Organised Sound (2026), 1–9
doi:10.1017/S1355771826101319

CAMBRIDGE UNIVERSITY PRESS

Article

From schizophonia to mediated auditory dislocation: Mediated tension in interactive VR soundscape composition

Boyi Bai
School of Languages, Arts and Societies, The University of Sheffield, UK

Abstract

The contemporary listening environment is characterised by extensive mediation, in which environmental recordings are continually detached from their original contexts and reconfigured elsewhere. Earlier soundscape research used the term schizophonia to describe the separation of sound from its source and the consequent result of a disembodied and unsettling sonic experience. However, this framing has come under growing criticism for regarding mediation as a defect. This article proposes an alternative concept, 'mediated auditory dislocation', to describe an analytically non-normative condition in which environmental sounds are recorded, displaced from their original time and place, and re-embedded within new audiovisual and interactive settings. The argument is developed through a practice-based case study of *Shifting Horizons* (Bai 2025), an interactive VR audiovisual soundscape work. The case study shows how control over sonic layering, electroacoustic processing, and audiovisual coupling allows listeners to reconfigure the relation between recorded sound and its mediated presentation, making mediation itself available for comparative and reflexive listening. The article concludes by introducing a three-layer collaborative media model that clarifies how VR can support comparative and reflexive listening through parameterised layer control. The model is proposed as a transferable analytic heuristic, intended to be tested and refined across further VR soundscape works.

Keywords: Mediated auditory dislocation; mediated tension; perceptual dissonance; schizophonia; VR soundscape composition

(Received 1 October 2025; revised 10 May 2026; accepted 11 May 2026)

1. Introduction

Every day listening is increasingly shaped by mediated sound, from streaming platforms and spatial-audio games to immersive virtual reality (VR) systems. These media do more than alter how sound is captured, spatialised, and reproduced; they also reconfigure how we understand the relation between sound and space and how we move between directly encountered environments and environments constructed through mediation. Consequently, the presentation of soundscape work can no longer be treated as a purely technical matter; it is entangled with ecological perception, cultural imagination, and social practice.

Emerging from the tradition of acoustic ecology, soundscape composition has long treated environmental recordings as creative material through which listeners are invited to attend to particular acoustic environments and their social–ecological implications. Within this lineage, R. Murray Schafer's concept of schizophonia describes the separation of sound from its source enabled by recording and amplification technologies (Schafer 1994). As sound studies have developed, however, the term has attracted sustained critique. Barry Truax (2001) argues that soundscape composition does not aim to faithfully restore an original scene; instead, it retains environmental recognisability while using temporal, spatial, and timbral transformation to intensify listeners' awareness of both environment and situated experience. John Levack Drever (2002) further connects soundscape composition to field recording and ethnographic practice, framing soundscape work as a situated and interpretive account of place rather than a neutral document. In parallel, an expanding body of practice has moved beyond the strictly auditory domain by combining field recordings and electroacoustic processing with installation and image-based media. This broader approach foregrounds how place can be co-constructed through sound and image within cross-media frameworks (Werner 2002; Fraisse et al. 2022).

VR is becoming a major platform for spatial audio, interactive music, and soundscape research. While spatial sound plays a critical role in VR immersion and presence, the broader functions and meanings of sound in VR still warrant further investigation (Bosman et al. 2024). Research at the intersection of VR soundscape composition and musical interaction has begun to examine how room-scale spatial audio, interactive sound objects, and navigable environments can support emerging modes of listening and creative practice (Turchet et al. 2023). However, much of this work either prioritises the technical implementation of spatial audio or treats sound primarily as a support for visual immersion. Limited theoretical engagement exists, particularly from soundscape composition and acoustic ecology perspectives, regarding how VR reorganises the relations among listening, place, and mediation.

When VR works bind sound to 360° video or highly realistic virtual scenes, this audiovisual coupling appears to reinforce the experience of returning to the source and has been adopted in applications such as urban soundscape assessment and

Email: bby9918@gmail.com
Cite this article: Bai B. (2026). From schizophonia to mediated auditory dislocation: Mediated tension in interactive VR soundscape composition. *Organised Sound*, 1–9, https://doi.org/10.1017/S1355771826101319

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

environmental education (Tang 2023). At the same time, it may also reassert an ideal of faithful reproduction, an ideal increasingly challenged in recent critiques of research on schizophonia. If contemporary listening is already embedded in pervasive mediation, the key question is not whether VR can restore an original unity between sound and place, but how it makes mediated relations perceptible through adjustable degrees of audiovisual alignment.

This article reframes schizophonia as 'mediated auditory dislocation' and examines the interactive VR soundscape composition Shifting Horizons (Bai 2025). Combining Ambisonics field recordings, electroacoustic transformation, 360° video, and interactive parameter control, the work does not attempt to eliminate 'mediated auditory dislocation'. Instead, it frames dislocation as a mediated tension that users can adjust through interaction. This makes it possible to examine how degrees of electroacoustic transformation and audiovisual alignment shape people's sense of place and environmental connection.

Based on this practice, the article proposes a three-layer collaborative media model comprising acoustic evidence, artistic intervention, and experience generation as an analytical framework. The model examines how VR's audiovisual and interactive structures establish relations among field recordings, electroacoustic transformations, and users' real-time operations. It is offered as a transferable heuristic for analysing and designing interactive VR soundscape composition in terms of comparative and reflexive listening.

2. Reframing schizophonia as mediated auditory dislocation

Schizophonia, one of the most influential concepts in Schafer's (1994: 88) soundscape theory, was originally coined to describe the separation of sound from its source produced by recording and amplification technologies, resulting in a disembodied, unsettling sonic experience. The term occupies an uneasy position in soundscape studies. While it has been conceptually generative, enabling early acoustic ecology to articulate an auditory crisis within modern media environments, it is increasingly difficult to sustain as a general diagnosis of mediated listening in contemporary contexts. Historically, the concept carried a clear ethical orientation; it cast technological reproduction as a threat to soundscape integrity while positioning the unmediated natural soundscape as normative.

