

Sound design in VR filmmaking: Opportunities and challenges in a new medium

Convergence: The International Journal of Research into New Media Technologies
2025, Vol. 0(0) 1–30
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/13548565251339304
journals.sagepub.com/home/con



Christos Manolas , **Panos Raptis**  and **Mark Durham**

Ravensbourne University London, UK

Abstract

Cinematic sound design workflows are well-established and refined over many years of practice and scholarly review, and the creative and technical approaches resulting from this are widely adopted and used. Until recently, these approaches were partially dictated by the conventions and limitations of the underlying visual medium: the cinematic screen. Nevertheless, the ongoing technological revolution brings radical changes in the way we approach and consume creative media content and the accompanying soundtracks. Long-established cinematic sound design conventions are challenged and transformed, as emerging cinematic sound systems allow for greater creative freedom and more precise sound spatialisation. At the same time, advanced virtual reality (VR) systems provide an opportunity to break the boundaries of the screen and allow filmmakers to explore new ways of creative expression. With this comes an increased demand for specialised audio content that utilises the unique technological and creative characteristics of the VR medium. Through discussions and commentary based on secondary research and a series of interviews with sound design and filmmaking practitioners, this study aims at identifying some of the creative and practical opportunities and challenges sound designers may encounter when working in VR projects with a cinematic narrative structure and orientation. Ultimately, the study seeks to assist sound designers, VR filmmakers, and immersive media practitioners in better contextualising and defining the potential roles of the soundtrack in VR filmmaking and the associated workflows, contributing to the ongoing dialogue as the language of the new medium is being developed.

Keywords

Virtual reality, VR, filmmaking, sound design, soundtrack

Corresponding author:

Christos Manolas, The Screen School, Department of Film and Media, Ravensbourne University London, 6 Penrose Way, London SE10 0EW, UK.

Email: c.manolas@rave.ac.uk

Introduction

Cinematic sound design workflows are well-established and refined over many years of practice and scholarly review. Today, the creative and technical approaches resulting from this process are adopted by the majority, if not the entirety, of sound designers and filmmakers. Until recently, these approaches were partially dictated by the conventions and limitations of the underlying visual medium: the cinematic screen. Nevertheless, the rapid technological advances of the past decades brought about radical changes in the way we approach and consume creative media content and the accompanying soundtracks. From the introduction of advanced immersive recording techniques and hardware and the adoption of object-based audio to the associated changes in producing deliverables for a multitude of reproduction systems, the sound production paradigms are constantly challenged and transformed. This is also affecting the long-established cinematic sound design conventions, with the popularisation and wide adoption of systems like Dolby Atmos ([Dolby Laboratories Inc, 2015](#)) allowing for much greater creative freedom, more precise sound spatialisation, and easier deployment of the soundtrack deliverables through all mainstream commercial reproduction systems.

One of the areas that pose particular interest in this context is the potential offered by *VR filmmaking*. With this comes an increased demand for specialised audio that utilises the unique technological and aesthetic characteristics of the VR medium. While many of the established sound design and production approaches are also applicable and effective in VR filmmaking, the medium is still in its infancy and sound designers are bound to face a new set of challenges that were not present in the production models of the past. Some aspects of sound design that may be affected by this are relatively easy to identify. For instance, the importance of sound as a guide of visual focus may be pronounced 360° viewing space as opposed to a screen, while the spatial expansion of the soundtrack around the audience may be pivotal in creating an immersive sonic environment that aligns with the spatial characteristics of the VR medium. However, sound design is a nuanced and intricate process that cannot be fully described and analysed thoroughly at such an abstract level. Uncovering the true potential and unique role of sound in the context of VR filmmaking requires an understanding of current practice coupled with an educated guess of the trajectories of the emerging medium.

This study aimed at identifying some of the technological and creative opportunities and challenges sound designers may encounter when working in VR film projects. Through preliminary analysis based on secondary research, the following areas of potential interest were identified:

- Guiding visual attention within the VR viewing space.
- Immersion and envelopment through sound spatialisation.
- The use of low-end frequencies in VR films.
- Sound design for VR films versus video games.

This formed the basis for a series of interviews with selected practitioners, culminating in the construction of the following themes:

- VR film space navigation through visual and auditory cues.
- Perceptual effects of low-end frequencies and their implications on VR film sound reproduction.
- Assisting the VR film sound designer through hybridisation of workflows and tools.

The identified themes were used for further discussion and reflection, with the aim to assist sound designers and filmmakers contextualising and defining challenges and opportunities associated with the creation and delivery of the VR film soundtrack.

Background

Cinematic sound spatialisation: common approaches and conventions

Any discussion of potential challenges and opportunities for sound design in VR filmmaking, to some extent, is bound to be referenced against established sound spatialisation approaches in screen-based cinema. While an exhaustive analysis of cinematic sound spatialisation workflows and techniques exceeds the scope of this study, three common themes, namely spatialisation of:

- Dialogue,
- Sound effects, and
- Ambience and background sounds, music, and reverberation effects

are briefly presented below. From the perspective of creative practice, the approaches described below generally apply to both the *discrete surround* systems of the past (Dolby Laboratories Inc, 1997) and contemporary *object-based* systems like *Dolby Atmos* (Dolby Laboratories Inc, 2015)

Dialogue

With the exceptions of highly experimental cinematic forms, it is generally accepted that *dialogue clarity* is paramount. This is not only because dialogue conveys essential narrative information, but also because it requires ‘additional cognitive processing [for the viewer] to derive both literal and implied meaning’ (Beauchamp, 2005: 146). It is, thus, unsurprising that one of the priorities of sound designers is to keep dialogue as clear and distinct as possible.

This inevitably influences dialogue spatialisation decisions (Kerins, 2011: 71). By positioning the dialogue on specific channels (e.g. centre-front channel in a surround system or an Atmos bed; Figure 1) and keeping it separate from the rest of the soundtrack components, filmmakers avoid several intelligibility and perception issues (Dolby Laboratories Inc, 2005: 5-1).

As an example, if dialogue were placed on the rear-left channel of a surround soundtrack to reflect the supposed position of a character in relation to the screen it could blend with other soundtrack components that commonly occupy this space, such as sound effects, ambiences, background music, or reverberation effects. In addition, dialogue clarity could be compromised for audience members seated away from the speakers carrying the spatialised channel signal.

From a compatibility perspective, sound designers must also ensure that the dialogue is clearly audible across a range of different sound systems with varying channel configurations (Kerins, 2011: 14, 44). Thus, dialogue is almost exclusively positioned in a central position aligned with the screen and guaranteed to exist in all possible reproduction configurations (i.e. centre-front, reproduced through a centre channel or as a stereo phantom image over the left and right channels). This ensures clarity, intelligibility, and scalability across different sound reproduction systems, such as theatrical and home cinema speaker arrays, mono or stereo speakers, and headphones.

Other factors also contribute to the rather conservative use of dialogue spatialisation in screen-based cinema. For instance, ‘audiences focus more on the content rather than the placement of

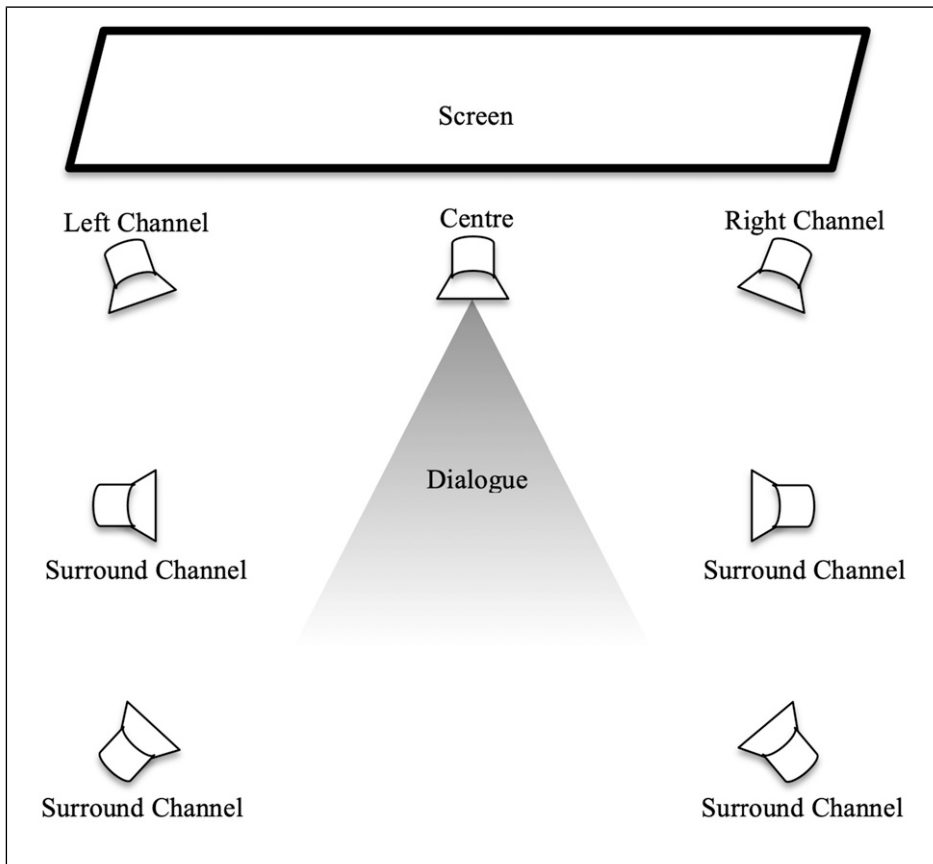


Figure 1. Typical dialogue spatialisation.

dialogue in the mix' (Beauchamp, 2005: 146) and sound is drawn towards the image when information is available to both the visual and auditory sensory streams (Chion, 1994: 69).

It is true that the evolution and wide adoption of advanced spatial audio systems like *Dolby Atmos* (Dolby Laboratories Inc, 2015), and the resulting change of audience expectations have led to less rigid approaches towards dialogue spatialisation (Kerins, 2011). Yet, spatialisation decisions are still commonly based on the tried and tested approaches of the past 60 years: dialogue usually emanates from the front channels for both on-screen and off-screen characters.

Sound effects

In contrast to dialogue, spatialisation of sound effects outside the screen space is standard practice in cinematic sound design. Since the introduction of surround sound systems (Dolby Laboratories Inc, 1997), spatial sound effects have been routinely used to enhance immersion, amplify the impact of visual events, or extend the cinematic world beyond the boundaries of the screen (Figure 2). Helicopters flying overhead, bullets and arrows whizzing through the theatre, or 'the roar of a monster in the back' are classic examples (Vanhoutte et al., 2010: 11).

