) and integrated (MR) audio augmentations are not differentiated.



Section 2: Case Studies of Musical XR

This second section investigates current creative practice by analysing selected examples of musical XR. The case studies exemplify the four categories identified in the framework (Figure 3): i) auditory and visual VR, in which visuals and sound from the real world are occluded and replaced by interactive virtual content; ii) auditory VR and visual AR, in which sound from the environment is replaced and there is visual augmentation; iii) auditory AR/MR and visual AR/MR, in which the real auditory environment is augmented by synthesised sound and where synthesised images can be seen/interacted with as part of the immediate physical environment; iv) auditory AR/MR without visual augmentation.[63]

Case studies were selected by searching musical XR recorded music experiences released between 2017 to 2022 using platforms, festivals and reports in academic, trade and online magazines, journals and websites. From these, four examples were chosen to represent different types of musical interaction (participatory/presentational, collective/individual), and musical structure (open/fixed, teleological/non-teleological). The analysis aims to better understand the kinds of musical XR applications which currently exist, the relationship between sound, spatiality and subjectivity manifest in the applications, and their affordances for musical expression and experience. The analysis draws on methods of analysing audio-visual media and music, and associated paratextual material including published and original interviews with creators. The research was approved via the University of Sheffield ethics review procedures and interviews were conducted online with informed consent.

2.1 Case studies of musical VR

2.1.1. Auditory VR with visual VR

In Auditory and Visual VR experiences, sight and sound from the local environment are replaced by computer synthesised visual and auditory information and enable some form of interactivity. Numerous music-centred VR experiences exist, within which category I include those enabling three degrees of freedom (3DoF e.g. spherical/360° video and film) and six degrees of freedom (6DoF). 360° video (also called ‘immersive videos’, ‘spherical videos’ and ‘VR Video’) is a

[63] Other categories shown in Figure 3 lie outside the focus of this paper because they are expressive forms associated with live performance rather than recorded music: visual AR/MR with real-world sound (e.g. a live concert with AR visual staging) and auditory AR/MR with real-world visuals (e.g. an acousmatic concert performance).

form of audio-visual media in which an omnidirectional view is captured, allowing the user to change the angle of view in 3DoF from a stationary point of view (POV). In a 6DoF VR experience the location or performance is filmed from multiple angles and then software is used to recreate the location/performance as a fully volumetric three dimensional image. These are often (although not always) paired with spatialized sound. The majority of consumer experiences are designed to be playable within domestic room-sized environments and therefore combine real movement of the user within this room-scale set up with the use of controllers to 'teleport' from one position to another within the larger-than-roomscale dimensions of the virtual world. With few exceptions (e.g. *Lune Rouge*[64] — a multiplayer, participatory musicking VR experience), these experiences are presentational formats for individual users.

One of the first 360° VR recordings by a major contemporary artist was Björk's 2015 *Stonemilker*,[65] which illustrates some of the specificities and challenges of music performance and arrangement for a stationary ('at seat') 360° context. Made by Björk and director Andrew Thomas Huang for the track of the same name from Björk's eighth studio album *Vulnicura* (2015), the song traces the chronology of a relationship breakdown. It was released in 2015 both as an at-seat 3DoF headset experience which formed part of the *Björk Digital* touring exhibition for *Vulnicura* (2015-2020) and as a 360° video on YouTube (viewed 7.3 million times, as of March 2024) and was subsequently released in 2019 as part of *Vulnicura Virtual Reality Album* on the Steam platform for VR systems.

The title 'Stonemilker' can be understood as a reference to the difficulty in getting someone to give or tell you something due to the character of the person, as in the English proverb 'Like getting blood from a stone', or the biblical description of Moses getting water from a stone. Björ said, 'It's about someone who's trying to get emotions out of another person.'[66] Björk's lead vocal is accompanied by the interweaving sub-melodies of a string ensemble. In the VR app version this instrumental palette is supplemented by the sound of waves crashing, and the sound is

[64] *TOKiMONSTA The Lune Rouge Experience*, dir. by TheWaveVR (2017) <https://vimeo.com/236682878> [accessed 3 July 2018].

[65] *Björk: Stonemilker (360 Degree Virtual Reality)*, dir. by Andrew Huang (2015) <https://www.youtube.com/watch?v=gQEyezu7G20> [accessed 13 July 2018].

[66] Michelle Gesiani, 'Björk Breaks down Vulnicura Single "Stonemilker" on Song Exploder — Listen', *Consequence of Sound*, 17 December 2015 <https://consequenceofsound.net/2015/12/bjork-breaks-down-vulnicura-single-stonemilker-on-song-exploder-listen/> [accessed 10 May 2018].

spatialised so it surrounds the listener in the virtual headphone space and reacts to the listener's position. The overall mood is melancholic, yet with a lightness of touch due to the grace note ornamented string melodies and a playful performance with the camera detailed below. The binary song structure corresponds to two visual scenes, each giving the impression of having been filmed in a single take on the same rocky beach[67]: Grótta Lighthouse on the Seltjarnarnes peninsula at the Northwestern most point of the greater Reykjavik region. In both scenes, Björk performs to camera from four locations at 90° separation and moves between each in a clockwise direction, with clones of her appearing and disappearing at irregular intervals. By circling the camera, Björk is effectively circling the viewer in the 360° surround, who is also encircled by the spatialised audio. This deceptively simple technique ensures visual and auditory dynamism across the 360° space, and was received by fans as offering an intensely personal experience.[68]

The choreography and camera work can be understood in relation to challenges of performance in a 360° setting, including directing the audience's attention, avoiding VR motion sickness and achieving technological transparency (in this case by avoiding obvious editing stitch lines between cameras). Björk walks clockwise between fixed performance locations against a relatively unchanging background which provides motivation to the viewer's (3DoF) gaze around the 360° surround (Figure 4). In a 360° setting the viewer can choose to direct their attention anywhere in the space (often deemed a challenge for normative directorial approaches), and so Björk's movement acts as a focal point for the viewer's gaze. Moreover, the video masks the limitations of 3DoF: the video provides an environment which minimises the desire a viewer might have to explore a virtual scene further because Björk's performance is so compelling and the rocky locations offer no obvious path out of the scene. In addition, Björk's walking pace between performance locations minimises risk of viewer motion sickness. The four performance positions likely correspond to the four cameras capturing 180° angles, with her movement between them minimising distortions typical of the stitch points in 360° footage. The moments when she changes her performance position are mostly synchronised with the sectional structure of the song (in Scene 1) and phrase breaks (in Scene 2) so the change of position can be

[67] The transition between the two scenes is marked by a fade to black, but continuity between the two is created by Björk walking away from the camera to the right along the shoreline with a match on attention in visuals and reactive spatialised audio at the fade up to Björk standing in the new location – a traditional technique in visual continuity editing here used in the 360° medium.

[68] Guy Morrow, *Designing the Music Business: Design Culture, Music Video and Virtual Reality*, Music Business Research (Springer International Publishing, 2020), p. 181, doi:10.1007/978-3-030-48114-8_1.

experienced as motivated by, and expressing, the changing emotional content of each section. At each of those 90° locations, she moves towards and away from the viewer/camera. This is likely constrained by the need to perform with face to front, central to each camera, and it takes advantage of the affordances of the different proximities: figures can quickly become too distant in 360° video, while closer positioning (up to a point) allows greater intimacy and playfulness as seen here.