Engaging disability studies, Mara Mills and Jonathan Sterne (2017) argue that Schafer explicitly intended schizophonia to function as an unsettling term linked to schizophrenia. This coupling of recording with psychiatric metaphor has been read as an ableist stance. It relies on imaginaries of mental illness rather than lived experience, and it frames technological mediation itself as a form of damage or impairment (Hall 2019).

Contemporary sound studies increasingly argue that schizophonia is ill-suited to account for pervasive mediated experience. The concept problematically treats an unmediated natural state as normative, thereby framing recorded experience as a deviation (Sterne 2003; Thompson 2004; Spring 2012). However, recorded listening is mediated from the outset. Sonic meaning and experience do not break at the moment of separation but are continuously reconfigured through the interplay of technology, culture, and listening conventions (Chion 1994; LaBelle 2006). What appears as separation is better understood as a historically specific listening condition and an emergent mode of listening shaped by technological infrastructures and cultural norms (Drever 2002; Lane and Carlyle 2013).

Critics within acoustic ecology also note that Schafer's notion of the soundscape is not a neutral description. It carries ideological and eco-ethical prescriptions about which sounds matter, often narrating soundscape history as a decline from nature and harmony towards machines and noise, and it can reflect a bias against urban soundscapes. In this way, ecological value becomes aligned with a normative imagination of the natural soundscape (Kelman 2010: 214). Even if such judgements had historical plausibility within early soundscape ethics, they benefit from repositioning within today's media ecology. Modern auditory experience, from broadcasting and cinema to headphone listening and platform-based audio, is fundamentally constituted through technical mediation, and the cultural meaning of sound is produced and circulated within these mediating structures (Sterne 2003).

Related insights also emerge from the electroacoustic tradition of 'acousmatic listening', which treats the loosening of sound from original reference as a positive aesthetic resource. Pierre Schaeffer's theory of 'reduced listening' brackets source-bonding so that sound can be encountered and analysed in terms of texture, energy, and form (Kane 2014). In his critique of Schafer, Francisco López (2001) argues that framing schizophonia as an abnormal condition misconstrues the artistic potential of working with recorded sound as autonomous material. The expressive force of environmental sound can arise precisely from forms of 'profound listening' that suspend habitual semantic and visual reference and focus on sound as such, allowing the listener to engage with its internal qualities and complex relations within a soundscape. Denis Smalley's (1997) theory of 'spectromorphology' likewise demonstrates that auditory meaning can emerge without reliance on identifiable sound sources, through the growth of sonic forms, their behavioural trajectories, and their spatial projection.

Taken together, these perspectives suggest that detaching sound from its source does not necessarily undermine the integrity of the soundscape experience. On the contrary, it provides a key methodological resource for contemporary soundscape practice by enabling a composable aesthetic tension between abstraction and reference and between form and context. Schizophonia is best understood as a historically specific mode of naming. It registers anxieties and value judgements about mediated transmission rather than offering a conclusion about the consequences of recording and transmission technologies. This article, therefore, proposes a more analytically neutral concept: 'mediated auditory dislocation'.

'Mediated auditory dislocation', as used in this article, does not name a pathological rupture. Rather, it refers to the structural condition whereby environmental sound, through recording, transmission, and reproduction, is displaced from its originating place and moment. Intermediality provides a useful point of departure for the present discussion. It refers to the intersections, relations, and crossings between different media forms and examines how media influence, depend on, and transform one another, thereby generating new modes of expression and perception that exceed the boundaries of any single medium (Grishakova 2024). From this perspective, 'mediated auditory dislocation' narrows the focus to the sonic dimension of such intermedial configurations. Once environmental sound is recorded and displaced, it is re-embedded within new audiovisual and interactive frameworks, where its relations to moving images and interface operations can become a site for negotiating anchoring and perceptual tension. Under these conditions of multi-layer media collaboration, listeners' understanding, attribution, and meaning-making are continuously shaped and redistributed.

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

The concept retains Schafer's core insight that mediation can, in a certain sense, separate sound from its source. However, it detaches this insight from the moral claim that such separation must be avoided or repaired. The point is less to mend a rupture than to confront an already established media structure and ask how it should be understood and mobilised. In this respect, the concept aligns more closely with Truax's (1996) view that environmental meaning arises from relationships among sound, context, and listener. Electroacoustic technologies, on this account, can alter these relationships and can also be used to strengthen them.

3. VR as relational reorganisation

Building on the framework of 'mediated auditory dislocation', VR soundscape composition can be understood as a media form that makes such dislocation explicit and allows it to be actively worked with. This section examines how VR reconfigures relations among sound, image, place, and listener.

3.1. Audiovisual and acousmatic mediation in VR

From a broader media perspective, modern auditory experience is already shaped by technical mediation through broadcasting, cinema, headphone listening, and digital platforms. When sound is recorded and reproduced, it does not simply lose its original place. Instead, it enters new relational networks in which spatial meaning, embodied perception, and cultural orientation are reconfigured within particular media environments (LaBelle 2006). In interactive VR soundscape composition, VR can turn the relation between sound and source into a mediatised parameter that can be designed, manipulated, and experienced, enabling users to configure and transform relations between sound and place deliberately.

Michel Chion (1994) uses the term 'synchresis' to describe how synchronised sound and image are perceptually fused into a single audiovisual event. On this view, sonic meaning is not confined to the sound's physical source. Visual information guides auditory attention and shapes temporal interpretation and causal attribution, producing what Chion calls the 'added value' of the audiovisual relation. In the VR soundscape composition examined here, this account helps specify one structural pole of the work. When the evidential recording remains closely aligned with the real visual scene, the sound is readily heard as belonging to what is seen, and the scene's sense of place is stabilised through audiovisual matching. Mary Ann Doane (1985) likewise shows that the semantics of sound can be reorganised within cinematic audiovisual structures. This matching does not merely reproduce a source but actively constructs spatial meaning within an audiovisual structure.