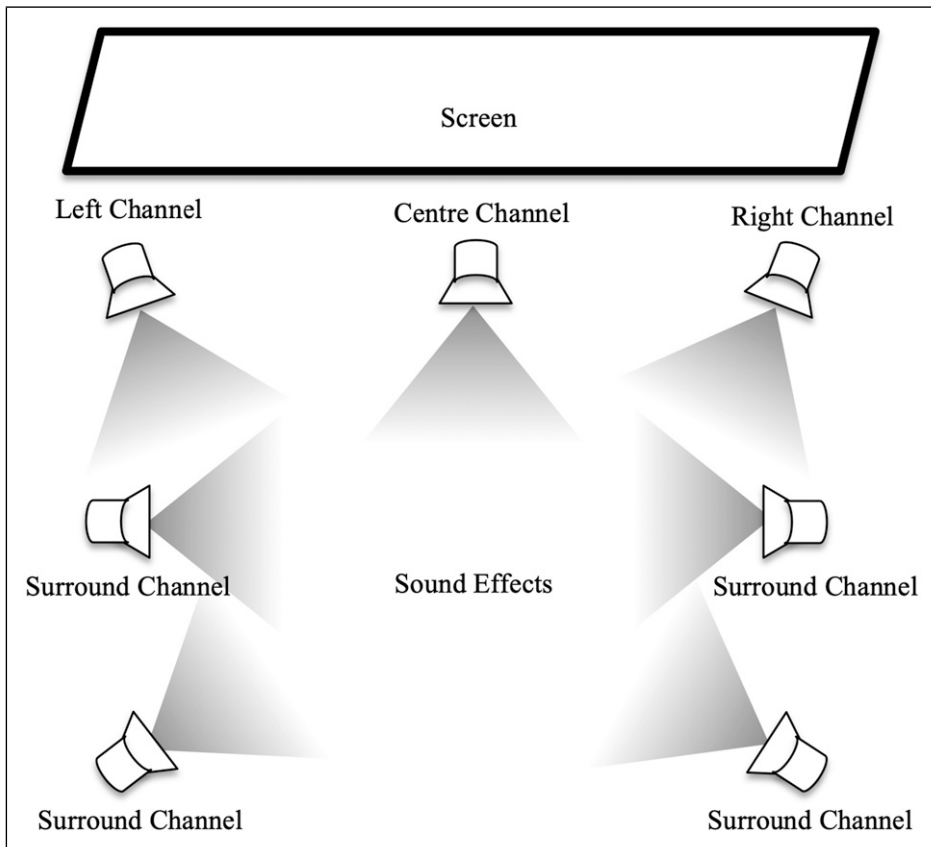


Figure 2. Spatialisation of on-screen and off-screen sound effects.

This is unsurprising, as sound effects are not as critical to the narrative as dialogue; filmmakers can be more experimental without risking important information being overlooked or obscured. Nevertheless, even though the spatialisation of sound effects is certainly more common than dialogue, it is still applied selectively and in moderation. There is a good reason for this: filmmakers do not want to divert viewers' attention away from the screen.

Ambience and background sounds, music, and reverberation effects

When it comes to ambience and background sounds, music, and reverberation effects, spatialisation approaches may differ significantly among sound designers and filmmakers. However, it is common practice to use spatialisation to create *acoustically immersive* spaces (Figure 3). The purpose of this is to immerse the viewer into the audiovisual experience through a sense of *envelopment* rather than to expand the cinematic world into an *acoustically realistic space*. For example, the expansion of a non-diegetic musical piece towards the rear of the cinema may be used to surround the audience but is not an attempt to recreate a realistic event or environment.

In other cases, the goal is indeed to reinforce the illusion of the cinematic world expanding beyond the screen. Typical examples are battlefield ambiances surrounding the audience during a realistic battle

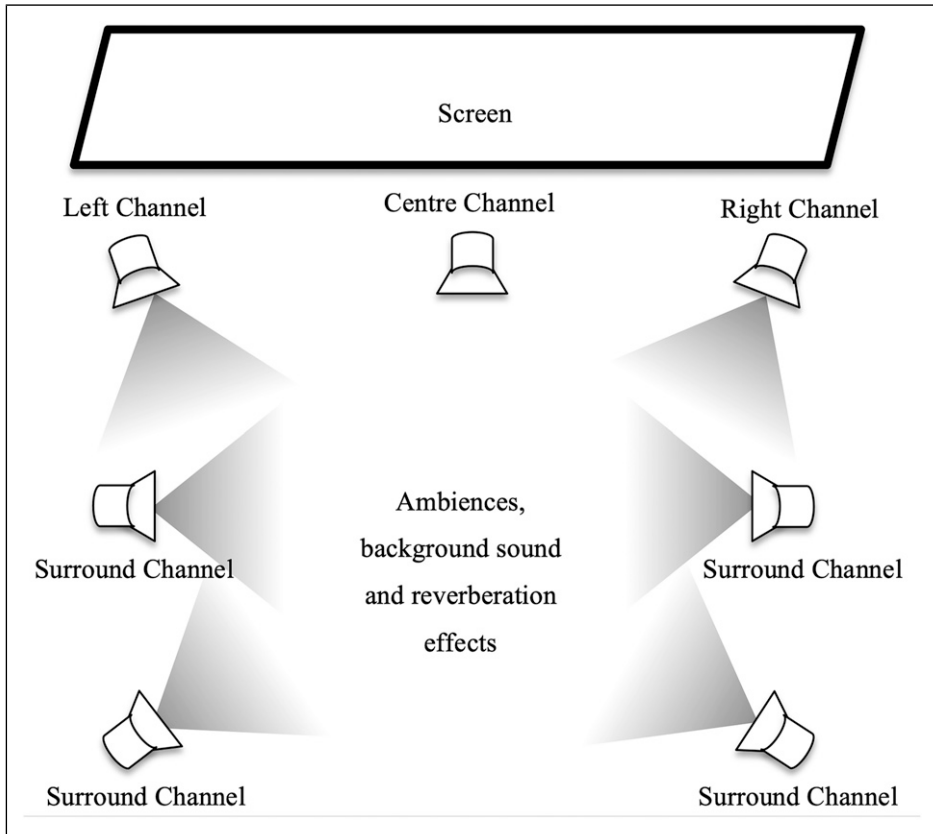


Figure 3. Spatialisation of ambience and background sounds, music, and reverberation effects.

scene or multichannel reverberation effects as the protagonist enters a cave. Although such uses of sound spatialisation are meant to create a more convincing and enveloping environment in a physical sense, this is done in a rather *abstract, pseudo-realistic* manner. The intention is to create a *general sense of expansion* of the auditory space rather than an accurate representation of exact locations or events.

With the emergence and wide commercial adoption of advanced spatial audio systems and workflows like Dolby Atmos (Dolby Laboratories Inc, 2015), object-based audio, and binaural rendering (Rumsey, 2016), sound designers are encouraged to adopt increasingly adventurous and bold spatialisation approaches. Nevertheless, in screen-based cinema, sound spatialisation decisions to a large degree are still informed by the (front-side) *biased visual orientation of the medium*. This is important in the current context, as the limitations and constraints of screen-based cinema are *either absent or significantly reduced in VR*. In VR filmmaking, sound designers may question and review many of the established sound spatialisation strategies and approaches that have been the conceptual basis of their work for decades.

Framing and guiding the narrative

In screen-based filmmaking, the frame is the primary device for the *mise-en-scene*. The action is framed according to well-established compositional techniques and devices, such as *depth of field*,

rule of thirds/golden ratio, leading lines, centering/symmetry, or internal framing (Zettl, 1990). The filmmaker expands the narrative world beyond the frame using both *camera movement* and the *soundtrack*, through *off-screen auditory cues* and *sound spatialisation*. In VR, the absence of a traditional frame and the boundless staging area – extending to the entire enveloping space – present both a challenge and potentiality. The viewer is surrounded by the narrative world and is able to look or walk around the environment. The field of vision – or the more restricted field of view of the headset – *becomes the frame*, which is directed by the viewers as they engage with the narrative. In this context, Mast's (1984) analysis on the nature of the film frame provides an insightful perspective: the film frame is unlike a painting frame or a theatre proscenium; frame boundaries can be breached, if needed. He describes the frame not as a *passive container*, but an *active signifier*, shifting, refocusing, reestablishing significations, thus referencing Bazin's (1967) view of a *window* with a dynamic centrifugal boundary and Münsterberg's (1916) analogy of the frame to the *mental act of attention*. Adopting this perspective in VR filmmaking, the absence of the screen-based frame allows for the orchestration of dynamic loci of attention – spatial and narrative – to direct the viewers' engagement with the story and characters.

There is a rich and growing body of research on guiding the viewers' attention in VR. While a systematic review of this subject exceeds the scope of this study, the reader can find an indicative taxonomy in the work of Rothe et al. (2019). Wang et al. (2021) have also assessed and rated visual cues, such as *localised brightness*, *saturation*, *blurring*, and *superimposed attention signifiers* (e.g. logos or graphics). Unsurprisingly, a number of these techniques are common in filmmaking where brightness, lighting, and colour variation are used to guide the eye to different parts of the image, through cinematography and grading.

Irrespective of the exact techniques used, for visual cues to be effective *they must originate within the field of view*, which in VR is constantly shifting following the story. Gödde et al. (2018) suggest placing the significant action in the *frontal 180 degrees*, referred to as the *initial viewing direction (IVD)* or the *main content zone* (Ozenua Oluwatobi, 2022; Figure 4), and then use *attention shifting cues* to move around the space.

Gödde et al. (2018) assess the viewers' preference for the IVD and the reluctance to turn towards actions behind them. They describe the technique as a compromise favouring the seated viewer, but they interestingly note that even standing viewers in an environment with little directional differentiation or emphasis, ended up mostly orienting themselves towards the IVD. Despite the viewers' bias to the IVD, the implementation must be *narratively consistent*. For example, leaving large parts of the environment devoid of action risks disrupting the authenticity of the story world.

One of the benefits of arranging the action within the field of view and shifting the viewers' attention is that it echoes the conventional filmic staging of the narratively significant action in front of the camera and subsequently motivating a camera movement. Nevertheless, this is not the only possible approach in VR. Palma Stade et al. (2023) cite Williams et al. (2022) distinguishing between:

“‘Newtonian’ approaches, where the action is predominantly arranged in front of the viewer, requiring little or no movement of the viewer’s head or body, and ‘Copernican’ approaches, where the action requires the viewer to turn to view the action’ (Williams et al., 2022, cited in Palma Stade et al., 2023).

The ‘Copernican’ approach is implemented in the film *HELP* (2016). The characters and fast-paced action surround the viewer with *sound cues accentuating actions and guiding attention*. The role of the soundtrack is emphasised – especially in spatially dense narratives – mitigating viewers' anxiety of missing out narratively significant actions and disengagement from the story.

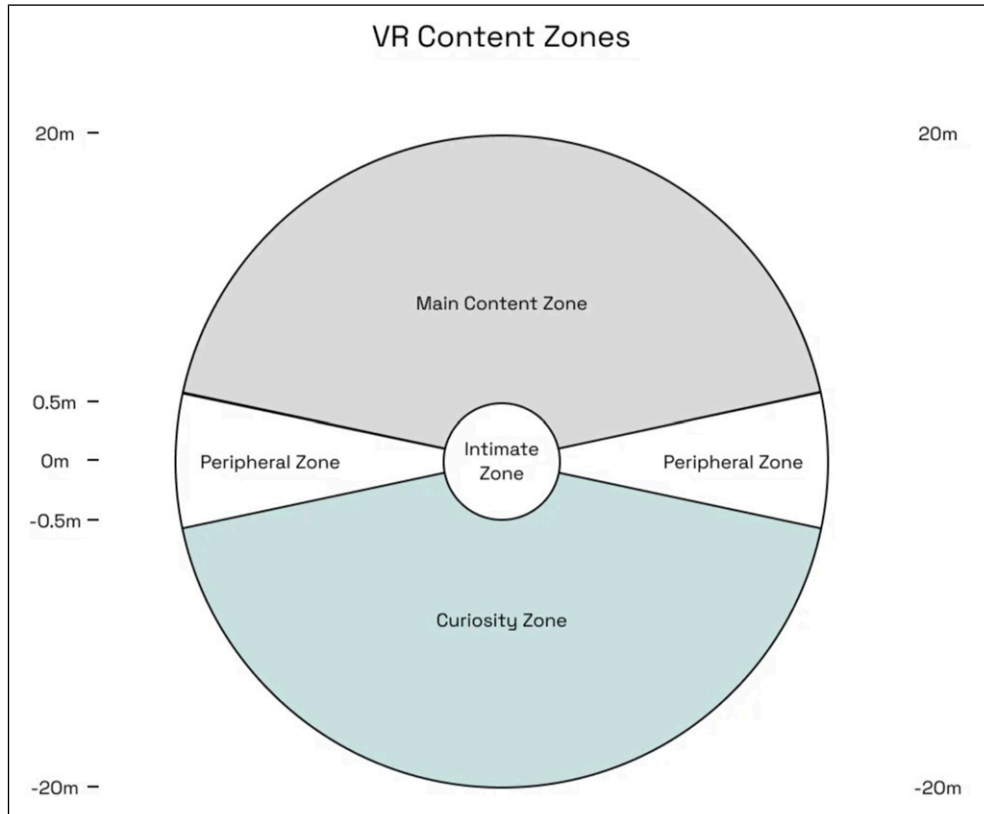


Figure 4. VR viewing space with content zones (Ozenua Oluwatobi, 2022).