The sound is spatialised and reacts to the listener's head position. The vocal track was recorded in binaural audio, and the strings recorded with a clip-on microphone on each instrument which were then mixed by Björk and BBC audio engineer Chris Pike to position sound sources in the 360° space. When listened to on mobile devices listeners can experience a real-time interactive mix which uses positional tracking to respond to the 3DoF movements of the listener in relation to the strings and the multiple vocals of the visual Björk clones, thereby heightening the interactivity of the experience.[69]

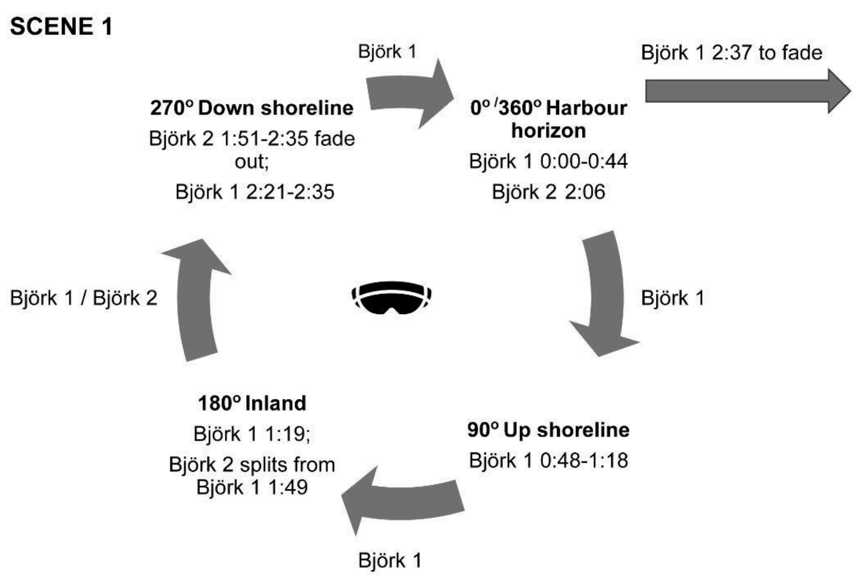
But what does this VR medium mean for the song's realisation? Björk's account of the song emphasises the importance of circular structures, both in the arrangement of the visual and auditory soundscape around the listener, and in terms of song structures: 'If the song has a shape it is sort of like a circle that just goes on forever.'[70] I suggest that the VR realisation exploits the 360° surround visually and sonically in ways which draw out cyclic structures in the cinematography and music, and across which linear processes create directionality.

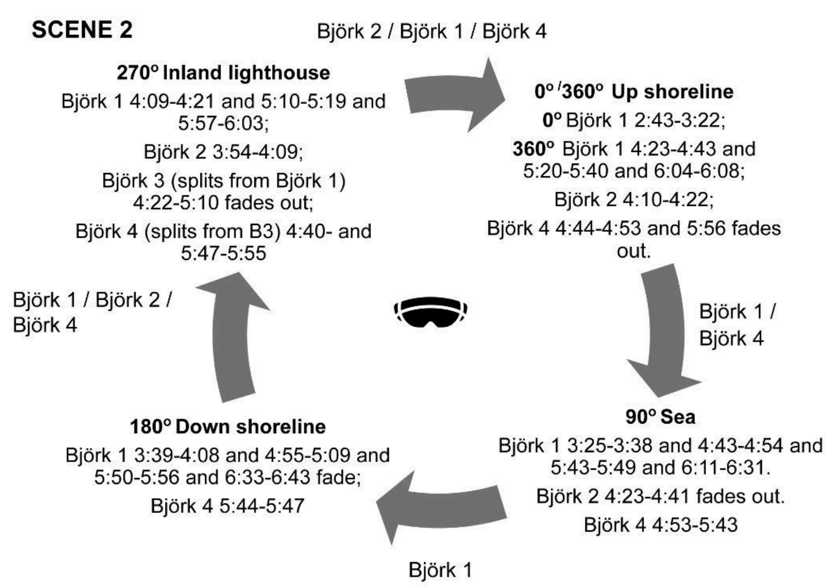
---

[69] Paul James, 'Björk's 360 VR Music Video "Stone Milker" Debuts Tomorrow, Powered by 3DCeption Audio', *Road to VR*, 21 March 2015 <https://www.roadtovr.com/bjorks-360-vr-music-video-stone-milker-debuts-tomorrow-powered-by-3dception-audio/> [accessed 13 July 2018].

[70] 'Bjork's First 360 Degree Virtual Reality Video: Stonemilker', Virtual Reality Reporter, 7 June 2015 <https://virtualrealityreporter.com/bjork-360-vr-video-stonemilker/> [accessed 21 March 2018].

Figure 4. Diagram of performer movement within 'Stonemilker VR' as seen from above with the viewer in the middle of the 360° performance space. Arrows show the direction of travel between fixed performance locations within each of the two scenes. The numbered performers 'Björk 1' to 'Björk 4' indicate which of the four 'clones' is present at any one time. Figure 4a represents Scene 1 (0:00-2:37). Figure 4b represents Scene 2 (2:43-6:44).





The song's form comprises a single repetition of a tripartite structure, book-ended by unaccompanied strings and Björk's halting delivery of the lyric 'Juxtapositioning fate | Finding our mutual coordinates'. A two-bar cello ground bass establishes an E minor to D to A to B minor chord progression whose avoidance of the tonic helps create non-climactic directionality. This is present throughout with the exception of the chorus when the key centre of D major emerges more clearly with alternation between D and E-minor harmonies.

The VR track opens with the sound of waves, immediately providing the soundscape to the virtual beach location. The continual crashing of the waves weaves a cyclic process from the natural world into the song arrangement. The repeated call-and-response (antecedent/consequent) of vocals and strings in the verse creates balance and symmetry and the cyclic pattern referenced by Björk in the quotation above. The waves are heard most clearly in the mix in those places where the voice is absent and it functions as another sound source, sometimes 'answering' the vocal line in places like the strings, and sometimes as an anacrusis to a structurally important vocal line (as heard immediately before 'Moments of clarity', the first line of the first verse at 0:48, and 'Show me emotional respect', the first line of the first chorus). Indeed, this crafted integration of the sound of waves into the arrangement allows them to take on the characteristics of a cyclic bodily process — respiration — and is an example of Björk's wider practice of integrating and embodying natural phenomena in her music.[71]

Linear processes cut across the visual and auditory cyclic structures creating a large-scale arch-shape trajectory. These include increased textural density and complexity, increased pace of events (the time Björk spends in each location in the 360° surround decreases, and the number of clones present at any one time increases), and increased duration of time that Björk looks directly into the camera.

The illusion of a single take in each of the two scenes means that the pace of the video is influenced by the length of time Björk and her clones stay in any one location rather than by the shot length as would be the norm in two-dimensional music video. Increasing visual dynamism is created during the temporal course of the song by the decreasing duration of time spent by each Björk in each location: the mean duration in each location is twice as long in Scene 1 as it is

[71] Nicola Dibben, *Björk*, Revised edition (Equinox, 2009), pp. 53–71.

in Scene 2.[72] The decreased duration spent in each successive location, and the increased number of clones works in tandem with the increasing sonic energy of the track created by increasing textural density and rhythmic complexity.

The final directional process I want to draw attention to is Björk's playful engagement with the viewer's gaze. She mostly gestures, sings and moves around the viewer, only briefly engaging and acknowledging the camera as if in a private reverie, but by the second chorus she presents a more sustained gaze. The length of time she looks directly into the camera gradually increases during the course of the song (duration of the gaze in successive sections is 9 s, 11 s, 21 s, 36 s, and 40 s) until the final line in the second chorus ('show some emotional respect') when all three Björk clones present at that point address the camera (Figure 5). Indeed, the moments where she holds the camera's gaze for longest coincide with the chorus where the lyrics are commands rather than questions. This could be interpreted as signalling a gradual change of psychological state, from personal reverie, to direct address of the implied viewer/listener. The skilled play with the camera is particularly evident in the instrumental outro from 5:43 onwards in which she addresses the camera within each phrase in each location, and looks away while moving between locations, until, with the exception of a brief glance at 6:13, her focus turns inward featuring self-directed stroking and gesturing. So when (at 6:32) she comes close to the camera, gives a quick look up and holds a steady gaze into the camera into the fade, it comes as a suddenly direct and intensely intimate contrast.