The same VR environment can also be structured to suspend this anchoring function of the visual without removing the visual itself. A contrasting pole emerges when less source-bound sound material is placed against the same real visual scene in a looser relation. Here, the image no longer serves as a reliable cue for causal attribution, so listening can shift towards an acousmatic tendency. In this mode, sound is attended to less as direct evidence of what is visible and more as composed material with its own texture, motion, and spatial behaviour. The work, therefore, stages both audiovisual matching and audiovisual separation within the same environment, making their contrast perceptible in experience.

3.2. Presence, immersion, and mediated tension

'Presence' is commonly understood as a subjective sense of 'being there', namely, the extent to which users experience themselves as located within a virtual environment rather than as external observers. Research on presence suggests that this experience is strongly shaped by sensorimotor consistency and coherence, not visual realism alone (Slater 2003). Spatialised audio, head tracking, and the coupling of visual cues with bodily action allow listeners to construct spatial structure through continuous feedback (Hendrix and Barfield 1996). Barry Blesser and Linda-Ruth Salter (2007) further argue that presence is not a passive replication of an original environment but is continually produced through bodily navigation and interaction with sonic events. From this perspective, spatial presence in VR is better understood as a perceptual practice that is continuously generated within a multimodal media structure.

Correspondingly, immersion is better approached as an experiential dimension shaped by system properties and interactional structure, rather than as a simple effect of audiovisual enclosure (Witmer and Singer 1998). In the VR soundscape composition discussed in this article, users' actions can reconfigure sonic organisation in real time, giving sound embodied malleability and turning listening into an exploratory practice (Slater and Wilbur 1997). At the same time, they introduce a controllable and variable degree of source-bonding, that is, the perceived attachment between sound and visible or imagined causes. This variable can be treated as a parameter for regulating the relation between sound and scene, creating a navigable range between strengthened representational coherence and sustained abstraction. Sound may at times align closely with visual objects, reinforcing presence (Grimshaw-Aagaard 2021), or be displaced, producing a layered field of reference. This continuous tension between matching and mismatching makes VR soundscape composition a processual experience, as listeners repeatedly revise their understanding of place (Chion 1994; Collins 2013).

To describe listeners' experience within these mediated configurations with greater precision, this article distinguishes three analytical levels. The first is 'mediated auditory dislocation' as a structural condition: environmental sound is extracted from its original spatiotemporal and sociocultural context through recording, editing, and reproduction, and placed within a new media structure. The second is 'perceptual dissonance' as a phenomenological effect: the confusion, conflict, and instability that arise when listeners encounter discrepancies between what they hear, see, and anticipate. The third is 'mediated tension' as a designable relation: the structured interplay between dislocation and dissonance that compositional strategies can stage and regulate. Perceptual dissonance is one route through which mediated tension becomes experientially legible, but tension can also arise through comparative parameterisation under relatively coherent audiovisual alignment.

For this article, the sequence of 'mediated auditory dislocation' (condition), 'perceptual dissonance' (effect), and 'mediated tension' (designable relation) functions as a set of core analytical tools for interactive VR soundscape composition. It enables analysis of how sound-image relations, temporal organisation, and interaction mechanics guide listeners across degrees of authenticity and artifice, prompting them to reconstruct both the soundscape's meaning and their own listening position within it.

4. Conceptual model for the VR soundscape composition: *shifting horizons*

4.1. Overview of the three-layer collaborative media model

Building on the preceding discussion, this article proposes a three-layer collaborative media model for analysing and designing

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

interactive VR soundscape composition. The model specifies three interdependent layers and the relations through which they co-produce meaning.

The first is the acoustic evidence layer. Grounded in original environmental recordings, it provides recognisable cues for place, events, and broader environmental relations. Through the selection and framing of recordings, it stabilises a set of sonic relations from a particular spatiotemporal segment. Those relations are extracted from their originating site and made available for repeated listening, comparison, and reinterpretation within a new presentation context.

The second is the artistic intervention layer. Through electroacoustic processing and sonic reorganisation, it makes mediation perceptible and foregrounds that sonic meaning does not emerge automatically from 'the environment'. Instead, meaning is continually arranged and reconfigured through compositional decisions and listening practices. By reshaping morphology, structure, and spatial distribution, this layer loosens habitual environmental semantics and opens alternative perceptual pathways.

The third is the experience-generation layer, constituted by VR spatial audio, panoramic visuals, head tracking, and interactive mechanisms. This layer translates tensions between the first two layers into a participatory interface of perceptual negotiation. Soundscape meaning is thereby treated less as a fixed endpoint to be reproduced than as an aesthetic variable that can be adjusted, compared, and reflected upon. Crucially, the three layers do not form a linear stack; they co-construct meaning through collaborative relations.

In *Shifting Horizons* (Bai 2025), the acoustic evidence layer is carried by the unprocessed field recordings anchored to each scene node. The artistic intervention layer is realised through the transformed strand and its associated sound processing strategies. The experience-generation layer is realised through the VR scene nodes, spatial audio rendering, head tracking, and the interactive controls through which listeners configure relations between the two sonic strands.

The following case analysis uses *Shifting Horizons* (Bai 2025) as a case study to demonstrate how these three layers are organised into an experiential pathway. Through a multi-scene structure, stratified sonic materials, and real-time interaction, the work guides listeners towards an auditory understanding of environment and place.

4.2. VR scenes and technical architecture

Shifting Horizons (Bai 2025) aims to operationalise the three-layer collaborative media model as an interface structure that listeners can directly manipulate. The work is based on field recordings made at Sheffield railway station in the United Kingdom, a major rail hub in South Yorkshire. Its soundscape offers a dense and readily recognisable set of sonic cues, including trains arriving and departing, public-address announcements, crowd movement, and the noise of surrounding urban traffic.