In this context, the soundtrack is used to carry additional tasks compared to screen-based filmmaking. The effectiveness of such approaches is based on *cross-modal integration*, the mechanism at the centre of human perception (Bolognini et al., 2005; Holt, 1997; Shams and Kim, 2010; Vroomen and De Gelder, 2000). At an abstract level, all our senses fundamentally serve the same function: providing information about our environment. Our brain anticipates that events or objects in our surroundings would be detectable through more than one sensory stream. This capability 'can enhance perceptual clarity and reduce ambiguity about the sensory environment' (Lippert et al., 2007: 102).

The VR film soundtrack as a visual attention guide

From a cinematic perspective, cross-modal integration can be examined both in terms of the *physical delivery of sensory information* to the audience and the *internal perceptual mechanisms* as the audience interprets and translates this information. The latter can attach meaning and context to the incoming sensory information, unifying and transforming it into a multisensory experience that transcends the boundaries of the individual sensory channels. This perceptual reading of the cinematic information, described by Chion as 'rendering' (Chion, 1994: 109-114), allows filmmakers to convey complex multisensory messages, although the *delivery of information* primarily

involves two sensory streams: *visual* and *auditory*. In more immersive formats, such as the so-called 4D-9D cinema experiences (Owatch, 2024), *olfactory* and *tactile cues* are also used to provide an increased sense of realism, while in VR the use of tactile feedback is standard practice. Irrespectively of the combination of sensory cues used, the purpose is generally the same: to provide information regarding objects or events considered *important for the narrative*. In this study, the focus is on *audiovisual integration*, arguably the most essential subset of cross-modal integration for the delivery of sensory information in both screen-based and VR filmmaking.

A key aspect of the audiovisual integration mechanism is to *guide visual attention* (Störmer, 2019). In everyday life, audition alerts us of possible events that *may require visual attention*. In screen-based cinema, the use of this mechanism is relatively limited. While there are instances where spatial audio cues are used to guide visual attention to a particular area of the screen, there are reasons to avoid using such approaches too heavily.

Firstly, *visual action* is contained within the boundaries of the screen, firmly within the field of vision and easily scanned even in the absence of auditory cues. Secondly, although off-screen sounds are commonly used for establishing settings and actions that are not visible on-screen, exaggerated spatial displacements of auditory cues can lead to the *in-the-wings* or *exit door effect* (Chion, 1994: 83; Kerins, 2011: 159), a distraction of attention away from the on-screen action. This can potentially lead to violations of the *suspension of disbelief* the film seeks to maintain (Abrams et al., 2001: 211). Thirdly, for events detected both visually and aurally, visual cues generally influence perception more heavily (Meares, 1993: 172). This phenomenon is known as *visual dominance* (Colman, 2009) and is widely used in cinema in the form of the *ventriloquism effect* or *spatial magnetisation* of sound by image (Bregman, 1994: 183, 290; Chion, 1994: 69-71; Woszczyk et al., 1995: 4). Once visual focus is established, audiences are likely to *attribute the associated sounds to the visual cues*, even if there are *significant spatial discrepancies between the two streams*. A typical example is the use of the centre channel for dialogue reproduction while characters are off-centre. In this case, the dialogue is:

‘perceived as if it were coming from a point in the visual environment where something is moving or changing [the actors], even when the actual source of the sound is elsewhere’ (Woszczyk et al., 1995: 4).

In VR, the visual action can take place anywhere around the audience. The spatial arrangement and orientation are much more closely related to everyday life. As the cinematic space becomes *spatially realistic*, audiences are more likely to resort to perceptual mechanisms used in everyday life. As cinematic events can occur anywhere *there is no definitive focal point*. The narrative and the cinematic action may not be followed solely through visual indicators, contrary to the fixed-screen scenario. This filmmaking challenge, unique to the VR medium, can be addressed through the audiovisual integration mechanism. This rather obvious realisation has profound implications for VR filmmaking, as the soundtrack can be *used in ways not necessary or applicable in screen-based cinema*. This does not necessarily mean that the cinematic experience should or would become *more realistic*, but rather that new conventions, better aligned with the characteristics of the new medium, should be established.

The role of immersion and agency in VR and screen-based filmmaking

In related literature, *immersion* is quoted as one of the three pillars of VR, along with *presence* and *interactivity* (Cipresso et al., 2018; Mütterlein, 2018).

Agrawal et al. (2020) define immersion as:

‘a phenomenon experienced by an individual when they are in a state of deep mental involvement in which their cognitive processes (with or without sensory stimulation) cause a shift in their attentional state such that one may experience disassociation from the awareness of the physical world’.

Nilsson et al. (2016) distinguish between four views of immersion in VR:

‘(a) immersion as a property of the system used to present the virtual world; (b) immersion as a perceptual response to that system; (c) immersion as a response to an unfolding narrative, the characters inhabiting the story world, or the depiction of the world itself; and (d) immersion as a response to challenges demanding the use of one’s intellect or sensorimotor skills’.

Immersion is associated, and sometimes used interchangeably, with *presence*. Nilsson et al. (2016) also define presence as ‘the sensation of being in a given virtual (or unmediated) environment’ and Williams et al. (2022) as ‘the sense of being in one location, whilst truly situated in another’.

From an auditory perspective, immersion and presence are also frequently linked to *envelopment*. Rumsey (2002) describes envelopment as the feeling of being surrounded by direct sound sources or the reverberant sound-field around the audience. Agrawal et al. (2020) distinguish between envelopment as a perceptual process and immersion as a cognitive one. While a systematic review of these concepts is beyond the scope of this study, the reader can refer to the extensive body of relevant literature, such as the work of Kiouisis (2002), Freude et al. (2020), Steuer (1992), and Liu (2023). In the current context, *narrative immersion* is used for associations with the cinematic context and grammar, and *system immersion* with the audiovisual system characteristics, such as display size and format, spectatorship parameters, spatial sound, and interactivity.

System immersion has been key in filmmaking since its inception. It has been the driving force behind technological innovation for both film production and delivery. Notable examples include the introduction of wider screens (MacGowan, 1957), colour and sound, multichannel and object-based spatial audio (Dolby Laboratories Inc, 2011, 2015), stereoscopic projection, and more recently VR films and immersive cinema (Owatch, 2024).

Virtual reality (VR) was actively developed in the 1950s, in the context of training and military simulation initiatives (Virtual Reality Society, 2017). Nevertheless, early attempts at creating immersive experiences can be traced back to 1897 with *Cineorama*: spectators in a basket of a hot-air balloon surrounded by projection screens, taken on a virtual journey through filmed environments (MacGowan, 1957; Figure 5).

In 1956 cinematographer Morton Heilig also developed *Sensorama* (Basso, 2017; IEEE Communications Society, 2024; Figure 6), a booth with a 3D screen and multi-sensorial stimuli (vibration, smell, wind effect) and soon after Sutherland (1965) presented his future vision of indistinguishable-from-real VR experiences providing full system immersion. The development of VR head-mounted displays and head and hand tracking in the 60s and 70s led to *VR arcades* and the adoption of the technology by video game companies in the 1990s (Edwards, 2024; Wiltz, 2019). Since 2010, technology companies like Oculus, Samsung and HTC started releasing VR headsets on a commercial scale (Virtual Reality Society, 2017).

The emergence of these technologies led to the concept of *VR filmmaking*, the creative practice at the convergence of VR and cinema. Mateer (2017) provides a definition of *cinematic VR*:

‘a type of immersive Virtual Reality experience where individual users can look around synthetic worlds in 360°, often with stereoscopic views, and hear spatialised audio specifically designed to reinforce the veracity of the virtual environment’.

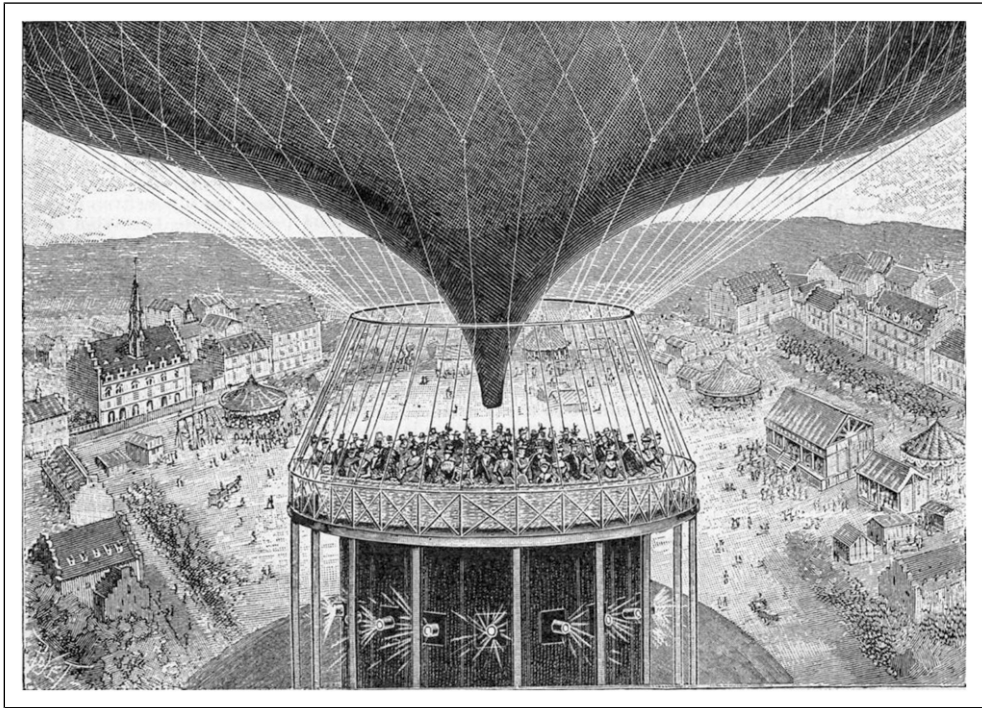


Figure 5. Illustration of Cineorama at the 1900 Paris Exposition ([Scientific American Supplement, 1900](#)).

It must be noted that in the context of this study, the term VR filmmaking encompasses both 360-degrees live-action and animated, graphics-based films, as long as the focus is on *narrative* and *storytelling* rather than the *entertainment value of the interactive experience*. Nevertheless, interactivity – a fundamental element of the VR medium – is still a key element of VR filmmaking, as it influences narrative immersion and, consequently, the overall experience and the underlying creative considerations ([Jicol et al., 2021](#)).

Viewer agency describes the ‘sense of being in voluntary control of actions within a VR environment’ ([Freude et al., 2020](#)) and can be conveyed through interaction with objects or characters, navigating the environment, or influencing the experience. In VR, viewer agency is intrinsically associated to the concept of *degrees of freedom* (DoF) of movement. VR environments accommodate either 3DoF, allowing the viewer to *rotate* and look around but not move in space, or 6DoF where the user can also *move* within a designated area ([Figure 7](#)). As exemplified by VR film [HELP \(2016\)](#), placing the viewer in the middle of the action amplifies viewer agency. The viewer determines which characters and actions to focus on and becomes the ‘live’ editor of the story, choosing the characters and the subjects to ‘cut to’.