[72] Mean performance location duration in Scene 1: Björk 1 = 30s; Björk 2 = 29s; mean performance location duration in Scene 2: Björk 1 = 15s; Björk 2 = 16s; Björk 3 = 24s (or 48); Björk 4 = 16s.

Figure 5. Analysis of gaze duration for each of the four Björk visual clones. Lyrics delivered by Björk directly into the camera are shown in bold for each clone. Unshaded blocks in columns labelled ‘Björk 2’, ‘Björk 3’ and ‘Björk 4’ indicate that the clone is not visible anywhere in the 360º surround.

Song section	Björk 1	Start time gaze into camera	End time gaze into camera	Gaze duration	Björk 2	Start of gaze into camera	End of gaze into camera	Björk 3	Start of gaze into camera	End of gaze into camera	Björk 4	Start of gaze into camera	End of gaze into camera
Intro 00:00													
Verse 00:18													
A	A juxtaposition in **fate**	00:23	00:24	00:01									
	Find our mutual **coordinates**	00:42	00:43	00:01									
B	**Moments of** clarity are so rare	00:49	00:53	00:04									
	I better document this			00:00									
	Across the view is fears			00:00									
	All that **matters** is	01:14	01:17	00:03									
Total gaze duration				*00:09*									
Prechorus 01:48													
C	**Who is** open-chested	01:20	01:22	00:02									
	And who has coagulated			00:00									
	Who can **share and Who has shut down the chances?**	01:40	01:49	00:09									
Total gaze duration				*00:11*									

Song section	Björk 1	Start time gaze into camera	End time gaze into camera	Gaze duration	Björk 2	Start of gaze into camera	End of gaze into camera	Björk 3	Start of gaze into camera	End of gaze into camera	Björk 4	Start of gaze into camera	End of gaze into camera
Chorus 01:50													
D	**Show** me emotional respect	01:50	01:52	00:02	**Show me emotional respect**	01:50	02:05						
	oh respect, oh respect	01:59	02:02	00:03	oh **respect, oh resp**ect								
	And I have emotional needs, oh needs, oh ooh			00:00	And I have emotional **needs,** oh needs, oh ooh	02:13	02:16						
	I wish to **synchron**ize our feelings, our feelings, ah ha **oh ooh** [brief looks into and away from camera.]	02:22	02:38	00:16	**I wish to** synchronize our feelings, **our feelings, ah** [fades out.]	02:21	02:34						
Total gaze duration				*00:21*									
[Fade out from scene 1 and into scene 2.]													
Verse 02:45													
B	What is it that I have			00:00									
	That makes me **feel your pain Like milking a stone To get you to say it** and	03:01	03:22	00:21									
Prechorus 03:23													
C	Who is open?			00:00									
	And who has shut up	03:29	03:37	00:08									
	And if **one feels** closed	03:39	03:44	00:05									
	How **does** one stay open?	03:47	03:49	00:02									
Total gaze duration				*00:36*									

Song section	Björk 1	Start time gaze into camera	End time gaze into camera	Gaze duration	Björk 2	Start of gaze into camera	End of gaze into camera	Björk 3	Start of gaze into camera	End of gaze into camera	Björk 4	Start of gaze into camera	End of gaze into camera
Chorus 03:55													
D	We have emotional needs, **oh needs, oh needs, oh** ooh	04:02	04:08	00:06	We have emotional needs, oh needs, oh needs, oh ooh								
	I only wish to synchronize **our feelings**, our feelings, **ooh** [brief looks into and away from camera.]	04:17	04:22	00:05	I only wish to synchronize our **feelings, our feelings**, ooh	04:17	04:22						
	Oh show some emotional respect, oh respect, oh ooh	04:25	04:42	00:17	**Oh show some emotional respect, oh respect, oh ooh**	04:24	04:41	Oh **show some** emotional respect, **oh respect, oh ooh**	04:26	05:10			
	[vocables]	04:57	05:09	00:12							Vocables	04:51	04:52
Total gaze duration				*00:40*									
Outro 05:10													
A	**A juxtaposition in fate**	05:12	05:19	00:07							A juxtaposition in **fate**	05:17	
	Find our mutual **coordinates**	05:32	05:40	00:08							**Find our mutual coordinates**		05:40
	[Movement clockwise between 90 degree camera positions: Björk addresses the camera within each phrase and looks away when moving between locations.]	05:41	06:44	01:03							[Movement clockwise between 90 degree camera positions: Björk addresses the camera within each phrase and looks away when moving between locations.]	05:41	06:00
[Fade out.]													

In addition to its expressive and structuring effects, this changing address of the camera is particularly significant for a reason specific to VR: it acknowledges the viewer within the virtual environment. By donning a 360° headset and headphones or viewing through the 'magic window' of a mobile device the viewer occupies a first-person perspective: they are at the centre of the visual and sonic surround, and at moments is addressed directly by Björk's gaze, gestures, and even potentially by the 'you' of the lyrics. Yet the viewer has no body representation within the virtual environment and they cannot act on that world other than by turning their head. So, while *Stonemilker* VR satisfies many sensorimotor contingencies (a 3D sound environment which reacts to head-position, and a stereoscopic display which provides depth perception) and affords a good sense of place and plausibility, it does not offer embodiment within the virtual environment: the player is both present (I can see and hear I am on a beach with Björk) and not present (I have no body within the virtual environment). I have what virtual environment researcher Mel Slater calls 'ghost presence'.[73] Björk's playful acknowledgements of the viewer may help overcome some of the potential experience of 'invisibility' and help create the spatial and emotional intimacy between performer and viewer congruent with the song and album's themes but I, the viewer, am still strangely absent. While on the one hand this can be viewed as a technological limitation of VR, it also highlights a facet of recorded musical experience that has always been there but which is generally overlooked – the voyeuristic/ eavesdropping character of audiences of recorded music where one listens in without being seen or heard oneself.[74]

Stonemilker VR is representative of a particular kind of seated, 3DoF experience with head tracked audio and while it was the first of its kind in terms of a commercial music release by a pop artist, the use of the technology remains much more extensive in gaming than in audio-first experiences. Beyond its limited use as a medium for pop releases, it has potential for music in heritage contexts. For example, the *Linlithgow Palace* VR experience reconstructs a ruined chapel that had once been the site of important musical activity.[75] The VR experience enabled users to see the lost space and hear it as it would have sounded by virtue of modelling the visual

[73] Mel Slater, *Mel Slater's Presence Blog: Ghost Presence*, 26 May 2017 <http://presence-thoughts.blogspot.com/2017/05/ghost-presence.html> [accessed 12 July 2018].
[74] Nicola Dibben, 'The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings.', in *Electrified Voices: Medial, Socio-Historical and Cultural Aspects of Voice Transfer* (V & R Uni Press, 2012), pp. 107–22.
[75] James Cook, 'Hearing History: Bringing to Life the Sounds of the Past through Virtual Reality', UK Research and Innovation, 2019 <https://gtr.ukri.org/projects?ref=AH%2FS010653%2F1> [accessed 24 October 2022].