The work comprises six scene nodes corresponding to different platforms, tram stops, and the station entrance. Each node is anchored by a site-specific 360° video viewpoint and uses first-order Ambisonics environmental recordings as the sonic base. Importantly, the audiovisual content within each node is fixed (i.e., it presents a pre-authored recording-based scene), and interaction does not generate new events or alter the recorded actions captured in the footage. The electroacoustic layer applies spectral processing, temporal manipulation, and spatial redistribution to the original recordings. These processed and unprocessed strands are layered onto the VR foundation so that the three layers of the model are presented side by side within a single interface. With a VR headset, listeners experience the preconfigured scene nodes and actively adjust the balance and parameters of each layer via sliders and effect controls (Figure 1), thereby reshaping the perceived relationships between source, transformation, and audiovisual anchoring without changing the underlying recorded material.

The final version was presented on a PICO 4 headset and implemented in Unity (2022.3.50f1), with interaction handled via the default handheld controllers. Users aim at a slider handle, press the trigger to engage it, and drag to update the parameter value continuously. Alongside the two primary strand-balance sliders, additional effect controls are implemented using Unity's built-in audio processors – Param EQ, Compressor, Echo, Reverb, Distortion, Pitch Shifter, Chorus and Flange – each exposed as a continuous UI slider and applicable to both strands.

4.3. Layered sound design and effect sliders

Shifting Horizons (Bai 2025) presents two parallel strands. The first is an unprocessed recording grounded in the original environmental capture (acoustic evidence layer). The second is a composer-preconfigured electroacoustic transformation layer, derived from the original recordings through spectral reconfiguration, time-domain stretching, or spatial redistribution (artistic intervention layer). These two strands are controlled by two independent sliders, allowing listeners to determine the balance between the unprocessed field recording and its strongly electroacoustic counterpart (Figure 2). At the extremes, a listener may retain only the original recording or listen only to the electroacoustically transformed sound. More commonly, the two coexist in intermediate states, ranging from faint permeation to dense superimposition. The result is a series of hybrid forms that preserve indexical reference to the site while simultaneously exposing pronounced traces of artifice.

The significance of this design lies less in fine-grained volume mixing than in reframing the opposition implied by traditional readings of schizophonia. Rather than presenting original sound and electroacoustic transformation as mutually exclusive terms, the work renders their relation as a continuous, explorable plane. Listeners do not receive a fixed, composer-determined point on a natural or processed axis; instead, they occupy positions within a parameter space produced by the joint presence of both layers. They can intensify both layers to produce a saturated texture, attenuate either so that only distant traces remain, or privilege one while the other persists as residue. The work thus offers an explorable media surface in which judgements about realism or abstraction emerge as contingent relations produced by each specific configuration of parameters.

In addition, the two primary sliders are paired with additional effect controls implemented using Unity's built-in audio processors. This pairing gives the parameter space genuine compositional agency and constitutes a central manifestation of the artistic intervention layer. As a result, the original recording is no longer positioned as a natural, untouchable reference point. It is revealed as technical material that can itself be reshaped, and the seemingly more natural end of the continuum is no longer a stable baseline but a sonic substance continually reformed through interaction. Meanwhile, the drone is also not a fixed development that merely follows a preset trajectory. Through effects, it can be

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press



Figure 1. *Shifting Horizons* (Bai 2025) VR environment.

pulled back towards environmental coherence or pushed further towards abstraction.

Within this structure, a portion of the composer's decisions is translated into parameters that listeners can manipulate in real time, which is one key respect in which it differs from traditional fixed-media soundscape works. Listening thereby becomes a micro-level practice of re-composition. Each slider combination and effect setting constitutes a provisional stance on how environmental sound and mediating intervention can coexist, and it generates, within the work, a temporary version of soundscape meaning.

4.4. Audiovisual relations and perceptual dissonance

This interactive design is not a simple accumulation of interface functions. It operationalises the three-layer collaborative media model proposed above as a structure that listeners can directly experience within *Shifting Horizons* (Bai 2025). The acoustic evidence layer offers recognisable environmental cues and a sense of place through the original recordings, yet its apparent objectivity is continually unsettled by ongoing intervention through effect controls. Over time, listeners may recognise that what is presented as unprocessed environmental sound is already shaped by choices such as microphone directivity, dynamic-range compression, and post-production equalisation. When they apply filtering or spatial expansion to the raw layer, they are not merely adjusting timbre. They are experiencing, first-hand, how environmental sound can be converted into material for artistic organisation.

The artistic intervention layer is likewise not a single, fixed aesthetic position pre-inscribed by the composer. It becomes a latent structure that is reactivated through interaction. Each slider movement and parameter adjustment exposes the compositional logic of the processing strategies, allowing listeners to hear technical and aesthetic choices that are usually hidden behind the apparent closure of a finished work.

At the layer of experience generation, Ambisonics spatial audio and head tracking support the perception that, as listeners turn their heads and shift their gaze, they perceive corresponding changes in directionality, distance, and spatial envelopment. The 360° image provides each scene node with spatial cues of varying intensity. The critical point is that, as the work transitions between preconfigured scenes while listeners continually adjust the balance sliders and effect parameters, the tensions across the three media layers are superimposed onto embodied experience. Under different parameter

