The Routedge Handbook of Sound Design

Edited by Michael Filimowicz

First published 2025

ISBN: 978-1-032-35147-6 (hbk)
ISBN: 978-1-032-35157-5 (pbk)
ISBN: 978-1-003-32556-7 (ebk)

Chapter 1

Listener Perspective in Games and Virtual Reality

DOI: 10.4324/9781003325567-1

The funder for this chapter is Tampere University.

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

1

LISTENER PERSPECTIVE IN GAMES AND VIRTUAL REALITY

Isak de Villiers Bosman, Kristine Jørgensen, Oğuz 'Oz' Buruk, and Juho Hamari

1.1 Introduction

Common use of the term 'perspective' as it relates to perception is generally synonymous with the visual sense, e.g., the correct application of perspective to create the illusion of a 3-dimensional image in art or the term being synonymized with a 'point of view' (Cambridge English Dictionary 2023; Merriam-Webster 2023). However, since the resulting perspective is the result of how we perceive the environment from a particular vantage point in space, the concept itself applies equally to the sense of sound. Sound is generated through physical movement that causes vibrations which reflect off the surrounding environment and reach our ears. These are then perceived and interpreted by our ears and brain in a way which encodes the resulting perceptual information in many ways, including its spatial qualities. Semantically speaking, however, the common use of the term 'perspective' does not extend to such a perceptually holistic definition.

This ocularcentric view is pervasive throughout the discourse in multimedia, such as film and games (Collins 2013a, 21–22; Altman 1992, 39). The result of this is a lack of pervasive terminology with which to discuss the concept and implications of perspective with an auditory focus, e.g., a 'point-of-audition' rather than a 'point-of-view.' This is especially true in the domain of virtual reality (VR) where the discourse on perspective is already preemptively imbued with an assumption of the visual first-person perspective (1pp) whereby a user perceives all visual information from the point of view of a character/avatar within the virtual environment (Monteiro et al. 2018). The accompanying assumption is thus that if the user is viewing the virtual environment from the 'eyes' of the avatar, they must also be listening from its 'ears.' However, considering that the role of sound in media such as film is guided by conventions that evolve over time (Altman 1992, 42), we posit that an absolutist approach to imposing the roles of audio and visual feedback as fixed in VR (not to mention other senses such as touch, smell, etc.) limits the evolution of VR as an expressive medium. The goal of this chapter is thus to provide an entry point for discussing the auditory perspective in VR by exploring the relationship between related concepts in 'traditional' media studies, specifically film and games played on so-called 'non-immersive' platforms, such

DOI: 10.4324/9781003325567-1

as PC and console. As such, we will often use existing concepts in film studies as a point of departure. We explore this concept of perspective and 'point-of-audition' through the following concepts: first, we discuss the idea of a perceptual frame within which media experiences are located and how this relates to existing theories in film and game studies. We then examine these against the often-misused concept of realism and refer to theories of embodiment as a basis within which to understand these concepts. Finally, it should be noted at this point that while the role of agency afforded by the ability of games to make decisions also plays a role in some of the concepts discussed here, e.g., embodiment, our focus is primarily on the perceptual affordances of the media discussed and, as such, we do not discuss the concept of agency here in depth.

1.2 The perceptual 'frame'

The technological capabilities of a VR headset, or head-mounted display (HMD), have several implications that pertain to how a player of VR games experiences the virtual environment. This includes the fact that the visual display itself blocks out the outside world and provides a large field-of-view, as well as the accurate sense of depth and scale provided by stereoscopic 3D. Furthermore, the sensorimotor contingencies (SCs), i.e., intuitive actions for performing actions such as looking around, reaching out and grabbing with one's hands, etc. work to create the so-called 'place illusion' of convincingly finding oneself in a different place than where one is actually located. Put differently, VR utilizes a strategy of 'immediacy,' i.e., it aims to hide the perceptual medium and convince the user that they are directly perceiving the environments and objects in the virtual environment (Jørgensen 2013, 7–8). This illusion is often referred to as 'being there,' which contributes to the construct known as 'presence' (Slater 2009). In terms of visual feedback, these technological capabilities effectively remove the 'boundary' or 'finite framing' (McArthur et al. 2017) between the player and the virtual environment which, in traditional platforms, is always visible in the form of the edges and the inherent 2-dimensionality of the screen and which constructs a distinct perceptual frame in which all information is located.

Although the terminology of the perceptual frame is perhaps more suited to visual feedback, we can extend this concept of finite framing to the audio domain as well. This, however, requires us to reconsider what these perceptual boundaries are, since headphones, which predate loudspeakers, already envelop the listener's ears and have the ability to block out external stimuli through noise isolation or active noise cancellation. The argument has even been put forth that headphones can be considered the original HMD, since they allow listeners to perceive virtual worlds in a way that far predates the capabilities of visual displays (Garner 2017, 201). Audio displays such as headphones thus allow for the sound of the virtual environment to extend beyond the screen and create a convincing acoustic ecology (Grimshaw 2007). Once again in reference to screen-based media, the visual imagery of film and games is tied to the limits of the screen within the larger viewing environment, such as the cinema, living room, etc., while sound, in contrast, creates the illusion of inhabiting the same physical space as the listener (Collins 2013b). It should also be noted, however, that the contrast between the enveloping spatiality of audio and visual feedback is less pronounced in VR, since head-based tracking also affords the powerful illusion of the digital environment 'bleeding into' the user's physical reality.