(and auditory) interior in 3D and the music that would have been performed within it. Choral music was recorded in an anechoic chamber and then overlaid with the acoustic modelling of the chapel, thereby recreating the soundscape of the chapel. The resulting performances were released as a linear audio recording, and were intended to form a stationary, onsite VR visitor experience. Inclusion of volumetric capture of the performer in VR music experiences offers greater interactivity by enabling the user to move around the performance space. *Hallelujah*[76] is the first such live action experience: in this performance of Leonard Cohen's song, all five spatialised voices in the a cappella arrangement are sung by (visual and auditory) clones of singer Bobby Halvorson, who encircle the central point of view, albeit within an otherwise conventional linear presentation of the song.

None of these early examples of musical XR exploit the full interactive potential of VR. One which begins to experiment with a 6DoF realisation is *Melody of Dust*[77] — a single-user VR music experience released in 2017 on the Steam VR platform by Viacom NEXT and American musician Nick Koenig (aka Hot Sugar). The creators describe it as an experiment in creating an interactive music album for XR: the user discovers 'melodies' within an VR audio-visual environment – music stems – by throwing different combinations of virtual objects into a central vortex which then triggers various tracks which play in linear format.

Play takes place inside a cavernous marble hall, the giant marble hand sculptures ranged around the circular room giving the space a temple-like appearance. A crystal ball sits on a stone pedestal which, when thrown into a central pool, opens double doors behind the visitor revealing a single room into which the player can teleport. Objects in the room (roses, vases, a cushion, a spider) make sounds when picked up and thrown into a central tornado above the pool where they circle. Sounds are spatialised congruent with the virtual space: the music created in the tornado and heard from the object room is quieter and has a low pass band filter applied giving the impression of distance; the crystal ball sound is spatialised to match its position near the visitor. According to Koenig, the visitor

> can hear the origins of the song in the objects that they picked up […] When you shake the dove, you hear the sound of a dove. Then you throw it into the tornado,

---

[76] *Hallelujah (VR)*, Music, Short, dir. by M. Zach Richter (Within, 2017).
[77] *The Melody of Dust*.

> that melody has the voice of that dove hitting different notes. That was the original idea, to romanticize what I'm trying to do with music by turning it into a fantasy game.[78]

When the right combination of elements is present to hear a specific track, the crystal ball reappears. Throwing the ball into the pool ends the track-building phase (the object room disappears and the visitor POV is restricted to 3DoF instead of previous 6DoF) and the experience segues into a full rendition of a track and 360° linear video. At the end of the climactic, linear video the player is told how many of the more than 80 possible melodies have been ‘discovered’.

The music of *Melody of Dust* VR experience and associated thirteen-track audio album have a restricted musical palette which may result from the need to create musical stems which could easily recombine into different ‘stand alone’ music tracks. Tracks share a common musical identity: copious amounts of reverberation, a tempo of 108 bpm, a pitch centre on D, and use a synth-based palette which is especially noticeable in the timbre and processing of instrumental and vocal sources. For example, the track 'I First Heard the Melody from Within the Womb' is dominated by swirling synthesised scalar passages over a minimal drum track emphasising the back beat, and lugubrious one bar (in 4/4) harmonic changes and no obvious found sounds. A trip-hop-like drum track, filtering and pitch bending, is combined with almost classical-sounding short melodic phrases.

Despite its limitations — the restricted musical palette, the interactivity limited to triggering stems and tracks, and sound spatialisation limited to proximity effects — it illustrates the potential for participatory VR music experiences. The process of the visitor constructing the songs, rather than simply pressing ‘play’ to hear each linear track, starts to get away from the idea of a single complete musical text or indeed that the ‘song’ is a linear performance, although both these elements are still present as this brief analysis illustrates. Moreover, it illustrates how a virtual environment can be specific to a track or album, to some extent akin to the function of a music video: as Koenig puts it, ‘You’re in a venue created for a song rather than being at a venue that’s having the song played within it.’[79]

---

[78] Sandra Song, ‘Hot Sugar’s Romantic Dream World Is Now Available Via Virtual Reality’, PAPER, 8 March 2017 <http://www.papermag.com/hot-sugar-vr-2305520028.html> [accessed 31 August 2018].
[79] ‘Hot Sugar’s “The Melody Of Dust” Is The Future Of Music’, NYLON, 23 March 2017 <https://nylon.com/articles/hot-sugar-the-melody-of-dust-virtual-reality-album> [accessed 4 July 2018].

These examples illustrate the variety of creative possibilities that artists, directors, engineers and developers have investigated. They share commonalities as regards the kind of spatiality and subjectivity they manifest: for example, the medium is sometimes used in ways which heighten the sense of a personalised performer-listener relationship. They allow the user more agency over what they hear and see at any given moment and, as one journalist remarks, provide a focus of attention for music listening which is particularly notable in the face of mobile music's ubiquity and simultaneity with other activities.[80] As this illustrates, VR provides an opportunity to create experiences that change according to audience choice and exploration within a virtual environment. Whether this is an experience that composers and audiences want is a question I return to later (see §3.2).

2.1.2. Auditory VR with visual AR

At the time of writing, the majority of consumer musical XR experiences augment the surroundings with visual overlays and replace the sound of the immediate environment with recorded sound. An example of this is *Moonmoons AR* which is an AR rendition of a four minute single of the same name by composer Anna Meredith from a studio album for band and electronics. The *Moonmoons AR* app for smartphone overlays visual AR as viewed through a camera lens and presents spatialised audio for headphones.[81] This is a presentational, linear performance which is interactive in so far as the listener can move around virtual sound sources in their environment, exploiting proximity effects to change the mix of the track.

British composer Anna Meredith's music is known for its distinctive blend of contemporary classical, electronic, and experimental elements. 'Moonmoons' features many of her music's characteristic features: intricate rhythmic patterns, pulsating textures, strong directionality via a gradual increase of sonic momentum and energy, and an interplay between acoustic and electronic instrumentation, which mark the influence of minimalist aesthetics and popular culture. The track has an arc-like structure, which starts and ends with an ethereal texture comprising a repeated synthesised/processed plucked rising contour arpeggiated gesture

[80] Marc Hogan, 'Is Secretive Virtual Reality Startup Magic Leap Dreaming Up the Future of Music?', Pitchfork, 2017 <https://pitchfork.com/features/article/is-secretive-virtual-reality-startup-magic-leap-dreaming-up-the-future-of-music/> [accessed 9 July 2018].

[81] Arthur Carabott, 'Moonmoons AR', Arthur Carabott, 2019 <https://www.arthurcarabott.com/moonmoons-ar/> [accessed 24 October 2022].

alternating between two chords in antecedent-consequent structure over which a cello enters. The cello melody initially avoids downbeats, giving it a floating quality, but transforms into a propulsive rhythmic build which emerges and dominates a contrasting middle section in minor mode. *Moonmoons AR* makes use of Meredith's characteristic 'build' technique: between 1:45 and 3:02 there is a thinner texture without percussion and bass, a gradual build-up of sound energy due to the addition of layers, an increased rate of activity, and a rise in overall pitch level and rising pitch contours, peaking with what Meredith refers to as the 'confetti' gesture — a stabbing cascade made from a compressed version of earlier material, and the reason for the track's title ('moonmoons' are the natural satellites that orbit a planet's moons, and reference the idea of the track also containing a compressed version of some of its own musical material).[82]