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

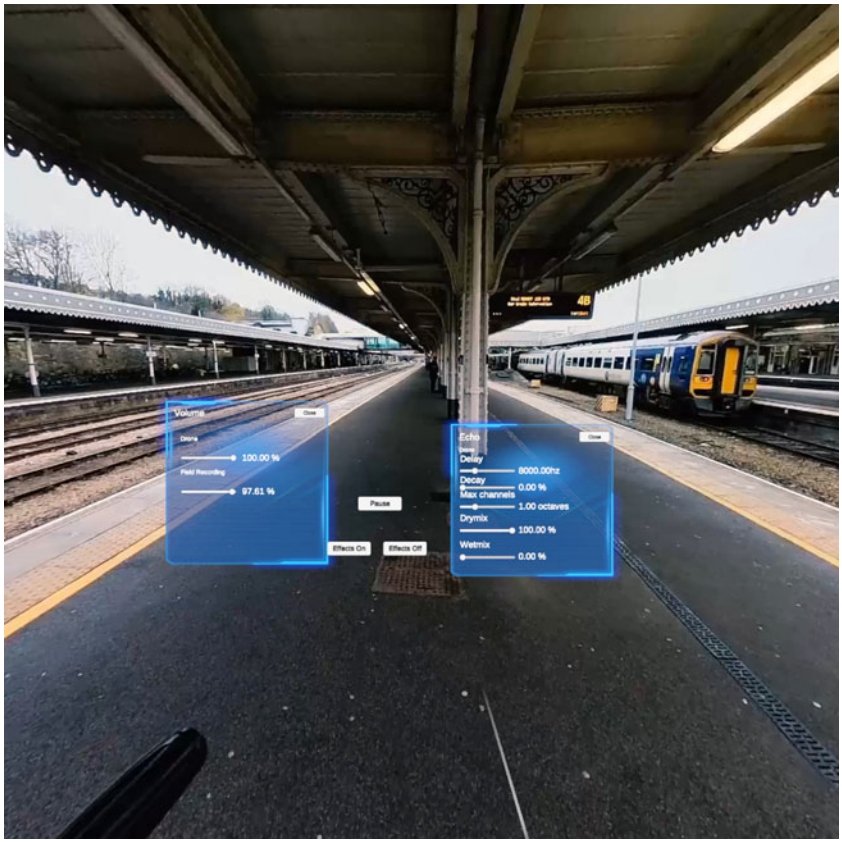


Figure 2. Interactive sliders in *Shifting Horizons* (Bai 2025).

configurations, the same site noise, rhythmic patterns, or sense of space can take on markedly different meanings.

Building on this, the visual layer of *Shifting Horizons* (Bai 2025) is also organised as a set of layered media cues rather than as a single, fixed explanatory image for sound. The 360° footage largely preserves the spatial structure of the recorded site. Yet beneath these seemingly faithful images, the sound may already have undergone pronounced time stretching and spectral reconstruction, creating subtle divergences from the events implied by the visuals. Listeners see a familiar railway station while hearing a sound field that is temporally dilated and spectrally exaggerated. For example, public-address announcements may be extended into harmonic clouds through granular synthesis and spectral processing, while crowd noise is compressed into continuous textures via dynamic-range processing. Here, the recognisability of the image and the abstraction of the sound produce 'perceptual dissonance'.

In some passages, the work deliberately brings sound and image into close temporal and causal alignment. The low-frequency rumble of an arriving train, the metallic impact of doors opening, and the trajectories of movement on screen correspond closely, while announcements broadly match the motion of passengers. This high degree of matching does not simply return the work to documentary representation. Instead, it may momentarily intensify presence by encouraging listeners to accept the premise of 'I am in this station'. As users adjust the sliders and introduce more electroacoustic material, this sense of being there may be disrupted or shift into a different condition. The same images can be overwritten by a newly configured sound field, and the same sound fields may acquire different interpretations under different images.

By continually moving between close alignment and explicit displacement, *Shifting Horizons* (Bai 2025) becomes an experimental site for testing how sound, image, and listening expectations shape one another. 'Mediated tension' is not presented here as a purely theoretical proposition. It is enacted through specific passages in which listeners can feel, at the level of embodied experience, how strengthening or weakening audiovisual coupling shifts their understanding of the station as a place and their judgement of what environmental sound is doing within the work.

4.5. Soundscape practice as an interface for media negotiation

From the perspective of sound composition, this design also extends how environmental sound can be listened to and understood. It neither dismisses the value of acousmatic approaches that attend closely to sonic form nor downplays the importance of visual and spatial cues for grasping environmental relations. It brings these orientations into co-presence within a single work structure through adjustable interaction.

When the sliders are set towards the original recording end, and visual cues remain clearly legible, listeners can more readily relate what they hear to infrastructures, spatial layout, and social behaviour. When the sliders move towards the electroacoustic end, and additional effects intensify abstraction, attention is redirected towards texture, energy, and trajectories of sonic motion. In the absence of a stable reference, listeners are prompted to imagine and project new interpretations onto these acoustic movements. By enabling comparative listening across multiple scenes and parameter settings, the work shifts environmental sound from an object to be faithfully reproduced into a perceptual event and listening practice that can be reinterpreted through multiple listening modes. In this sense, mediated tension is sustained not only

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

through audiovisual mismatch but also through comparative parameterisation across scenes and interaction states.

5. Discussion

5.1. Reconsidering VR soundscape composition through mediated auditory dislocation

Shifting Horizons (Bai 2025) is not merely a technical demonstration. It functions as an experiment in how soundscape composition might rearticulate its aims under contemporary conditions of mediation. As a practice rooted in acoustic ecology, soundscape composition has often been framed through ethical discourses of noise control and environmental protection. Under the premise of 'mediated auditory dislocation', VR soundscape composition can retain these concerns while shifting emphasis towards how soundscape meaning is generated, compared, and critically reflected upon through collaboration among multiple media layers.

This reorientation resonates with Truax's context-based composition (Truax 2018), where the central issue is not faithful reproduction of place but how contextual information and listeners' knowledge are mobilised to form a relational framework for listening. *Shifting Horizons* (Bai 2025) extends this approach by parameterising and visualising contextual relations, externalising configurations that are often implicit in compositional decision-making. This enables listeners to test different contextual formations within a single work rather than receive a single composer's stance.

Taken together, the two concepts proposed in this article carry implications that extend beyond their immediate application to VR soundscape composition and speak to a broader reconceptualisation of media relationships within an intermedial framework. 'Mediated auditory dislocation' reframes the sonic dimension of intermedial configurations not as a failure of fidelity but as a structural condition through which media actively produce new relational fields. Once environmental sound is recorded, displaced, and re-embedded within audiovisual and interactive settings, it does not simply coexist with image and interface but enters into mutual dependencies with them, reshaping how each medium's perceptual affordances are registered.