Viewer agency is also conveyed through interaction with the virtual environment. This can range from passive viewing only allowing for 3DoF movement to extensive interactions with story elements and characters. Such interactions can also influence the story, as in the case of immersive VR films using *branching narratives* ([Reyes, 2017](#)). While more passive interaction approaches may be used in a VR film compared to other VR experiences ([Gödde et al., 2018](#)), the VR filmmaker must still strike a balance between following the narrative and maximising immersion and engagement through viewer agency.

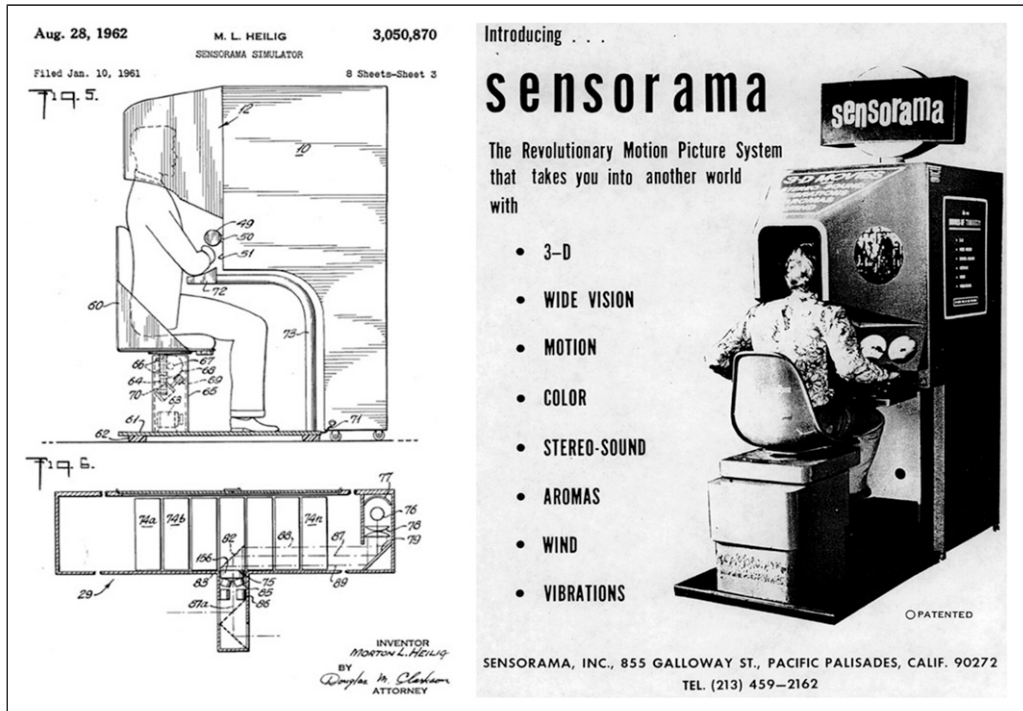


Figure 6. Sensorama, the first reported VR machine (IEEE Communications Society, 2024).

As discussed, cross-modal (audiovisual) integration is routinely used in cinema to create or reinforce a sense of immersion and envelopment. The goal is to create an immersive and enveloping soundscape *without drawing attention away from the screen* (Beauchamp, 2005: 145-149; Candusso, 2019; The National Academy of Recording Arts and Sciences Inc, 2004: 4.1-4.6). The underlying mechanisms are related to both *narrative* and *system* immersion (Nilsson et al., 2016).

In the case of narrative immersion, this typically takes the form of a soundtrack that elicits emotional responses and evokes a closer connection between the audience and the story. Such mechanisms are based on the *conceptual* aspects of the story rather than the physical orientation of the medium. Thus, they are transferrable between screen-based and VR cinematic forms with minimal or no changes to the associated production workflows or deliverable formats. For instance, the stereo reproduction of a powerful, emotional piece of music or the sound of the main character's heartbeat can be equally effective through the speaker arrays of a cinema or a set of headphones.

In the case of system immersion, cross-modal integration is used for placing the audience deeper into the narrative space in a *physical* sense. In screen-based cinema, associated auditory effects are mainly related to the expansion of the cinematic space beyond the boundaries of the screen through *ambiances*, *background sounds*, and *reverberation effects* (Beauchamp, 2005: 145-149; ITU-R BS.775-4, 2022). The main goal is the reinforcement of the *suspension of disbelief* (Oxford University Press, 2024) through the creation of *semi-realistic* spatial audio environments; a standard practice when working with both the standard discrete surround systems of the past (Dolby Laboratories Inc, 1997) and the more recent object-based systems like Dolby Atmos (Dolby Laboratories Inc, 2015).

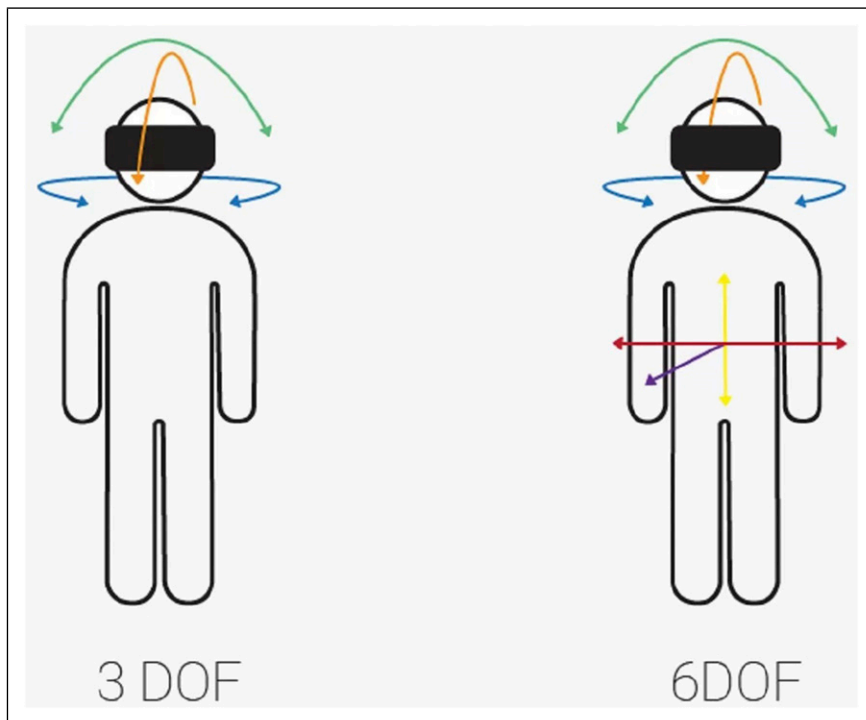


Figure 7. 3DoF and 6DoF in VR (Owatch, 2021).

From this perspective, the screen-based and VR cinematic spaces are significantly different. In the former, while the auditory environment may exceed the field of vision, it is always aligned with a *fixed visual space*. The cinematic illusion may be more convincing as a helicopter flies across the theatre, but it is unlikely that viewers will look away from the screen to visually identify it. In VR, spectatorship is an *intrinsically interactive* process, even in cases where viewer agency is limited. Unless the visual action is explicitly limited to a specific area of the VR space or the viewers' ability to move is constrained, there is no fixed visual attention point. Misaligned or unbalanced spatial audio cues may encourage viewers to turn their attention to the point of the *perceived sound source* (Bregman, 1994; Chion, 1994).

These unique characteristics of the VR film space present various aesthetic and technical challenges that were absent or irrelevant in screen-based cinema. From a sound design perspective, this has direct implications on key artistic and practical aspects like the spatial orientation of the soundtrack and the ways in which sound elements respond to the viewers' actions or movement.

Uses of low-end frequencies in screen-based and VR films

In addition to sound spatialisation, a key aspect of the cinematic soundtrack has to do with its *spectral content*, and particularly the *low frequencies*. The reproduction of low frequency content has been a focal point for both technology companies and sound designers since the introduction of *multichannel sound systems* (Dolby Laboratories Inc, 1997; Hope, 2009). This is unsurprising, as the emotional impact of the deep, resonant sounds of powerful subwoofers on audiences is

profound. By strategically manipulating low frequencies – most frequently 20Hz-120 Hz, but also below 20 Hz in the form of a tactile, vibratory effects – filmmakers can *heighten tension, emphasise dramatic moments, and provide a more realistic representation of large-scale events or characters.*

Low frequency sounds are commonly associated with *size* and *magnitude*. This stems from everyday life experience, where sounds emitted by large-scale objects or events generally contain more low frequency content. The examples of using this in cinema are numerous, ranging from the alien mothership floating above the theatre to a larger-than-life musical performance.

The capability to manipulate and reproduce low frequency content may also allow for a more realistic sense of *distance* and *space*. Without getting into a detailed discussion of the underlying perceptual mechanisms, one can refer to Moore's statement regarding *perceived distance calculations* based on auditory spectral cues:

'Over moderate distances, the spectrum of a complex sound may be changed, owing to the absorbing properties of the air; high frequencies are attenuated more than low' (Moore, 2003: 265).

Such connotations are deeply ingrained in the cinematic language and widely used for sound design purposes. In VR filmmaking, as the cinematic space becomes more expansive, immersive, and enveloping, a soundtrack that fully exploits the power of low frequency content *may be of heightened importance.*

In addition to low frequencies within the human hearing range, *infrasound* or *near-infrasound* frequencies – frequencies below 20 Hz, the lower end of the human hearing spectrum (Merriam-Webster Inc, 2024) – are also being used in the cinematic context. This can induce profound perceptual and emotional reactions without being directly detectable through audition:

'For a person with healthy hearing, the auditory sensation of the frequency range below approximately 20Hz alters to a feeling of pressure, changing the experience from one of listening with ears to one of sensing with other parts of the body' (Hope, 2009).

Hope (2009) also provides two notable historical examples of the use of infrasound effects: 'an almost inaudible 27 Hz drone tone' to magnify the sense of dread and confusion in Gaspar Noé's *Irréversible* (2002) and the use of 4 Hz to 50 Hz signals (Collins, 2024) to 'bring the audience into the experience' of earthquakes or explosions in *Earthquake* (1974) and *Rollercoaster* (1977).

Cinematic sound systems and sound design workflows have come a long way since these films were released, but the idea of using infrasound for supporting and enhancing the cinematic action persisted. In an interview with Blair Jackson (2010), Richard King, supervising sound editor in Christopher Nolan's *Inception* (2010), recalls how the sound crew used a system of oscillators and subwoofers to record:

'[transitions] from 10 or 12 Hz up into the audible hearing range [to create a] wave of sound that comes from nowhere that becomes quite massive as the sound comes up into 18, 20, 25Hz range' (Jackson, 2010).

Similar approaches are also used in immersive cinema shows (Owatch, 2024), where infrasound vibrations are used to enhance physical immersion. Irrespectively of whether the medium is powerful subwoofers or vibrating seats, it is important to note the interest of technology companies and cinematic content creators in this type of effect.

Admittedly, the heavy use of low frequency content is traditionally associated to specific cinematic genres like action, horror, or science fiction, where the experience is largely based on exaggerated emotional states, an element of surprise and mystery, and intense dramatic transitions. Nevertheless, the power of a full-range soundtrack experienced in an appropriately equipped theatre is undeniable, even if the genre or style of the production does not directly call for such obvious and explicit uses of sound.

The creative manipulation of low frequency content may be of *an even higher importance in VR films than in screen-based cinema*. This is due to the more expansive, immersive, and enveloping nature of the medium. This rather obvious and intuitive idea poses some important questions regarding the future direction and commercial viability of the sound systems used for VR films.