The track 'Moonmoons' was partially written when British developer Arthur Carabott, a personal and work contact of composer Anna Meredith, approached her with the idea for an AR app. 'Moonmoons' was chosen for the AR realisation because they needed a track where they could group instrumental sources into a small number of stems that could be attached to separate AR visual objects with some temporal consistency.[83] This stem-based approach was a logical continuation of Carabott's previous work which had explored spatialisation using multiple speakers; instead of speakers he took advantage of the convergence of AR technologies (SLAM and Spatial Audio), mobile phone power which enabled real-time running, the Unity 3D editing environment and Google's Resonance Audio Unity plugin. Six types of musical material (stems from the track) correspond to six volumetric captured sculptures made by Eleanor Meredith (Anna's sister) which the user places in their environment such that they function as virtual loudspeakers: 'Bass', 'Soft synth', 'Perc' (thundersheet), 'Cello' (bowed with vibrato), 'Arpeggios' (a plucked-like electronic synth sound) and 'Textures' (owl-like hooting and 'confetti' gesture). The perceptual effect of the sound appearing to emanate from the virtual sculptures arises from 3D audio-visual localisation cues, and congruence between visual and auditory event onsets: each sculpture is lit and pulsates when its associated stem plays, and the three structural divisions of the music arising from changes of texture and rate of movement coincide with change in the

[82] Anna Meredith, 'Interview', interview with Nicola Dibben, 23 June 2021, Online.
[83] Arthur Carabott, 'Interview', interview with Nicola Dibben, 1 March 2022, Online.

visuals from colour to black and white (1:43) and back again (2:54).[84] This ternary structure of the track was another reason for its choice. Carabott noted that ‘A lot of 3D music stays in one world’ and he was interested in introducing change: ‘Even though it was abstract instrumental music and graphics it would have a narrative arc rather than just hit play and walk around, and hit play and stop when bored.’[85] The combined visual AR and spatialized audio enable the listener to remix the music, altering the relative balance of stems, by moving through the physical space.

Carabott noted the interesting perceptual issues that emerged in the making of the AR experience whereby 'realistic' acoustic treatment was insufficient to create the best AV experience. These experiences illustrate the point that sound reproduction technologies are about producing perceptual effects rather than producing accurate models of 3D reality: ‘Realism is nothing more and nothing less than an aesthetic effect.’[86] For example, the relationship between the volume of sound sources in the app and their proximity was managed so as to provide maximal loudness at a distance at which the whole of the virtual sculpture could be seen through the phone screen rather than when the phone/user was right next to or inside the virtual object.[87]

The interactivity afforded by the AR app raises some of the same issues as noted with Stonemilker VR, namely the tensions inherent in the listeners’ ability to move around a 360° space (in 3DoF in this case) and thereby miss significant events in the audio-visual material. For example, in *Moonmoons AR* the sound level of stems in the mix is attenuated and boosted according to proximity to each virtual sound sculpture, but this means it is possible to miss the structurally climactic moment in this track, i.e. the ‘drop’ of the ‘confetti gesture’: if one moves further away from the ‘Textures’ sculpture and closer to other sculptures then the sound of the confetti gesture is masked. One can view this purely as a technical issue which requires resolution, such as by altering sound attenuation so that regardless of where one stands the listener never misses that structural moment, or by using visual cues to direct the viewer’s attention to relevant virtual sound sources. However, it is also an artistic, aesthetic and

[84] Despite their conceptual relationship to the album, there is incongruity between the virtual sound sources and the sounds because the objects do not look as though they could physically produce the sound allocated to them. Other apps deal with this potential perceptual incongruence by using abstract solids or orbs without definite form, e.g. *Fields AR*.
[85] Arthur Carabott, ‘Interview’, 1 March 2022.
[86] Sterne, ‘Space within Space’, p. 126.
[87] Arthur Carabott, ‘Interview’, 1 March 2022.

compositional problem, which belies the fact that *Moonmoons AR* was made for a pre-composed track: one could instead compose music without climatic peaks assigned to individual sources, or in such a way that there is a deliberate compositional strategy to the assignment of instrumental materials to particular locations so that the aesthetic result exploits this affordance of the technology. Discussing this compositional problem, Carabott remarked on the experimental character of what they were doing: 'With those things you don't know what's good yet. So this was a good place to start. We need some explorers to go into the metaphorical jungle and find where the safe places to go are to figure out what works well in the medium'.[88]

Similar principles to those in *Moonmoons AR* operate in the case of *Fields*,[89] but this app functions as a platform for numerous individual AR experiences and artists and affords participatory music-making. Similar to *Moonmoons AR*, *Fields* positions stems in the user's surroundings, representing the virtual sound sources as coloured, transparent orbs. In addition, it allows the user to record their own sounds and make original 3D spatialised compositions. While the orbs are a form of visual AR the music is once again spatialised auditory VR which is translocal in character. Geolocated spatialised three-dimensional sound experiences enable sound installations to be fixed in specific locations, opening up new artistic possibilities for music's relationship to place. For example, in heritage sites visual AR and auditory VR could be used to enable visitors to experience auditory reconstructions of lost architecture or landscape while walking through and viewing the real site (as opposed to seeing a visual simulation of the site in VR realisations such as the *Linlithgow Palace* project[90]). Another possible application of auditory VR combined with visual AR is to 'musicalise' one's physical surroundings by attaching sounds to specific objects within a space – either human-made places and objects, or those of the natural world. Or one could create a listening party, with tracks visually dispersed throughout a venue or in different rooms in a house. Both these examples require a medium-specific compositional approach which treats sounds as spatial and temporal from the start rather than breaking up a pre-existing track into stems.

The examples discussed above manifest a particular kind of spatiality and subjectivity. The spatialisation is three dimensional and of a scale that maps onto the listener's domestic space.

[88] Arthur Carabott, 'Interview', 1 March 2022.
[89] Particle Incorporated, *Fields - Spatial Sound in AR* (2018), iOS <https://fields.planeta.cc/>.
[90] Cook, 'Hearing History: Bringing to Life the Sounds of the Past through Virtual Reality'.

The listener is positioned at the centre and arranges sound around themselves. It affords exploration of sounds in relation to self, but not in relation to the environment they are in because the sounds are deaf to the real acoustic environment. These experiences construct the listener as their own ‘audio mix engineer’, rearranging pre-existing musical components into spatialized relationships within which the listener is the centre of a domestic sized, individualised experience. The musical material does not interact with real-world sound in the environment, and can’t be dynamically controlled in three dimensions. For that, other kinds of experiences exist as described below.

2.2 Case studies of musical AR/MR

2.2.1 Auditory AR/MR with visual AR/MR

In auditory AR/MR, the real auditory environment is augmented by synthesised sound, sometimes with synthesised visual representations of these sounds which function as interactive visual sources in a space. Examples of the first commercially available experiments with this idea are *Bloom: Open Space,* a room-sized mixed-reality experience by Brian Eno and Peter Chilvers for Hololens and *Tónandi* for Magic Leap headset with Icelandic post-rock band Sigur Rós.

Bloom[91] was one of the first smartphone apps available when the iPhone came out in 2008: it is a procedural music app which emits a bell-like sound and displays an expanding coloured circle when the screen is tapped. The sequence of taps is repeated, builds up and fades away creating ambient music. Ten years later this concept was realised as *Bloom, Open Space*[92] — a ten minute HoloLens gallery installation for up to ten people at a time. The installation used HoloLens MR headsets (which have speakers positioned above the ears) worn by users standing within a circle of 6 screens and 6 speakers. Users pinched in the air with their thumb and forefinger to create a pitched sound and a three-dimensional bubble which repeated and faded.[93] Like the touchscreen version, pitch height was mapped to spatial elevation such that placing a bloom higher up produced a higher pitch, but in this case sound spatialisation was incorporated such that the mix differed with proximity to the sound bubbles. The MR version was a multiuser, networked

[91] Brian Eno and Pete Chilvers, *Bloom* (Opal, 2008), iOS.
[92] Brian Eno and Peter Chilvers, *Bloom: Open Space*, February 2018, Mixed Reality for HoloLens.
[93] *Behind the Scenes of Bloom: Open Space*, dir. by Microsoft (2018) <https://www.youtube.com/watch?v=-vQ_DYWh734&t=5s> [accessed 1 April 2021].

experience meaning that people could in theory hear and see each others' Blooms, and interact in creating a whole musical experience.[94] The visual domain of coloured 'blooms' is an example of MR because they are created using hand gestures and they appear to move within the three-dimensional space of the installation. The sound of the Blooms are spatialised so they appear to emanate from the visual blooms within the three-dimensional space, while also being heard alongside other real gallery sounds. It can therefore be thought of as an instance of auditory and visual MR since the open backed headphones of the Hololens mean that users can see and hear their own and others' creations.