'Mediated tension', in turn, names the designable interplay through which different media layers negotiate their alignment, divergence, and mutual transformation. This mode of relation enables sound, image, and interactive control not merely to supplement one another but to continuously transform one another's meaning as the listener navigates the work. In this way, the model does not merely describe how sound and image interact; it also explains how multiple media forms collaborate to generate emergent perceptual modes and contextual meanings that cannot be reduced to any single medium. This proposition shifts intermedial analysis from describing how media cross or influence one another towards examining how their collaborative configuration can be actively composed and reflected upon.

5.2. Ethnographic listening and parameterised contextual configurations

In this sense, *Shifting Horizons* (Bai 2025) does not conduct systematic social ethnography in a conventional research-design sense. However, its adjustable two-strand structure and comparative listening across multiple preconfigured scenes offer a procedural parallel to ethnographic listening at the level of perception. Listeners are not instructed in advance about 'how this train station ought to be heard'. By repeatedly recalibrating the balance between raw and electroacoustic layers as the work moves between nodes, they encounter how the same infrastructure can appear, under different acoustic configurations, as mundane background activity or as an abstract sonic field. This comparative pathway resonates with the ethnographic work Drever (2002) advocates, in which one moves back and forth between environmental sound and its social context. The difference is that this movement no longer occurs between an artwork and an accompanying text; instead, it is compressed into the interactive structure of a single VR work.

From the standpoint of ecological listening, the model also offers a pragmatic way to speak about ecology under conditions of mediation. On the one hand, it acknowledges the unavoidable chain of technical intervention from recording to processing to reproduction. Rather than imagining a return to an unseparated original soundscape, it treats this chain as a lever for organising perception and reflection. On the other hand, it does not collapse into a general relativism in which all versions are treated as equivalent constructions. By providing multi-scene and multi-parameter contrasts, it offers concrete points of comparison that allow listeners to distinguish and evaluate different soundscapes. As Drever and Rennie suggest, contemporary soundscape and ethnographically inflected electroacoustic practices indicate that environmental and social acoustic meaning is often built gradually through cross-site juxtaposition and contrast, repeated replay, and shifting listening positions among composer, listeners, and audiences, rather than through a single immersive encounter with one location (Drever 2002; Rennie 2014). The design of *Shifting Horizons* (Bai 2025) embeds this comparative listening within the work itself. By listening across scenes and repeatedly returning to the same site under different parameter configurations, listeners are prompted to consider whether what they notice is simply a routine train station soundscape or whether they are beginning to register deeper environmental and social questions that the station's soundscape can disclose.

5.3. Immersion and interaction mechanisms in VR soundscape composition

At the level of immersive experience, the design of *Shifting Horizons* (Bai 2025) also resonates with research on game audio and immersive media. Grimshaw's research on first-person shooter (FPS) sound and immersion suggests that, in 'realism' FPS games, a sense of being in the virtual world is supported by the dynamic coupling between player action and the game engine's sonification processes. He further argues that this immersive effect does not only depend on strict sonic realism but can also be achieved through caricature rather than realism (Grimshaw 2008). Garner similarly emphasises that immersion is not determined unilaterally by device performance, such as rendering fidelity or tracking accuracy. It is also shaped by users' perceptual strategies and task structures. Immersion, in this view, is largely an emergent perception that depends on how listeners actively seek meaning within sonic cues (Garner 2017). Recent reviews focusing on head-mounted VR likewise indicate that spatialised audio and interactive sound design can significantly influence presence, affective intensity, and patterns of attention (Bosman et al. 2024).

Within this research trajectory, the contribution of *Shifting Horizons* (Bai 2025) lies not in proposing novel acoustic technologies but in reorganising established mechanisms into a parameterised apparatus for exploring soundscape meaning. The work is discussed here as a sound-led interactive audiovisual practice; while the

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

panoramic video remains fixed, its semantic function is continually reconfigured by adjustable sonic layers and their degrees of audio-visual coupling. Head tracking and Ambisonics provide baseline spatial coherence, while the 360° image offers visual anchors that can be partially aligned and partially displaced. The layered-sound sliders and effect controls translate questions such as 'What kind of place am I hearing?' and 'How close am I to that place?' into listening tasks that can be directly enacted through interaction. This approach points towards productive connections between VR soundscape composition and game audio studies.

VR's distinctive contribution, in this sense, is not the technical restoration of an original site. It lies in the integration of spatialised audio, audiovisual alignment, and interactive control into a unified interface that sustains listeners within a perceptible and adjustable field of 'mediated tension'. Within this field, the oppositions of authenticity and artifice, site and recording, and 'acousmatic listening' and visual reference are no longer treated as binary alternatives. They become continua that can be probed and negotiated through practice. While many of the perceptual mechanisms discussed here can also be achieved in multichannel loudspeaker settings or non-VR binaural systems, VR's distinctiveness lies in integrating action, audiovisual cues, and interactive control within a unified interface, enabling these mechanisms to be systematically linked and repeatedly tested.

5.4. Positioning and future directions of the three-layer collaborative media model

The three-layer collaborative media model proposed in this article can be read as an extension and specification of what Truax terms context-based composition. Truax (2018) argues that the central issue in soundscape composition is not whether an environment is faithfully reproduced, but how a work mobilises listeners' experience and imagination of a particular context so that internal sonic organisation can enter into a productive tension with knowledge of the external world. In *Shifting Horizons* (Bai 2025), the three layers, namely the acoustic evidence layer, the artistic intervention layer, and the experience-generation layer, operationalise this relational mechanism as a set of media structures that can be analysed and designed. The acoustic evidence layer stabilises auditory relations from a specific spatiotemporal segment. The artistic intervention layer makes audible how these pieces of evidence are reorganised in morphology and structure. The experience-generation layer then enables listeners, through head-tracked orientation and parameter adjustment, to continuously test their own judgements. In this sense, soundscape meaning is not fixed in advance; it emerges as a relational effect repeatedly produced within specific media configurations.