Presently, VR audio is commonly reproduced through the built-in speakers of the VR headset (Meta, 2024). Due to the size and positioning of the drivers, this delivery method limits the ability to reproduce low frequencies accurately. Nevertheless, most VR headsets offer some form of external headphone compatibility. Listening through headphones can offer improvements in terms of low frequency reproduction, but this largely depends on the characteristics and design of the headphones. However, even when high-grade headphones with accurate low-end extension are used, there are limitations on how well the intended low frequency effects will translate compared to a system using sizable subwoofers or dedicated vibration devices. This is especially true for infrasound effects where the aim is to create physical vibrations, strong enough to induce full-body sensations. It can be argued that watching movies through headphones or on mobile phones is not preventing audiences from enjoying their favourite films. This delivery mode, though, is usually an *alternative* to the full-scale theatre experience, which is arguably the closest interpretation of the filmmakers' and sound designers' vision. From a sound design perspective, this is an intriguing observation, as it introduces a new set of considerations relating to the production and reproduction of the VR film soundtrack.

Designing sound for VR films vs video games: common ground and differences

Many concepts and workflows related to sound design and audio production for VR have been thoroughly explored in the context of video games. Middleware like *FMOD* (Firelight Technologies and Logic, 2024) and *Wwise* (Audiokinetic Inc, 2024) provide detailed control of *interactive* and *procedural soundtracks* without the need for coding or engaging with the complexities of game development. Contemporary game development platforms like *Unreal Engine* (Epic Games, 2024) and *Unity Engine* (Unity Technologies, 2024) offer powerful audio manipulation tools to those willing to familiarise themselves with the rather complex production environments, and advanced audio encoding and deployment options ranging from discrete surround to Dolby Atmos and head-tracked binaural audio rendering through headphones (Dolby Laboratories Inc, 1997, Dolby Laboratories Inc, 2015; Epic Games, 2024). Specialist audio roles like *game audio designers* and *game audio programmers* have emerged, bridging the contextual and technical gap between sound designers and game developers.

From a technical standpoint, VR filmmakers and sound designers can rely on such tools and the consultancy of game audio experts. Nevertheless, a key idea behind VR filmmaking is that the experience is largely driven by the *narrative* and dictated by the *intentions and decisions of the filmmaker*, as opposed to its interactivity features and entertainment value. Conceptually, this entails that the creation of the VR film soundtrack is bound to involve *processes and workflows deeply rooted in traditional cinematic production models*, something not always applicable, or even desirable, in game development. This places VR film sound designers at the crossroads of two

distinct approaches: a *linear* and a *non-linear* one originated in film and video game audio production, respectively.

When creating sound for film, the contemporary approach is for audio editors or sound designers to edit or add additional sounds to a *linear timeline*, using a *digital audio workstation* (DAW; Figure 8).

Non-linear, procedural sound design workflows have developed primarily around game audio, where the lack of a fixed timeline requires sound playback to be controlled on the fly as the game is running (Figure 9). Nevertheless, whilst games are a distinct medium, parallels exist between games and traditional cinematic forms that have led to shared techniques and production methods. Some thoughts on both the common ground and notable differences between the two forms of production in terms of sound design are briefly discussed below.

In line with traditional media production, when designing sound for *VR video* or other cinematic VR formats, sound designers are commonly required to add additional elements to the soundtrack, such as Foley or sound effects (Turner et al., 2021). A common workflow for integrating these additional sounds is to record them as mono sound assets and then position them within the spatial audio scene. For VR productions that use recorded video or other existing footage as the primary source material, this mixing process currently requires manual placement of the sound source within the 3D scene to the corresponding azimuthal location and height (Riddershom Bargum et al., 2022). Influencing the perceived spatial proximity of sounds can be further affected by adjustment of reverberation and frequency content through reverb and equalisation effects.

When it comes to *3D game engines*, the purpose is to create a three-dimensional virtual environment or scene that can depict either a real-world or fictional space. This environment is populated with *game objects*, and these can be manipulated *through code* to create behaviours and generate the mechanics that the gameplay is built upon. From an audio production perspective, this object-based paradigm also allows for a convenient method to model three-dimensional sound

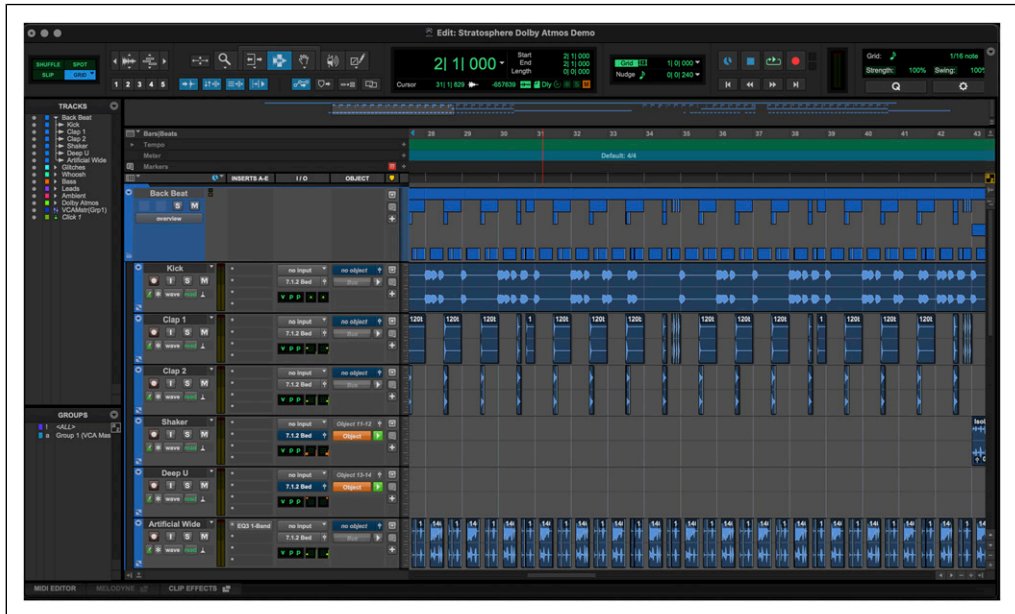


Figure 8. Linear DAW (Pro Tools) editing environment (Avid Technology, 2024).

interactions within virtual environments, as the spatial relationship between virtual sound sources and audio listeners is analogous to a real-world experience. This is notably different to mixing spatial audio *linearly* through discrete or object-based spatial audio systems (Dolby Laboratories Inc, 1997, 2015), as sound is processed *dynamically* based on the *positioning* and *actions* of the user.

Game sound sources typically comprise of pre-recorded or designed audio assets, and these are positioned within the virtual environment as *audio emitters*. *Audio listeners* function as virtual microphones within the environment, sampling the soundscape at a particular point (Goodwin, 2019). Sounds are shaped by specifying equalisation curves that can be automated over distance and orientation, and sound levels and reverberation mix can be controlled in the same manner. In practice, a combination of *amplitude reduction*, *increased filtering*, and *increased reverberation* can simulate sound source proximity to the listener (Sinclair, 2020; Stevens and Raybould, 2015). From this functional base of sound source and audio listener positional relationships, a three-dimensional interactive sound mix can be created in real-time, in sharp contrast to the manual approach required during linear VR video production (Turner et al., 2021).

In addition to *gameplay mode*, where the sound mix is created dynamically based on the player's actions and positioning as described above, games also employ *non-interactive sequences* or *in-game cutscenes* where player freedom of movement is reduced. These often function to deliver narrative elements to the player and feature traits that are natively cinematic, such as camera movements, voice acting or pre-composed score. In early games these were completely non-interactive, pre-rendered sequences, but as game technology developed so did the potential for them to be rendered within the game engine directly. This can be done by recording audiovisual content that keeps the player's perspective with less impact on gameplay, sound continuity, and player agency, or by manipulating this perspective by taking control with camera movements that:

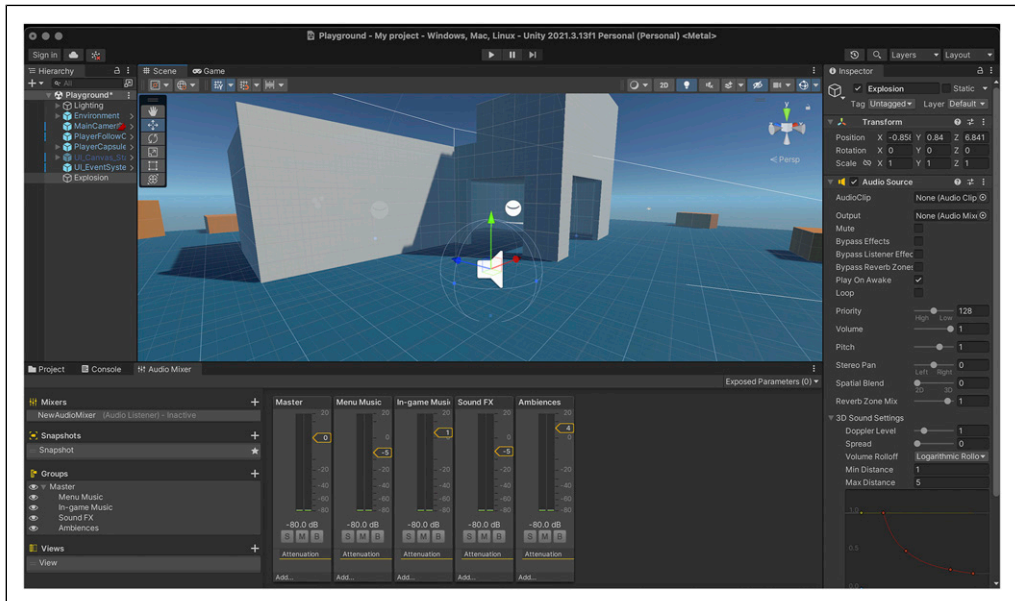


Figure 9. Non-linear game engine environment (Unity) with audio emitter and native audio mixer (Unity Technologies, 2024).

‘Place a precedence upon accentuating the scene and creating a filmic style with the possibility of reducing gameplay-cohesion and immersion’ (Garner, 2013: 131).

Alongside the development of spatialisation approaches to VR sound, there is an additional focus on creating more plausible spatial acoustics for virtual spaces within game environments, seen as a key contributor to immersion (Potter et al., 2022). Historically, the approach to this problem has been to apply reverberation to virtual spaces utilising a film sound workflow: using *auxiliary sends* to route audio to a set of pre-defined reverberation plugins. Current industry approaches include the *Spatial Audio Toolkit* within the Wwise audio middleware (Audiokinetic Inc, 2024) and *Project Audio* (Microsoft, 2024) that compute geometry-informed reverberation and implement a range of sound propagation effects. Whilst these new tools are designed for use within a game engine, plugins like *Indoor* (Audio Ease BV, 2024) also introduce graphical, game-like interfaces for linear media sound post-production. Such plugins model acoustic locations using impulse responses, allowing the user to specify source and listener spatial relationships *graphically* within the selected location. Moving either source or listener around the location adjusts the sound processing, but the signal flow is hidden from the user behind the plugin interface.

An important point stemming from these observations is that the workflows and tools used in game audio are notably different than those utilised by sound designers working in screen-based films. While it is worth noting that hybrid workflows and tools for mixing for VR in a DAW environment have already emerged (Dear Reality, 2024), they are still largely built around the game engine, thus, posing similar challenges to those used to more conventional workflows. If VR filmmaking is to be approached as a unique branch of filmmaking, it will be still rooted in the *narrative* and in *cinematic principles and aesthetics* that are not necessarily as important or relevant in production workflows rooted in game audio and game engines. This poses a unique challenge: sound designers who understand the filmmaking process may be called to use game audio workflows and tools that are not part of their current skillsets.

Methodology

Design

This study employed a qualitative research design based on secondary research and a series of targeted, semi-structured interviews (Knott et al., 2022). The secondary research part consisted of the background overview presented above, which formed the basis for identifying the following areas of interest:

- Guiding visual attention within the VR viewing space.
- Immersion and envelopment through sound spatialisation.
- The use of low-end frequencies in VR films.
- Sound design for VR films versus video games.