Analysing the relationship between the musical characteristics of *Bloom* and its media forms as a touchscreen app and MR experience reveal some of the creative potential of auditory and visual AR/MR and its implications for user and compositional experience. *Bloom* is an example of ambient music; the music is dynamic, in the sense of continually changing, and designed to be part of and to function as an environment. To that end many of its features afford ignorability, and openness to semi-attentive listening. The work is open-ended and not fixed: it sounds as though it could continue indefinitely, there are no textural surprises, and while the randomisation in the procedures means events might only happen once, the lack of change can be experienced as calming because there is no sense the listener will miss anything.[95] This is afforded by the restricted timbral and pitch palette heard against a drone but without a strong background/foreground effect. This restricted palette creates a particular kind of continuous variation through low dynamics, soft pitched percussive attack and soft long decay, creating a wash of sound with blurred edges, and although the rate of repetition and exact pitches can be adjusted these are from a restricted pitch collection. The same diatonic pitch structure is mapped to the vertical axis (of the screen in the Bloom apps, and of the space in the MR installation) comprising the root, fifth degree, the fifth above that, and the full scale over five octaves. Separation of the root notes ensures clarity and consonance. These appear in permutations of four modes (dorian, lydian, mixolydian, lydian flat 7), starting on five different pitches (B flat, F, C, G, D, A): mixolydian (C, G, D), lydian (B flat, F, C), dorian (G, D, A), lydian flat 7 (F, C, G). Notably, the harmonic relationship between starting pitches means that harsh changes between

[94] 'Reinventing Brian Eno and Peter Chilvers' "Bloom" for Mixed Reality', *Microsoft In Culture*, n.d. <https://www.microsoft.com/inculture/musicxtech/bloom-open-space/> [accessed 19 March 2021].
[95] Paul Roquet, 'Ambient Landscapes from Brian Eno to Tetsu Inoue', *Journal of Popular Music Studies*, 21.4 (2009), pp. 364–83 (p. 379), doi:10.1111/j.1533-1598.2009.01208.x.

them are avoided. As Eno describes it, his intention is to create variation 'akin to sitting by a river. So it's continuous, albeit within quite a restricted range […] That means the river isn't going to suddenly turn blue, or burst its banks, or turn to concrete'.[96]

Bloom: Open Space extends Eno and Chilver's underlying idea of ambient music as an environment rather than an object. Moreover, it realises a number of Eno's previous techniques: AI and generative music, generative art, light sculptures and music created for spaces that are public and meant to be moved through, such as airports and installations. *Bloom: Open Space* enables the listener to move around inside the music as if the piece of music itself is the environment. Consequently, the music is partly a composition and art work, but also partly an instrument to be played by the listener. For Eno this blurs the distinction between audience and artist and changes the role of the composer: 'What this means, really, is a rethinking of one's own position as a creator. You stop thinking of yourself as me, the controller, you the audience, and you start thinking of all of us as the audience, all of us enjoying the garden together. Gardener included.'[97]

But what is also revealed in this analysis of *Bloom: Open Space* is a primarily ocularcentric approach to MR. There are currently few examples of this medium due to the early stage of MR technology production and commercialisation.

2.2.2. Auditory AR/MR without visual augmentation

Auditory XR which augments real-world sound in ways which are blended into the auditory environment ('auditory AR/MR') are rare. The most well known example of this was *RjDj* ('Reality Jockey DJ' 2008-2013), a free app for iPhone described by its founder Michael Breidenbruecker as 'a new format that could react to context'.[98] The app produced reactive music in real-time using the sensors of the mobile phone and dependent on the movements of the user: ambient/environmental sound was modified using Pure Data digital signal processing so that the listener heard environmental sound plus mutations, additions or subtractions, together with precomposed audio. It can be thought of as an instance of Auditory AR because the listener's

---

[96] Tom Fenwick, 'Brian Eno', Flaunt Magazine, 2017 <https://flaunt.com/content/people/brian-eno> [accessed 30 April 2021].
[97] Kingsley Marshall and Rupert Loydell, 'Thinking inside the Box: Brian Eno, Music, Movement and Light', *Journal of Visual Art Practice*, 16.2 (2017), pp. 104–18, doi:10.1080/14702029.2016.1195073.
[98] Charlie Burton, 'Mod Your Sounds with RjDj', Tags, *Wired UK*, 16 December 2009 <https://www.wired.co.uk/article/mod-your-sounds-with-rjdj> [accessed 28 June 2022].

auditory environment is integrated with the music to create a hybrid experience. Subsequent apps from the same parent company used the same basic design (e.g. *Inception the App* which was linked to the movie of the same name).

Although reactive music existed prior to and since *RjDj,* it was the first mass adopted consumer reactive music product and exemplifies key characteristics of this medium. The user interface in this case is the user's movement and interaction with their environment rather than with a visual interface.[99] As a consequence, it produces a form of sound art in real-time which has its own forms rather than being tied to previous recorded music formats; it is 'somewhere between an album and a game'.[100] It also destabilises conventional ways of thinking about the commodity form of the musical work: what is distributed in reactive music is not the music but the software to generate the music, which some see as opening up business models formerly associated with gaming.[101]

The technologies used to create these AR/MR experiences — 'hearables' and the reactive systems they are paired with — have been used to enable environmental sound to be filtered according to preset selections appropriate to different environmental contexts (e.g. restaurant, office), and for music listeners to mix their own versions while listening live. For example, Active Listening system by Doppler Labs (2016) comprises ear buds and an app which captures and processes audio from the environment as well as playing back music. The emphasis in these types of experiences is on a close mapping between the real and virtual to achieve an 'optimisation' of the sound environment for individualisation and personalisation. But there are affordances for other kinds of experiences, namely those which exploit the compositional potential in combining hear-through capability with location and context awareness to create music experiences which are adaptive/reactive to their specific context.

2.3 Case study conclusions

The examples analysed above exemplify a proposed typology of musical XR; examples in which the sounds of the environment are *replaced* by music (auditory VR with visual VR; auditory VR

[99] David Barnard and others, 'Epilogue: Reactive Music and Invisible Interfaces', in *iPhone User Interface Design Projects* (Apress, 2009) <https://doi.org/10.1007/978-1-4302-2360-3_11>.
[100] David Collier, 'In Your Own Time: A Mobile Music Composition' (unpublished, Trnity College Dublin, 2012), p. 50 <http://davidcollier.ie/wp-content/uploads/2012/08/4_InYourOwnTime.pdf>.
[101] Florian Waldner and others, 'Cross-Industry Innovation: The Transfer of a Service-Based Business Model from the Video Game Industry to the Music Industry', *2011 International Conference on Emerging Intelligent Data and Web Technologies*, September 2011, pp. 143–47, doi:10.1109/EIDWT.2011.30.

with visual AR) and those in which the sounds of the environment are *augmente*d by music (auditory AR/MR with visual AR/MR; auditory AR/MR without visual augmentation). The analysis highlights additional dimensions along which these experiences operate: interactivity (the extent to which the experience is participatory or presentational, and to which it is collective or individual), and linearity (the extent to which the structural form is open or fixed, and teleological or non-teleological). The search for case studies revealed that certain kinds of experiences (e.g. visual AR with VR sound) are more common than others; mixed-reality experiences are rare at present but offer a promising domain for future interactive musical experiences.