At the same time, both the model and the case analysis suggest directions for future development. First, *Shifting Horizons* (Bai 2025) currently focuses on transport infrastructure as an initial case study. Testing the model across other soundscape types, such as protected natural areas, religious spaces, markets, or community environments, would clarify how mediated structures shape soundscape meaning in these different contexts and what ecological implications follow. Second, the analysis is primarily grounded in the perspective of the maker and in theoretical construction. While discussion of listeners' experience is inferred from plausible perceptual pathways based on design intent, systematic audience research would provide empirical validation. This position complements, rather than competes with, recent VR audio literature that adopts psychoacoustic and user-research paradigms. Where that literature often foregrounds measurable presence and task performance, the present article focuses on how media structures reorganise soundscape meaning. Future work could build on this foundation by adopting mixed-method user studies that combine qualitative interviews with quantitative evaluation in order to examine more closely how listeners, in practice, interpret the relations among mediated place, place itself, and ecological understanding.

In addition, the present discussion remains focused on single-user head-mounted VR. This focus necessarily brackets the social affordances of shared loudspeaker listening and collective audience co-attention that have historically shaped the reception of soundscape work. Recent work on extended reality and collaborative environments has begun to explore multi-user sharing, collective presence, and distributed perception (Garner 2024). A key question for future research is how the three-layer collaborative media model might be extended to multi-participant soundscape practices and how listening might move from individual experience towards collective experience.

In sum, *Shifting Horizons* (Bai 2025) and the analytical framework proposed here treat schizophonia as a foundational tension in contemporary soundscape practice, a structural condition that must be acknowledged, worked with, and aestheticised. By materialising this tension in VR as a layered media structure and as adjustable pathways of listening, the article seeks to open a new line of dialogue among soundscape composition, interactive VR audiovisual practice, and acoustic ecology. Soundscape meaning is no longer imagined as an entity hidden within an original source awaiting faithful disclosure. It is treated as an ongoing process of generation, comparison, and revision across media configurations, listening strategies, and modes of participation.

6. Conclusion

This article reconsidered the analytical value of schizophonia in contemporary soundscape composition and argued that 'mediated auditory dislocation' provides a more analytically non-normative account of how environmental sound is routinely displaced and re-situated through recording, reproduction, and immersive media. The central hypothesis was that VR does not resolve dislocation; instead, it renders dislocation experientially negotiable, turning it into an explicit resource for composition and reflection.

Through a practice-based analysis of *Shifting Horizons* (Bai 2025), the study advances three findings. First, VR soundscape composition operates less as a project of restoring an original site than as a relational environment in which captured sound, transformed sound, and situated perception are continuously rebalanced. Second, the work's interactive architecture makes relations between audiovisual cues and sonic materials adjustable, generating perceptual friction that this article terms 'mediated tension'. Third, the three-layer collaborative media model clarifies how place-related meaning in VR is co-constructed across (1) evidential capture, (2) compositional intervention, and (3) experience generation through spatial audio, panoramic visuals, head tracking, and real-time control.

These findings contribute to the field in two ways. Theoretically, the article shifts discussion away from a deficit framing of mediation (in which separation is treated as pathology) towards an analytic vocabulary that specifies how mediation structures listening and place-relation. Practically, it offers a transferable framework for both designing and evaluating VR soundscape composition, foregrounding interaction and audiovisual alignment as parameters

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press

through which ecological and critical listening can be staged, compared, and reflected upon.

The study establishes a foundation for future work in several directions. The analysis centres on transport infrastructure as an initial case study; future testing of the model across diverse genres, site types, or production workflows would further validate its generalisability. In addition, the argument is primarily design-analytic rather than empirical; integrating systematic audience research would establish how different listeners, cultural contexts, or levels of VR familiarity shape perceptions of 'mediated tension' and patterns of control use.

Future research can therefore extend this work by (1) conducting comparative analyses across multiple VR soundscape projects to assess the generality of 'mediated auditory dislocation' and the three-layer collaborative media model; (2) integrating mixed-method audience research, including cross-cultural listening studies, to examine how interaction conditions attention, memory, and place-connection; and (3) exploring expanded configurations such as multi-user VR, accessibility-orientated design, and alternative spatial audiovisual strategies to investigate how 'mediated tension' changes under shared, constrained, or assistive listening conditions.

Overall, the article positions VR soundscape composition not as a route back to an unmediated original but as a rigorous medium for articulating and interrogating the mediated conditions through which contemporary experiences of place are formed.

References

Bai, B. 2025. *Shifting Horizons* (2025). First-order Ambisonics version (online video). https://www.youtube.com/watch?v=l0aUA3oaEk8 (accessed 18 December 2025).

Blesser, B. and Salter, L.-R. 2007. *Spaces Speak, Are You Listening? Experiencing Aural Architecture*. Cambridge, MA: MIT Press.

Bosman, I. d. V., Buruk, O., Jørgensen, K., and Hamari, J. 2024. The Effect of Audio on the Experience in Virtual Reality: A Scoping Review. *Behaviour & Information Technology* **43**(1): 165–99. https://doi.org/10.1080/0144929X.2022.2158371.

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

Collins, K. C. 2013. *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9442.001.0001.

Doane, M. A. 1985. The Voice in the Cinema: The Articulation of Body and Space. In E. Weis and J. Belton (eds.) *Film Sound: Theory and Practice*. New York: Columbia University Press, 169–71.

Drever, J. L. 2002. Soundscape Composition: The Convergence of Ethnography and Acousmatic Music. *Organised Sound* **7**(1): 21–27. https://doi.org/10.1017/S1355771802001048.

Fraisse, V., Giannini, N., Guastavino, C., and Boutard, G. 2022. Experiencing Sound Installations: A Conceptual Framework. *Organised Sound* **27**(2): 227–42. https://doi.org/10.1017/S135577182200036X.