These areas of interest were used as guides for the generation of the following set of questions for a series of interviews with field experts:

- The VR space allows the filmmaker to expand the visual action beyond the field of vision of the audience, unlike screen-based cinema. Do you think the importance of the soundtrack in guiding visual attention may be heightened in VR films compared to traditional cinema? Can you provide some ideas about how this can be implemented?

- In the context of VR filmmaking, would you adopt traditional cinematic sound design techniques or take a different approach? Why?
- How do you approach the soundtrack in VR filmmaking to align with the spatial characteristics of the visual medium?
- Do you commonly use low-end frequency content and effects to portray assumed physical size, contextual gravity and importance, or emotional tone? What is the most important use of low-end content in the soundtrack?
- How important is the reproduction of low frequency audio in the context of the cinematic experience?
- The ability of most VR headsets and headphones to reproduce low-end frequencies is rather limited compared to speaker-based cinematic sound systems. How important do you think this may be in the development of VR filmmaking?
- What are your views on the possibility of viewing VR films in designated spaces where the visuals come from the headset but the sound through hybrid headphone/speaker systems? For instance, a ‘dark room’ type auditorium for VR viewing with full-range audio reproduction capabilities but no screen?

The interview questionnaire also included the following abstract, open-ended questions:

- What are the main challenges sound designers may face when working on a VR film?
- What future directions do you anticipate for sound design in VR filmmaking?
- What unique opportunities does VR filmmaking offer for sound design that were not present in screen-based cinema?
- What are the main challenges for sound design in the context of VR filmmaking?

These questions were aimed at addressing the initial assumptions of the authors that led to the identified areas of interest, while also providing further insights and context for analysis that may have been overlooked.

Participants

The interviewees were selected using convenience sampling, through the authors’ extended networks and connections. The main selection criteria were:

- A background in one or more of the identified areas of interest, and
- Some variation in terms of subject specialisation and areas of expertise.

For instance, instead of choosing sound designers who specialise in similar types of work, the authors chose two with no prior experience in VR production working in different formats (i.e. advertisements and short form films and long form and feature films, respectively) and two actively working in VR projects. The team also chose an immersive media producer with experience in VR production to capture a perspective different than that of a sound designer.

Procedure

The participants were contacted prior to the interviews and were given an overview of the study and consented to participating in the interviews. The interviews were carried out online and separately

from each other. At the start of each interview, interviewees were informed that the online meetings will be recorded, and a transcript will be automatically generated using the built-in functions of the online meeting software. They were also made aware that the recordings and transcripts would be used for analysis and dissemination. Interviewees consented to the intended use of their names, recordings, and transcripts for the purposes of the study. Interviews lasted between 40 and 60 minutes.

Data analysis

The interview data was used for thematic analysis (Willig and Rogers, 2017: pp.17-37). Both recordings and transcripts were used for familiarisation, while the transcripts were also used for coding and theme development. The authors identified and extracted the parts of the interviews that contained information relevant to the study, with an emphasis on the four areas of interest identified through secondary research. This resulted in the extraction of the following themes:

- VR film space navigation through visual and auditory cues.
- Perceptual effects of low-end frequencies and their implications on VR film sound reproduction.
- Assisting the VR film sound designer through hybridisation of workflows and tools.

Discussions and interview excerpts related to these themes are provided below.

Discussion

VR film space navigation through visual and auditory cues

Reflecting on his practical experience, immersive filmmaker Jack Ostergaard-Churchill (2024a) stated that visual cues, and specifically direction of movement and lighting, are commonly used methods of guiding the viewers' attention. Ostergaard-Churchill described the example of sequentially lighting parts of the environment in the production of Cortadellas's VR film *Symphony* (2021):

'You are in black. The conductor comes up in light in front of you and then does a piece of camera and then walks off to the side, kind of 90 degrees and then gestures at an orchestra, which is behind you, which then comes up in light. So, the movement of that means you follow him to 90 degrees. And then he points and you follow the point to the orchestra. And now you are 180 degrees behind. [...] You are moving through it, bringing more lights on as you kind of discover more things [and] as you look around and presume that the audience looks round to follow it' (Ostergaard-Churchill, 2024b).

In this case, the use of localised lighting is *motivated by and integrated into the narrative and location setting*.

The use of *extra-diegetic visual cues* requires a more discerning and genre-based approach. In games, the use of such cues to highlight active areas and guide interaction is commonplace and, often, necessary for playability. Arguably, in a highly interactive VR film, these techniques are acceptable and even expected by the viewer, whereas in a more linear VR film the risk of compromising the immersion and presence is higher. The narrative and aesthetic impact of extra-diegetic visual cues needs to be assessed on a case-by-case basis. For example, in a mystery or detective

story, where the character searches for clues, the use of overlays to highlight interactive items that convey information can contribute to the engagement with the story and character.

Ostergaard-Churchill proposes an intermediate approach of guiding the visual attention around the enveloping *mise-en-scene*, while respecting the expectations of the audience of a linear VR film:

‘With a headset and probably mobile [phone] as well [...] you have a general kind of sense of where front is, where you started. I like to pull people back to centre. If you send someone off for a little bit, then the next few shots kind of pull people back into the middle and then you send them off another way and pull them back in. Kind of having one region as their home or original viewing direction’ (Ostergaard-Churchill, 2024b).

This aligns with comments related to a general change of attitude of creators and audiences in terms of the use of the expanded cinematic space:

‘Early on in 360 people would put stuff all around because suddenly their frame is a lot wider. As audiences have progressed and as film makers have progressed that has narrowed down so your point of focus is usually a kind of main point within the 360 space. [...] People aren’t feeling pressured anymore to utilise the whole space for a narrative point. You can just have one area that is the focus and then everything else is supplementary; is kind of a background to that’ (Ostergaard-Churchill, 2024b).

Constraining and clearly defining the visual action space provides a viable solution, as it resembles a more traditional cinematic or theatrical environment with a well-defined focal area. Nevertheless, this is only part of the answer, as decisions about the focal points within the 360-degree space are largely dictated by the *storyline*, *genre*, and *setting* of the piece:

‘If you have a landscape with one thing happening, you know that there is nothing else there because it is off in the distance and there [are] no points of interest. But say you were in a shopping centre or something and there were lots of things happening. You could start to get that sense of, OK, maybe actually the action is going to jump out behind me or something like that. And maybe in a horror genre as well’ (Ostergaard-Churchill, 2024b).

In such cases, the lack of a definitive focal point is likely to encourage audiences to scan for auditory cues in the surrounding space more actively. This is an area where sound cues can take a pivotal role. Liberated from the concern of distracting the audience away from the screen, sound designers can be bolder and more creative with their sound spatialisation decisions, guiding viewers’ attention aurally through the visually complex scene. Nick Olsouzidis (2024a), a senior sound designer and composer interviewed during this study, aligned with this view, alluding to cases where the attention is intentionally drawn away from a previously established focal point to increase the dramatic effect:

‘VR ideas are there to be explored. If it is screen-based cinema, you are kind of confined into the screen space. You know, the exit door effect was something to be avoided inside the cinema because it takes you out of the action. In non-screen-based experiences, if the story demands that zombies are chasing you, it would be very interesting to get the [sound] out of the back, to reproduce the sound of zombies chasing you and make you turn your head and see them through the VR headset’ (Olsouzidis, 2024b).

In addition to the stereotypical use of transient sound effects for the occasional dramatic moment, the liberation from the constraints of the screen may encourage sound designers and filmmakers to experiment with attention-guiding through sound in a broader context. While perceptual effects related to visual dominance and the spatial magnetisation of sound by image may be still relevant when occurring within the visual field, they may take different forms and meanings when the audiovisual disparity expands beyond it. *Musical cues*, *persistent sounds within the surrounding ambiances*, or *spatialised dialogue* and *speaking voice* can function as key visual attention guides instead of merely confirming and following the action in the constraint space of the screen.

Perceptual effects of low-end frequencies and their implications on VR film sound reproduction

The association between low frequency content and size and magnitude is typical and widely used in cinema. In Nick Olsouzidis' words:

'If you do the sound design for Godzilla, you certainly want [to portray] physical size as it is a big character, a big monster' (Olsouzidis, 2024b)

Interviewee James Hynes, sound designer at [The Sound Directors \(2024\)](#), also refers to this well-known perceptual effect of sound:

'In the soundtrack in general, low-end is something that gives the audience a perception of volume. [For instance, in the case of] music in a feature film. Again, it is helpful if it is full, if [it contains] all the frequencies in the spectrum including the low frequencies to give it gravitas and impact' (Hynes, 2024).

In addition, low frequency sounds are also used to signify key moments, conceptual importance, or emotional charge:

'[Low-end frequencies can be used for signifying] adrenaline during a fight scene. There will be low-end on the weapons, whether it is cannons, explosions, or a single focused shot from a gun, or on body impacts. [You will also use low-end content] for contextual gravity and importance; in key moments where the director will cut to an object of significance or to a close-up of a character of significance. Emotional tones are exactly the same, whether it is a montage or the entire sequence. There will be extensive use of low frequency content in all of these scenarios' (Olsouzidis, 2024b).

This may have direct connections to the enhancement of *immersion*, *engagement*, and *envelopment* – key aspects of the VR experience – and to the suspension of disbelief, which is equally or potentially more important in VR compared to screen-based cinema.

For sound sources that are easily identifiable (e.g. explosions, animals, aeroplanes), the proportion of low frequency content compared to mid/high frequencies encourages the perceptual placement of the source further away from or closer to the listener. James Hynes also refers to a less known use of this effect:

'Low-end is something that I use all the time for proximity, especially for dialogues. [...] You know, [to portray] if something is clear or distant to you' (Hynes, 2024),

While he also brings up the case of creating a general sense of space by manipulating the low frequency content:

‘If the scene is in a small room, you will have a different bass response compared to a warehouse with a large hum throbbing away; you will put lots of bass in there. The audience, without really having to watch the pictures, will immediately understand that they are in a very large space’ (Hynes, 2024).

In relation to the technical capabilities of the VR sound production and reproduction systems used, Hynes provides a pragmatic and insightful perspective:

‘Knowing that you have got all the frequencies to play with is a lovely thing. If all of a sudden you are told, OK, actually they are going to be listening to it on headphones or earplugs that do not carry all the sounds, that are possibly data compressed or sent via Bluetooth or something, you’ve got to work with that. And that’s probably a different workflow. You would master differently knowing that it is going to be on small speakers with not all the frequencies present and no air to actually activate the sound’ (Hynes, 2024).

This is worth noting and discussing further in the context of VR filmmaking, as it may have *creative*, *practical*, and *commercial* implications as the VR film medium evolves. At this stage, one can only speculate about its future as a distinct, viable cinematic form. Nevertheless, it is relatively easy to envisage scenarios where the VR film soundtrack is delivered through appropriate, specialised systems that provide extensive low frequency reproduction and infrasound effects. This could take the form of *specialised headphones* and *tactile devices*, *hybrid delivery* using combinations of *open-back headphones*, *close-range speakers* and *dedicated low frequency reproduction hardware*, or *public viewings in appropriately designed spaces* equipped with *dedicated audio systems*. This aligns with Nick Olsouzidis’ projection:

‘We will definitely see better sound for VR. [...] It will be an amalgamation, or more of an extension of what we already use. You will walk into a room where you’ll wear headsets and [you will have the sound coming from] an array of speakers around you, whether it’s Dolby Atmos, a 5.1 system, or any other formats to reinforce that experience. And there is stuff like this being currently developed and released’ (Olsouzidis, 2024b).