XR experiences can also be characterised in terms of their affordance of 'being' versus 'doing'.[102] Most VR experiences emulate the gaming industries' values in so far as they focus on a 'doing' mode in which the user takes intentional action to affect the virtual world. This is true of musical XR experiences such as *Melody of Dust* in which the user's actions are necessary to trigger sound stems and linear renditions of tracks. This contrasts with 'being' in XR experiences — states of reflection, stillness, and immersion very similar to what in music might be called contemplative listening. *Melody of Dust* includes both 'doing' (actions to trigger sound sources and tracks) and 'being' (listening to non-interactive playback of a recorded track) in different parts of the experience, whereas *Moonmoons* and *Stonemilker* are more contemplative in character due to their limited affordance of interactivity, and *Bloom* achieves a more equal balance between the two.

Two models of listening can be inferred from the design, function, and mode of address of the musical XR analysed above. The first is so normalised that it is easily overlooked, namely AR audio in which music is broadcast from speakers, or superimposed via hear-through headphones, onto the real acoustic environment. Musical AR affords a potentially shared, simultaneous experience, albeit one which is non-interactive and presents minimal opportunity to experience spatiality *within* the music presented. The second model of music listening is manifest in auditory VR. Building on Gascia Ouzounian's analysis of the cultural, social and political production of space,[103] I suggest that current auditory VR privileges a particular model of spatial

[102] Jack Atherton and Ge Wang, 'Doing vs. Being: A Philosophy of Design for Artful VR', *Journal of New Music Research*, 49.1 (2020), pp. 35–59, doi:10.1080/09298215.2019.1705862.
[103] Ouzounian, *Stereophonica*.

hearing, namely a navigable, three dimensional, Euclidean geometric space, frequently domestic-scale, with the listener at its centre. The occluded sound world of auditory VR is akin to (stereo) headphone listening in so far as the majority of these VR experiences offer contemplative solitary listening in an alternate space. Multiplayer VR experiences such as *Lune Rouge*, or AR/MR experiences involving acoustic transparency, can by contrast afford a shared auditory experience and social connection, yet instances of these are few.

Current musical XR experiences are notable for their congruence with a trajectory of sound in music recordings which helps produce a particular listening subjectivity — the centred, singular perspective which others have noted is characteristic of stereo production and reception.[104] One difference of listening experiences in XR as opposed to other kinds of sound reproduction is that the headtracked experience places sounds outside the head. Moreover, unlike other sound recording practice, this currently early stage of XR affords auditory experiences which are congruent with the real/virtual space they are experienced within, rather than seeking to create spaces which could never exist in the real world. Future creative explorations could become more experimental in both regards. There is also the question of the kind of subjectivities being constructed in XR; as media theorist Frances Dyson asks in her critique of VR — 'what kind of eyes and ears are being constructed here?'.[105] Whose realities are being extended in XR? And whose spaces are simulated or created? William Fourie has highlighted the invisible labour performed by the acoustic presets of DAWs which are predominantly modelled on Western spaces.[106] The spaces and subjectivities of XR are just as implicated.

This analysis of case studies illustrates a new typology of musical XR which itself highlights factors differentiating these media and specific new compositional opportunities. The next and final section considers the extent to which musical XR experiences might cause us to rethink recorded music and what the future of music in XR might be.

---

[104] Dibben, 'The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings.'; *Living Stereo: Histories and Cultures of Multichannel Sound*, ed. by Paul Théberge, Kyle Devine, and Tom Everrett (Bloomsbury Academic, 2015), pp. 15–16.
[105] Dyson, *Sounding New Media*, p. 130.
[106] Personal communication, 2021.

Section 3: Rethinking Recorded Music Through the Lens of XR Experiences

3.1 Recontextualising pre-XR recorded sound and music listening experiences.

The analysis of XR in recorded music presented above identifies many continuities with prior compositional and listening practices. Taking a very loose/inclusive view of XR (as some people do) we could argue that recorded music (in fact acousmatic music) was the first form of virtual- and augmented-reality sound. In the realm of VR, computationally generated sensory information replaces the sensory percepts from the user's physical environment. I argued above that this is akin to listening practices associated with recorded music when, for example, sound is played so loudly over speakers that it blocks out real-world sound, or when performance conventions and listening practices subordinate real-world sound to musical sound. Similarly, standard headphones replace real environmental sound with recorded sound, just as a VR headset blocks out the real world and replaces it with a virtual environment. However, the acoustic overlay of a standard mono or stereo reproduction does not afford the sensorimotor-contingencies of VR because it lacks the interactivity of 3DoF or 6DoF virtuality and, when wearing headphones, the location of the sound moves with the user's head rather than being fixed in space. The full auditory VR experience (equivalent to the 360º 6DoF of visual VR) is realised only with spatial audio using binaural recording and positional playback.

This analysis of recorded music in XR also highlights some oddities of pre-XR experiences of recorded music which have become so normalised we hardly notice them anymore. Writing at the turn of the twentieth to twenty first centuries, Steven Feld remarks on the extent to which we have become habituated to this form of sonic virtuality: 'Not only does contemporary technology make all musical worlds actually or potentially transportable and hearable in all others, but this transportability is something fewer and fewer people take in any way to be remarkable.'[107] Moreover, what is considered a natural or surreal recorded auditory environment has changed over time through processes of naturalisation.[108]

The head-locked sound reproduction characteristic of mono and stereo headphone listening is a very particular experience encountered nowhere else in daily life, and yet it seems to cause listeners little cognitive dissonance. One could speculate that the path to headlocked audio was

---

[107] Steven Feld, 'A Sweet Lullaby for World Music', *Public Culture*, 2000, p. 27 (p. 14).
[108] Brøvig-Hanssen and Danielsen, *Digital Signatures*.

eased by the congruence between practices of contemplative listening to recorded music with the relatively stationary and limited interactivity of listening to Western concert music live in seated concert venues. Indeed, Gascia Ouzounian has argued that the foundations of today's spatially-oriented listening cultures derive from nineteenth century scientific, technological and aesthetic concerns.[109] Perhaps what is remarkable is that the headlocked form of listening is now so normalised that when we put recorded sounds back into specific places in the world via spatial audio it seems like a conceptual (if not perceptual) marvel.

In addition, Auditory AR/MR seems to run counter to the schizophonic trajectory of recorded sound. As many have noted, the development of sound recording enabled sounds to be split from their original space and place of performance, and subsequently, through the invention of the magnetic tape recorder, it extended the spatial and temporal parameters of recordings to allow the creation of patchworks of different times and places.[110] By contrast, location-based auditory AR/MR can afford the integration of music with the real auditory environment in which it is experienced. Arguably, we have done the difficult, imaginative work required to experience stereo sound as a virtual environment and events; what head-tracked spatial audio does is provide an experience of recorded music which is more akin to perceptual experience of the real acoustic environment.