Garner, T. A. 2017. *Echoes of Other Worlds: Sound in Virtual Reality – Past, Present and Future*. Cham: Palgrave Macmillan (Palgrave Studies in Sound).

Garner, T. A. 2024. The Fundamental Frequency: Extending Sound Perception Theory to Extended-Reality Collaborative Environments. *Computers & Education: X Reality* **5**: 100080. https://doi.org/10.1016/j.cexr.2024.100080.

Grimshaw, M. 2008. Sound and Immersion in the First-Person Shooter. *International Journal of Intelligent Games & Simulation* **5**(1): 119–24.

Grimshaw-Aagaard, M. 2021. A Step Back from Reality: Sound and Presence in Computer Games and Other Worlds. In M. Fritsch and T. Summers (eds.) *The Cambridge Companion to Video Game Music*. Cambridge: Cambridge University Press, 269–83. https://doi.org/10.1017/9781108670289.017.

Grishakova, M. 2024. Intermediality: Introducing Terminology and Approaches in the Field. In J. Bruhn, A. L. V. Azcárate and M. de Paiva Vieira (eds.) *The Palgrave Handbook of Intermediality*. Cham: Palgrave Macmillan, 13–29. https://doi.org/10.1007/978-3-031-28322-2_12

Hall, M. C. 2019. Critical Disability Theory. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy (Winter 2019 edn)*. Stanford, CA: Stanford University. https://plato.stanford.edu/archives/win2019/entries/disability-critical/ (accessed 18 December 2025).

Hendrix, C. and Barfield, W. 1996. Presence within Virtual Environments as a Function of Visual Display Parameters. *Presence: Teleoperators and Virtual Environments* **5**(3): 274–89. https://doi.org/10.1162/pres.1996.5.3.274.

Kane, B. 2014. Pierre Schaeffer, the Sound Object, and the Acousmatic Reduction. In *Sound Unseen: Acousmatic Sound in Theory and Practice*. New York: Oxford University Press, 15–41. online edn, Oxford Academic, 19 June 2014. https://doi.org/10.1093/acprof:oso/9780199347841.003.0002.

Kelman, A. Y. 2010. Rethinking the Soundscape: A Critical Genealogy of a Key Term in Sound Studies. *The Senses and Society* **5**(2): 212–34. https://doi.org/10.2752/174589210X12668381452845.

LaBelle, B. 2006. *Background Noise: Perspectives on Sound Art*. New York: Continuum.

Lane, C. and Carlyle, A. 2013. Introduction. In C. Lane and A. Carlyle (eds.) *In the Field: The Art of Field Recording*. Axminster: Uniformbooks, 7–14.

López, F. 2001. Profound Listening and Environmental Sound Matter. In C. Cox and D. Warner (eds.) *Audio Culture: Readings in Modern Music*. New York: Bloomsbury Academic, 103–08. https://doi.org/10.5040/9781501318399.ch-012.

Mills, M. and Sterne, J. 2017. Afterword II: Dismediation—Three Proposals, Six Tactics. In E. Ellcessor (ed.) *Disability Media Studies*. New York, NY: NYU Press, 365–78. online edn, NYU Press Scholarship Online, 24 May 2018. https://doi.org/10.18574/nyu/9781479867820.003.0017.

Rennie, T. 2014. Socio-Sonic: An Ethnographic Methodology for Electroacoustic Composition. *Organised Sound* **19**(2): 117–24. https://doi.org/10.1017/S1355771814000053.

Schafer, R. M. 1994. *The Soundscape: Our Sonic Environment and the Tuning of the World*. Rochester, VT: Destiny Books.

Slater, M. 2003. A note on presence terminology. *Presence Connect* **3**: 2–4. https://s3.amazonaws.com/publicationslist.org/data/melslater/ref-201/a%20note%20on%20presence%20terminology.pdf (accessed 18 December 2025).

Slater, M. and Wilbur, S. 1997. A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* **6**(6): 603–16. https://doi.org/10.1162/pres.1997.6.6.603.

Smalley, D. 1997. Spectromorphology: Explaining Sound-Shapes. *Organised Sound* **2**(2): 107–26. https://doi.org/10.1017/S1355771897009059.

Spring, K. 2012. Walk This Way: The Pedagogical Value of Soundwalking to the Study of Film sound. *Music and the Moving Image* **5**(2): 34–42. https://doi.org/10.5406/musimoviimag.5.2.0034.

Sterne, J. 2003. *The Audible Past: Cultural Origins of Sound Reproduction*. Durham, NC: Duke University Press.

Tang, J. 2023. Research on the Application of Sound in Virtual Reality. *Highlights in Science, Engineering and Technology* **44**: 206–12. https://doi.org/10.54097/hset.v44i.7323.

Thompson, E. 2004. *The Soundscape of Modernity: Architectural Acoustics and the Culture of Listening in America, 1900–1933*. Cambridge, MA: MIT Press.

Turchet, L., Carraro, M. and Tomasetti, M. 2023. FreesoundVR: Soundscape Composition in Virtual Reality Using Online Sound Repositories. *Virtual Reality* **27**(2): 903–15. https://doi.org/10.1007/s10055-022-00705-8.

Truax, B. 1996. Soundscape, Acoustic Communication and Environmental Sound Composition. *Contemporary Music Review* **15**(1): 49–65.

Truax, B. 2001. *Acoustic Communication*, 2nd edn. Westport, CT: Ablex.

Truax, B. 2018. Editorial: context-based composition. *Organised Sound* **23**(1): 1–2. https://doi.org/10.1017/S1355771817000218.

Werner, H. U. 2002. MetaSon #5 Skruv Stockholm: Turning Schizophonic Sound into Audiovirtual Image. *Organised Sound* **7**(1): 73–78. https://doi.org/10.1017/S1355771802001115.

Witmer, B. G. and Singer, M. J. 1998. Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments* **7**(3): 225–40. https://doi.org/10.1162/105474698565686.

https://doi.org/10.1017/S1355771826101319 Published online by Cambridge University Press