Developments like these have the potential to create entirely new markets. As a public exhibition, one can envisage *VR film viewing venues*, reminiscent of the cinema theatre. As a private experience, it is not too far-fetched to imagine *bespoke versions of home theatre systems*, fine-tuned for VR film viewing. With today’s exponential rates of technological advancement, one can start making even bolder assumptions like James Hynes’ description of an ideal future system:

‘The biggest thing that gets in the way of mass adoption [for VR] apart from price is going to be the size of the headsets and the sound. So, I was sort of speculating what would be my ultimate system. I’d come up with something where you don’t need to put anything on, and I know that sounds completely crazy but you’re not putting on goggles, you are just turning it on and it is just happening. The second one is not only are you getting discrete sound for yourself, that is 100% discrete, not just leaky headphones, and you are also feeling the sound as well as hearing; the sound goes back to the infrasound’ (Hynes, 2024)

Irrespectively of the format and level of sophistication of the systems we will be using to deliver the VR film soundtracks of the future, low frequency reproduction capabilities will be as important as – or, quite likely, more important than – they are for screen-based cinema.

Assisting the VR film sound designer through hybridisation of workflows and tools

Creating a dynamic soundtrack based on realistic spatial representations of objects and events around the viewer may be appropriate for video games and other types of interactive VR experiences, as the participant is given time and space to explore the environment. In filmmaking, however, this can be problematic if it does not align with the filmmaker's intentions, as it may attract attention away from the visual action or key narrative points. From a sound design perspective, considerations related to *volume levels*, *spatialisation*, or *reverberation* of specific elements of the soundtrack may become more challenging and nuanced in VR films compared to both screen-based cinema and video games. This can be exacerbated when the soundtrack is complex, involving numerous sound or music cues that can be dynamically placed anywhere around a freely moving audience. James Hynes highlights how this can lead to new and distinct sound design roles:

'In the [VR filmmaking sound design] production process, you are going to need to be able to actually separate all the layers of every single object. At the moment, there is a quite laborious way of doing it. I think there is probably a role for something like a 3D sound engineer, a 3D mixer in there. It will be a specific task, not just a dubbing or a re-recording mixer or a sound engineer. It will be a person who specializes in creating and curating the environment. [Someone like] Walter Murch or Gary Rydstrom who came up with something very different rather than just mixing sound, they *designed sound*' (Hynes, 2024).

Hynes also suggests that this could be reflected in the development of more advanced tools and workflows, specifically targeting the unique spatial and interactive characteristics of the VR film soundtrack:

'That is probably where AI comes in and helps you and say, OK, the stream is over there, the birds are over there, the cars are passing over there and it will work it out for you [...], but probably on a more sophisticated level. As you turn or as a car goes past it will also prompt you to turn and look at the car. These things sort of happen at the moment, but they are a little bit raw in the way they happen I think' (Hynes, 2024).

To some extent, such roles and workflows already exist in the game audio space. Interviewee Paul Weir (2024) described how game sound designers use a range of techniques to sequence and spatialise audio for *in-engine cutscenes*, from fully pre-rendered stems to individual sound elements triggered along a timeline. In the case of the latter, these can be mixed through *the same audio engine* that is *already in use for regular gameplay*, preserving mix levels, equalisation, effects, spatialisation, and sound continuity.

The mastering process and the generation of the final deliverables is also different when working with immersive media sound. For example, online streaming services provide a platform for hosting and playback of VR video, but their capabilities are limited. Tom Slater (2024) from Call & Response Studios Ltd (2024) points at the potential benefit of working natively in a game engine instead of relying on the rather limiting encoding and mastering tools provided by streaming platforms:

‘Generally I find streaming platforms [...] very limiting when it comes to spatial audio and 360 video. [With some of the big platforms] you are limited to 1st and 2nd order [Ambisonics¹] and no options to decode to anything other than stereo. [...] I find working in game engines offers far more decoding options’ (Slater, 2024).

Collectively, the approaches and tools highlighted here suggest the development trajectory of further procedural tools for game sound, moving towards more convincing spatialisation and environmental modelling approaches. Linear media workflows for VR video remain complicated and often require manual spatialisation (Turner et al., 2021), suggesting both automatic panning and distance emulation as key potential areas where the toolset could be improved. In contrast to most linear media tools, once these are setup, many parameters are automated based on positional and environmental data within the virtual environment of the game. Considering this, and the well-developed approach to designing sound for game cinematics in-engine, there exists potential for a convergence of these production approaches that both simplifies sound authoring process for linear VR and extends the possibilities for spatial modelling. The potential missing link for this process is either real-world spatial data from VR productions or the virtual equivalent in virtual productions and 3D-rendered animated work. Parallel developments in spatial computing may also provide the viable link between these two workflows, leading to the development of hybrid DAW and game engine editing environments that could provide more solid structures and support for sound designers working in VR film projects.

Conclusion

The main goal of this study was to identify some of the creative and practical opportunities and challenges filmmakers and sound designers may encounter when working in VR projects with a cinematic narrative structure and orientation. Due to the complexity and breadth of the associated subjects, the authors explored selected areas rather than attempting a thorough review of the underlying concepts. The following areas of interest were identified through secondary research:

- Guiding visual attention within the VR viewing space.
- Immersion and envelopment through sound spatialisation.
- The use of low-end frequencies in VR films.
- Sound design for VR films versus video games.

The identified areas informed a series of interviews with selected practitioners. Through thematic analysis of the interview data the following themes emerged:

- VR film space navigation through visual and auditory cues.
- Perceptual effects of low-end frequencies and their implications on VR film sound reproduction.
- Assisting the VR film sound designer through hybridisation of workflows and tools.

These themes were used for further discussion and reflection, highlighting areas of potential interest for further work.

In terms of the use of the soundtrack for guiding visual attention in VR films, while perceptual effects and mechanisms used in screen-based cinema may be still relevant when occurring within the visual field, they may take different forms and meanings in a VR environment where the audiovisual disparity expands beyond the field of view. Musical cues, persistent sounds within the

surrounding ambiances, or spatialised dialogue and speaking voice can function as key visual attention guides instead of merely confirming and following the action in the constraint space of the screen.

Low frequency content is commonly used in cinema to create a sense of size and magnitude, and signify key moments, conceptual importance, or emotional charge. This may have direct connections to the enhancement of immersion, engagement, and envelopment in VR – all key characteristics of the medium. The low frequency reproduction capabilities of the VR systems we use to produce and deliver the VR film content will be as important as – or, quite likely, more important than – they are for screen-based cinema.

The hybridisation of the linear sound design workflows of screen-based cinema and the non-linear ones stemming from game audio, although promising, poses challenges to those used to more conventional workflows. Approaches and tools highlighted here suggest the development trajectory of further procedural tools for game sound, moving towards more convincing spatialisation and environmental modelling approaches and more user-friendly environments for those coming from a linear, screen-based sound design background.

Collectively, these observations point to an intriguing and timely realisation. While many of the long-established principles, workflows, and approaches will still apply to VR filmmaking, the very nature of the medium may call for the development of a new cinematic language, distinct from that used in screen-based cinema. This study aimed at assisting sound designers, VR filmmakers, and immersive media practitioners in better contextualising and defining the potential roles of the soundtrack in VR filmmaking as the new language is being developed.

Acknowledgements

The authors would like to thank the following creative practitioners who kindly accepted the invitation to be interviewed for the purposes of this study: James Hynes, Nick Olsouzidis, Jack Østergaard-Churchill, Tom Slater, Paul Weir. This study would not be possible without your invaluable contributions.

ORCID iDs

Christos Manolas  <https://orcid.org/0000-0003-0057-8374>

Panos Raptis  <https://orcid.org/0009-0002-8867-2954>

Note

1. Ambisonics is a spatial audio system offering highly accurate 3D audio capturing and reproduction. The system is based on *spherical harmonic decomposition* and its accuracy and spatialisation precision is measured based on the level of analysis or *order* used. Audio spatialisation resolution increases in line with the order of the system, ranging between 1st and 7th order occupying 4 to 64 audio channels respectively (Bertet et al., 2013; Blue Ripple Sound Limited, 2024).

References

- Abrams N, Bell I and Udris J (2001) *Studying Film*. Arnold.
- Agrawal S, Simon A, Bech S, et al. (2020) Defining immersion: literature review and implications for research on audiovisual experiences. *Journal of the Audio Engineering Society* 68(6): 404–417.
- Audio Ease BV (2024) Indoor. Available at: <https://www.audioease.com/indoor/> (Accessed 11 June 2024).
- Audiokinetic Inc (2024) Wwise, the most advanced, feature-rich interactive audio solution. Available at: <https://www.audiokinetic.com/en/wwise/overview> (Accessed 28 May 2024).

- Avid Technology (2024) Pro Tools. Available at: <https://www.avid.com/pro-tools> (Accessed 14 July 2024).
- Basso A (2017) Advantages, Critics and Paradoxes of Virtual Reality Applied to Digital Systems of Architectural Prefiguration, the Phenomenon of Virtual Migration. *Proceedings of the International and Interdisciplinary Conference IMMAGINI* 1(9). DOI: [10.3390/proceedings1090915](https://doi.org/10.3390/proceedings1090915), In press.
- Bazin A (1967) *What Is Cinema? Volume I*. Berkeley: University of California Press.
- Beauchamp R (2005) *Designing Sound for Animation*. Taylor & Francis.
- Bertet S, Daniel J, Parizet E, et al. (2013) Investigation on localisation accuracy for first and higher order Ambisonics reproduced sound sources. *Acta Acustica United with Acustica* 99(4): 642–657.
- Blue Ripple Sound Limited (2024) Technical Notes. Available at: https://www.blueripplesound.com/technical_notes (Accessed 30 October 2024).
- Bolognini N, Frassinetti F, Serino A, et al. (2005) Acoustical vision of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160: 273–282.
- Bregman SA (1994) *Auditory Scene Analysis: The Perceptual Organization of Sound*. The MIT Press Massachusetts Institute of Technology.
- Call and Response Studio Ltd (2024) Call & Response Studios. Available at: <https://callandresponse.org.uk/> (Accessed 27 October 2024).
- Candusso D (2019) Creating and designing the contemporary soundtrack: a case study. In: Batty C, Berry M, Dooley K, et al. (eds) *The Palgrave Handbook of Screen Production*. Switzerland: Palgrave Macmillan, 221–230.
- Chion M (1994) *Audio-Vision: Sound on Screen*. New York: Columbia University Press.
- Cipresso P, Giglioli IAC, Raya MA, et al. (2018) The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in Psychology* 9: 2086.
- Collins (2024) SENSURROUND Definition and Meaning. Available at: <https://www.collinsdictionary.com/dictionary/english/sensurround> (Accessed 24 March 2024).
- Colman AM (2009) *A dictionary of psychology A Dictionary of Psychology*. Oxford University Press. Available at: <https://www.oxfordreference.com/display/10.1093/acref/9780199534067.001.0001/acref-9780199534067> (Accessed 25 October 2024).
- Dear Reality (2024) dearVR SPATIAL CONNECT. Available at: <https://www.dear-reality.com/products/dearvr-spatial-connect> (Accessed 30 October 2024).
- Dolby Laboratories Inc (1997) S97/10133/11496: Surround sound past, present, and future: a history of multichannel audio from mag stripe to Dolby digital. Available at: https://projectionniste.net/docs/Surround_Sound_-_Past_Present_and_Future.pdf (Accessed 20 March 2024).
- Dolby Laboratories Inc (2005) Dolby Surround Mixing Manual. Available at: <https://archive.org/details/manualzz-id-840859> (Accessed 05 May 2024).
- Dolby Laboratories Inc (2011) S11/22805/24230: Dolby Surround 7.1. Available at: <https://professional.dolby.com/siteassets/pdfs/dolby-surround-7-1-for-theater-tech-paper.pdf> (Accessed 20 March 2024).
- Dolby Laboratories Inc (2015) S15/27842/28833: Dolby Atmos Specifications. Available at: <https://professional.dolby.com/siteassets/cinema-products—documents/dolby-atmos-specifications.pdf> (Accessed 20 March 2024).
- Earthquake (1974) Directed by Mark Robson. *Mark Robson, the Filmmakers Group*. Universal Pictures.
- Epic Games (2024) Unreal Engine. Available at: <https://www.unrealengine.com/en-US> (Accessed 28 May 2024).
- Edwards B (2024) Virtual Boy: the bizarre rise and quick fall of Nintendo's enigmatic red console. *Ars Technica*, Available at: <https://arstechnica.com/gaming/2024/05/virtual-boy-the-bizarre-rise-and-quick-fall-of-nintendos-enigmatic-red-console/> (Accessed 26 October 2024).
- Firelight Technologies (2024) Sound. Logic Made for games – FMOD is the solution for adaptive audio. Available at: <https://www.fmod.com/> (Accessed 28 May 2024).