3.2 Auditory XR futures

Considering that XR is now part of the contemporary music industry's cultural production alongside music videos, album art, stage design, and merchandise, what is the potential future of musical XR, particularly in the forms of VR and AR/MR? Turchet et al's review of the musical metaverse identifies various related opportunities and challenges across technological, regulatory and artistic domains, the latter being most relevant here.[111] As they note, a key challenge is XR adoption, which hinges on the social acceptance of hardware and its utility for creators. The reception and uptake of audio-first XR experiences have been limited, even within the smartphone market where XR is relatively accessible and affordable. Composer Anna Meredith described the underwhelming experience of releasing her *Moonmoons* app which received

---

[109] Ouzounian, *Stereophonica*, p. 36.
[110] Ragnhild Brøvig-Hanssen, 'The Magnetic Tape Recorder: Recording Aesthetics in the New Era of Schizophonia', in *Material Culture and Electronic Sound*, ed. by Frode Weium and Tim Boon (Smithsonian Institution Scholarly Press/Rowman and Littlefield Publishers, 2013), pp. 131–57.
[111] Turchet, 'Musical Metaverse'.

minimal press attention and relatively few downloads despite being free.[112] This could be due to several factors specific to that release: Meredith's profile as an emerging cross-over artist; consumer reluctance to engage with via dedicated, stand-alone apps; the lack of a supporting industry partner or hardware launch with their promotional capabilities; and/or the app's limited interactivity (the AR visual overlays have restricted real-world integration, hindering their ability to convincingly represent sound sources). However, the limited adoption of musical XR may also stem from its perceived lack of additionality. The spatial dimension of sound is already effectively used in electroacoustic and concert performances, studio recordings, cinema sound, and computer games. It may be that sound spatialisation is insufficiently noticeable — perceived by the listener as 'merely a strengthening of the stereophonic effect'[113] or the immersion music already affords[114] — to justify new kinds of music media.

A significant challenge, and opportunity, is musical XR's potential to leverage the spatial, embodied and interactive capabilities of the medium and to enable creators to reimagine musical practices. Reflecting on the future artistic possibilities of the musical metaverse, Turchet and colleagues urge musical creators to 'focus on the aspects that are peculiar to the MM [musical metaverse], and all its underlying technologies', meaning the distributed character of musicians and audiences, and the multimodality of musical content. Within XR, there is already evident enthusiasm for the accessibility and immersivity that VR concerts can offer.[115] Notably, XR experiences, like many new media, bear traces of preceding media.[116] For instance, early 'VR concerts' (visual VR) feature familiar staged settings with a proscenium arch and a first person perspective from the venue floor; spherical videos present linear video of a performer from a fixed viewpoint; and AR tracks realise a mix by distributing its stems spatially around the listener. It may be that musical XR has yet to find a compelling use case for recorded music which exploits its unique affordances. For example, *MoonmoonsAR* offers limited interactivity by adjusting the relative amplitude of stems within a (pre-existing) track. But consider if musical

---

[112] Anna Meredith, 'Interview', 23 June 2021.
[113] Hodkinson, 'Creating Headspace: Digital Listening Spaces and Evolving Subjectivities', p. 170.
[114] Dyson, *Sounding New Media*.
[115] Kelsey E. Onderdijk and others, 'Concert Experiences in Virtual Reality Environments', *Virtual Reality*, 27.3 (2023), pp. 2383–96, doi:10.1007/s10055-023-00814-y.
[116] A parallel could be drawn with other digital innovations: for example, the first iterations of online supermarket shopping were virtual stores whose shelves shoppers could browse, much as they would in a physical store, but these were superseded by list-like browser interfaces because the virtual simulation of a physical store was redundant to the functional needs (and economic and social priorities) of online consumers.

form were specifically conceived to inhabit a space, utilising spatial placement and (geo)location as integral elements. *Melody of Dust* requires users to locate sound sources and combine them to trigger tracks, but it lacks conceptual coherence: the triggering sound objects are not stems of the triggered tracks, and the visual virtual environment is somewhat disconnected from the musical content. Imagine instead a conceptual unity between musical events and their representation in a virtual environment.

The preceding analysis highlights underexploited affordances of musical XR where music is integrated with its location, and/or is interactive and navigable. Listening experiences in AR/MR often employ visually represented, volumetric sound objects that are spatialised and responsive to listener movement. Yet audio augmented reality in which the digital mediated/synthesised sound is part of the surrounding environment remains largely unrealised. For example, neither *Bloom: Open Space* nor *Moonmoons* incorporates environmental sounds as part of the sound art. The convergence of spatialisation, computer listening, interactivity and geo-positioning technologies could make recorded music more responsive to its listening context, thereby expanding the scope of AR/MR.

Sound spatialisation in XR also provides further opportunities for interactivity and navigability, with significant implications for composition and listening. As in other forms of spatialized and installation-based sound art, musical XR may require listener movement to fully apprehend the work. When spatial configurations are preserved as the listener moves, agency is transferred to the listener, who can determine proximity to sound sources and shape the temporal unfolding of their experience. This contrasts with conventional listening, where music typically moves past a stationary listener, or sometimes induces a feeling of self motion.[117] Interactivity may also arise from environmental data detected by sensors. This approach might be preferable where direct listener interaction is undesirable (e.g. while driving).[118] Jack Atherton's call for a synthesis of 'doing' (intentional action) and 'being' (contemplation) in VR design suggests that XR experiences can foster human flourishing by balancing affordance of agency with reflection. In this spirit, XR offers the opportunity to create experiences that enable spatial and embodied

---

[117] Eric Clarke, 'Meaning and the Specification of Motion in Music', *Musicae Scientiae*, 5.2 (2001), pp. 213–34, doi:10.1177/102986490100500205.
[118] Collier, 'In Your Own Time: A Mobile Music Composition'.

interaction, shared participation in virtual and augmented environments, and contemplative immersion[119], perhaps by experiencing emancipatory subjugation to the artwork.

The interactivity afforded by musical XR signals a shift from linear to non-linear conceptions of musical form. As with interactive storytelling, XR invites musical structures that reward listener-centred participation. The case studies discussed illustrate how creators in 360° interactive contexts relinquish control over the listener's auditory and visual attention, movement and sequencing of events. XR's affordances of navigability and interactivity thus demand compositional approaches attuned to non-linear, participatory forms of musicking. Yet this redistribution of agency challenges established notions of authorship in Western art and popular music. Here, experimental and electroacoustic traditions, where similar issues have already been explored, may offer productive models for innovation.[120]

This paper investigated music in augmented, mixed and virtual reality as part of the broader ecology of recorded music culture. It has sought to reframe musical XR as both a continuation and a transformation of recording and listening practices. The novelty of this research is threefold. First, it offers an original synthesis of theories, studies and practices in musical XR and recorded music, consolidating insights previously scattered across disciplines and XR types into a unified conceptual framework. Second, it presents some of the first close readings of musical XR case studies, revealing how their aesthetic attributes reflect the medium's socio-technological affordances. Third, by contextualising recorded music within musical XR (VR, MR, AR) it reconsiders recording culture itself. Musical XR should be understood as part of the ongoing history of sound reproduction (and music-making technologies) and as a new iteration of the music commodity. It introduces possibilities for spatial composition, listener and sound source mobility, interactivity and non-linearity. However, it also compels us to reconsider the cultural processes through which listening practices become normalised, and to recognise the continual redefinition of what it means to hear, move, and make music in mediated space.

Acknowledgements

I would like to thank creators Anna Meredith and Arthur Carabott for giving up their time to participate in interviews for this research. I would also like to thank the participants in music

---

[119] Atherton and Wang, 'Doing vs. Being'.
[120] Parry, 'Navigating Sound'.

research colloquia (2018-2022) to whom I presented this research and whose questions prompted further reflection: City University, Durham University, Newcastle University, Rhodes University, Stanford University, University of Malta, University of Oslo, University of Oxford.

Copyright statement