- Freude H, Ressing C, Müller M, et al. (2020) Agency and body ownership in immersive virtual reality environments: a laboratory study. In *Hawaii International Conference on System Sciences*, University of Hawaii at Manoa. Vol. 2020.
- Garner TA (2013) *Game Sound from behind the Sofa: An Exploration into the Fear Potential of Sound & Psychophysiological Approaches to Audio-Centric, Adaptive Gameplay*. PhD Thesis. Denmark: Aalborg University.
- Gödde M, Gabler F, Siegmund D, et al. (2018) Cinematic narration in VR – rethinking film conventions for 360 degrees. In: Chen JYC and Fragomeni G (eds) *Virtual, Augmented and Mixed Reality: Applications in Health, Cultural Heritage, and Industry. Lecture Notes in Computer Science*. Cham: Springer International Publishing, 184–201.
- Goodwin SN (2019) *Beep to Boom: The Development of Advanced Runtime Sound Systems for Games and Extended Reality*. New York: Routledge.
- HELP (2016) *Directed by Justin Lin. 360 google spotlight stories*. 360 Google Spotlight Stories.
- Holt NJ (1997) *Perceptual Lateralisation of Audio-Visual Stimuli*. University of York.
- Hope C (2009) The bottom end of cinema: low frequency effects in soundtrack composition. *Sound Scripts* 2(1).
- Hynes J (2024) *Interviewed by the Authors*. London: Ravensbourne University.
- IEEE Communications Society (2024) The first VR device. *Sensorama, Created by Morton Heilig*, Available at: <https://www.comsoc.org/node/19151> (Accessed 13 July 2024).
- Inception (2010) *Directed by Christopher Nolan*. Warner Bros Pictures.
- Irréversible (2002) *Directed by Gaspar Noé. StudioCanal / Universal*. Mars Distribution.
- ITU-R BS.775-4 (2022) Multichannel stereophonic sound system with and without accompanying picture.
- Jackson B (2010) Christopher Nolan's inception. In: *Mix Magazine*. Available at: <https://www.mixonline.com/sfp/christopher-nolans-inception-369267> (Accessed 25 March 2024).
- Jicol C, Wan CH, Doling B, et al. (2021) Effects of emotion and agency on presence in virtual reality. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, Yokohama Japan, 6 May 2021, pp.1–13.
- Kerins M (2011) *Beyond Dolby Stereo: Cinema in the Digital Sound Age*. Indiana University Press.
- Kiousis S (2002) Interactivity: a concept explication. *New Media & Society* 4(3): 355–383.
- Knott E, Rao AH, Summers K, et al. (2022) Interviews in the social sciences. *Nature Reviews Methods Primers* 2(1): 1–15.
- Lippert M, Logothetis N and Kayser C (2007) Improvement of visual contrast detection by a simultaneous sound. *Brain Research* 1173: 102–109.
- Liu Y (2023) Analysis of interaction methods in VR virtual reality. *Highlights in Science, Engineering and Technology* 39: 395–407.
- MacGowan K (1957) The wide screen of yesterday and tomorrow. *Quarterly of Film Radio and Television* 11(3): 217–241.
- Mast G (1984) On framing. *Critical Inquiry* 11(1): 82–109.
- Mateer J (2017) Directing for Cinematic Virtual Reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice* 18(1): 14–25.
- Meares DJ (1993) Perceptual attributes of multichannel sound. In: *Proceedings of the AES 12th International Conference: Perception of Reproduced Sound*, Copenhagen, Denmark, 28 June 1993–30.
- Merriam-Webster Inc (2024) Medical definition of INFRASOUND. Available at: <https://www.merriam-webster.com/medical/infrasound> (Accessed 24 March 2024).
- Meta (2024) Meta Quest 3. Available at: <https://www.meta.com/gb/quest/quest-3/> (Accessed 27 October 2024).
- Microsoft (2024) Project Audio. Available at: <https://github.com/microsoft/ProjectAcoustics> (Accessed 11 June 2024).

- Moore CJB (2003) *An Introduction to the Psychology of Hearing*. 5th edition. Emerald Group Publishing.
- Münsterberg H (1916) *The Photoplay: A Psychological Study*. New York and London: Appleton.
- Mütterlein J (2018) The three pillars of virtual reality? Investigating the roles of immersion, presence, and interactivity. In: *Proceedings of the 51st Hawaii International Conference on System Sciences*, Hawaii, USA.
- Nilsson NC, Nordahl R and Serafin S (2016) Immersion Revisited: a review of existing definitions of immersion and their relation to different theories of presence. *Human Technology* 12(2): 108–134.
- Olsouzidis N (2024a) Sound Design. Available at: <https://nickolsouzidis.com/> (Accessed 27 May 2024).
- Olsouzidis N (2024b) *Interviewed by the Authors*. London: Ravensbourne University.
- Ostergaard-Churchill JA (2024a) Jack Aaron Ostergaard-Churchill. Available at: <https://jackaaronostergaardchurchill.com/> (Accessed 27 May 2024).
- Ostergaard-Churchill JA (2024b) *Interviewed by the Authors*. London: Ravensbourne University.
- Owatch (2021) 3-Dof vs 6-dof for VR headset selection. Available at: <https://www.stekiamusement.com/2021/3-dof-vs-6-dof-for-vr-headset-selection-21902/> (Accessed 12 July 2024).
- Owatch (2024) What is the difference between 4D, 5D, 6D, 7D, 8D, 9D Cinema Theater. Available at: <https://www.stekiamusement.com/2018/whats-the-difference-between-4d-5d-6d-7d-8d-9d-cinema-theater-3465/> (Accessed 24 May 2024).
- Oxford University Press (2024) Oxford reference. Available at: <https://www.oxfordreference.com/display/10.1093/oi/authority.20110803100544310> (Accessed 27 May 2024).
- Ozenua Oluwatobi J (2022) Designing for VR in Figma Medium. Available at: <https://medium.com/pintsizedrobotninja/comfortable-viewing-distances-in-xr-834aaec10df9> (Accessed 14 July 2024).
- Palma Stade T, Schofield G and Moore G (2023) Narrative perspectives and embodiment in cinematic virtual reality. In: De Paolis LT, Arpaia P, Sacco M, et al. (eds) *Lecture Notes in Computer Science*. Cham: Springer Nature Switzerland, 232–252.
- Potter T, Cvetković Z and De Sena E (2022) On the relative importance of visual and spatial audio rendering on VR immersion. *Frontiers in Signal Processing* 2: 904866.
- Reyes MC (2017) Screenwriting framework for an interactive virtual reality film. In: Immersive Learning Research Network. Coimbra, Portugal, June 26–29, 2017.
- Riddershom Bargum A, Ingi Kristjánsson O, Babó P, et al. (2022) Spatial audio mixing in virtual reality. In: Geronazzo M and Serafin S (eds) *Sonic Interactions in Virtual Environments, Human–Computer Interaction Series*. Cham, Switzerland: Springer, 269–302.
- Rollercoaster (1977) *Directed by James Goldstone*. California, USA: Universal Pictures.
- Rothe S, Buschek D and Hußmann H (2019) Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction* 3(1): 19.
- Rumsey F (2002) Spatial quality evaluation for reproduced sound: terminology, meaning, and a scene-based paradigm. *Journal of the Audio Engineering Society* 50(9): 651–666.
- Rumsey F (2016) Objects, mixing, and rendering. *Journal of the Audio Engineering Society* 64(7/8): 584–588.
- Scientific American Supplement (1900) *Illustration of the Cineorama Balloon Simulation at the 1900 Paris Exposition*. Scientific American Supplement No.1287.
- Shams L and Kim R (2010) Crossmodal influences on visual perception. *Physics of Life Reviews* 7(3), pp.269–284.
- Sinclair JL (2020) *Principles of Game Audio and Sound Design: Sound Design and Audio Implementation for Interactive and Immersive Media*. New York: Routledge.
- Slater T (2024) *Interviewed by the Authors*. London: Ravensbourne University.
- Steuer J (1992) Defining virtual reality: dimensions determining telepresence. *Journal of Communication* 42(4): 73–93.

- Stevens R and Raybould D (2015) *Game Audio Implementation: A Practical Guide Using the Unreal Engine*. New York: Routledge.
- Störmer VS (2019) Orienting spatial attention to sounds enhances visual processing. *Current Opinion in Psychology* 29: 193–198.
- Sutherland I (1965) The ultimate display. *Proc. of the IFIP Congress*: pp506–508.
- Symphony (2021) *Directed by Igor Cortadella*. CaixaForum.
- The National Academy of Recording Arts and Sciences Inc (2004) *Recommendations for Surround Sound Production*. Report The Recording Academy's Producers & Engineers Wing.
- The Sound Directors (2024). *Home*, Available at: <https://www.thesounddirectors.com/> (Accessed 27 May 2024).
- Turner D, Murphy D, Pike C, et al. (2021) Spatial audio production for immersive media experiences: perspectives on practice-led approaches to designing immersive audio content. *The Soundtrack* 13(1): 73–94.
- Unity Technologies (2024) Unity Engine. Available at: <https://unity.com/products/unity-engine> (Accessed 28 May 2024).
- Vanhoutte K, Joris E, Debackere B, et al. (2010) Cinematography of surround video, assuming a passive spectator. 2020 3D Media.
- Virtual Reality Society (2017) History of virtual reality Virtual Reality. Available at: <https://www.vrs.org.uk/virtual-reality/history.html> (Accessed 3 June 2024).
- Vroomen J and de Gelder B (2000) Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance* 26(6): 1583–1590.
- Wang H, O'Fearghail C, Zerman E, et al. (2021) Visual attention analysis and user guidance in cinematic VR film. In: 2021 International Conference on 3D Immersion (IC3D), Brussels, Belgium, 8 December 2021, pp. 1–8.
- Weir P (2024) *Interviewed by the Authors*. London: Ravensbourne University.
- Williams J, Shepstone S and Murphy D (2022) Understanding immersion in the context of films with spatial audio. In: AES 2022 International Audio for Virtual and Augmented Reality Conference, USA, August 2022.
- Willig C and Rogers WS (2017) *The Sage Handbook of Qualitative Research in Psychology*. Sage Publications Ltd.
- Wiltz C (2019) The story of Sega VR: Sega's failed virtual reality headset. *Designnews.com*, Available at: <https://www.designnews.com/testing-measurement/the-story-of-sega-vr-sega-s-failed-virtual-reality-headset> (Accessed 26 October 2024).
- Woszczyk W, Bech S and Hansen V (1995) *Interaction between Audio-Visual Factors in a Home-Theater System*. New York: Audio Engineering Society - 99th Convention.
- Zettl H (1990) *Sight Sound Motion: Applied Media Aesthetics*. 7th edition. Belmont, CA: Wadsworth Publishing Company.