The recent resurgence of VR technology has also sparked a resurgence of interest in spatial and immersive audio (Serafin et al. 2018). This chapter will not provide an in-depth

discussion of these terms (the reader is referred to chapters 15 and 17 of this book for topic-specific overviews), but the goal of these approaches can be briefly summarized as aiming to provide sound feedback that reproduces our perception of audio in everyday 3D space. By mimicking the underlying psychoacoustical processes that we use to discern spatial properties of sound, such as their location, the combination of different audio processing techniques allows us to localize properties of audio sources in 3D space by discerning their orientation and distance relative to ourselves (Liaquat et al. 2021). These techniques include introducing subtle differences in timing and volume for each ear to emulate a sound source reaching each ear at a slightly different time and volume, known as interaural time and level differences (ITDs and ILDs), and the spectral processing of sound sources to mimic a similar change as a result of sound reflecting off one's physical anatomy, known as head-related transfer functions (HRTFs). In other words, both of these techniques include adjusting audio based on an individual's anatomy, i.e., the distance between their ears and the shape of their torso, shoulders, head and ears.

This ability to localize an audio source in a virtual 3D space is in contrast to that of "inside head locatedness" (Rieger and Zingler 2023) or "lateralisation" (Garner 2017, 76) where the lack of spatial properties embedded into an audio signal leads to the listener perceiving the sound as if spatially originating from within their own head, e.g., monophonic audio. The spatial affordances of audio feedback and its omnidirectional quality thus allow for sound to envelop a listener and create not only the meanings encoded in the sounds themselves, but also the broader sense of a space wherein these sounds reside (Grimshaw and Schott 2007). It should be noted at this point that similar arguments can be made for the use of loudspeakers for VR audio, which has also received renewed attention as a method for creating 3D binaural spatial audio including the use of filtering mechanisms to provide unique audio streams to each ear (e.g., Kurabayashi et al. 2014). However, this chapter will focus on the use of headphones or near-field speakers, such as those built into the Meta Quest, since these approaches are more common in the form of commercially available VR headsets.

Drawing contrasts between headphones and what can now be considered a 'standard' HMD (such as a commercially available headset from Meta or HTC Vive), commercial off-the-shelf headphones lack positional tracking capabilities of the moment-to-moment rotation and movement of the user's head in 3D space. This limits the ability of headphones to allow for a convincing illusion of spatially located audio, since the 3D location of an audio source is fixed to the rotation of a user's head. In other words, even though a sound source might be processed using ITDs, ILDs, and HRTFs to give a sense of spatiality, the perceived location of the sound source is still relative to the orientation of the listener's headphones and ears and, as such, the 3D environment created by the sound source rotates along with the listener's head. Thus, a finite framing still exists in traditional games played with headphones, albeit one that is tied to the listener's own head rather than that of a screen.

Similar to the boundary of the edges and two-dimensionality of the screen being overcome by a VR display in the visual domain, this 'boundary' is also overcome in various degrees in the auditory domain, since the head-tracking provided by an HMD can be used to provide spatial audio based on the head movements of the wearer. In addition to standard HMDs using head-tracking to update the rotation and coordinates of the avatar in virtual 3D space, various audio-only implementations have been created which only provide audio feedback based on users' head rotation and/or movements. Examples of these

include adding six-degrees-of-freedom (6DOF) tracking solutions on top of headphones that can track position and orientation (Siu et al. 2020) or using standard webcams to track users' heads in 3D and update audio accordingly (Carvalho et al. 2021). Another notable example is that of Apple AirPods, which allow for head-tracked spatial audio (Apple 2020), which might signal a shift toward larger-scale adoption of spatial audio in commercial audio and affords entirely new possibilities for commercial interactive audio (Rieger and Zingler 2023), provided new games are able to make use of these capabilities. However, in addition to head-tracking, there are still significant barriers that need to be overcome for a high-fidelity spatial representation of audio in 3D space. Part of this is simply a result of the extreme complexity of recreating the physical processes of sound propagation and reflection in space, which is a very computationally expensive process (Liaquat et al. 2021). Furthermore, since each individual's anatomy affects their perception of sound, this creates a unique listening experience for each person (Jenny and Reuter 2021). While much work has been done on faster and more accurate approaches to creating personalized sound reproduction based on individual anatomies (e.g., Zhao et al. 2022), it is still an open problem in the design of sound hardware and software for VR. As such, the lack of positional tracking in audio hardware for traditional games and the general limitations in accurately recreating the spatial properties of sound both highlight the fact that the perceptual frame is still subject to an invisible finite framing in the audio domain. To illustrate the implications of this perceptual frame and its relationship with the concept of auditory perspective in immersive media, we will refer to three related theoretical concepts in film and games.

1.3 The film-, game- and surrogate body

When watching a film, the audience is generally assumed to be located outside the narrative world of the film, whereas their ocular window into the world of the film in the form of the camera/viewpoint might adopt a more ambiguous position from a static invisible bystander to a moving perspective that provides the illusion of viewing the world as if from a character's own eyes (Garner 2017, 116). This concept is described by the theory of a 'film subject' or 'film body' that inhabits the space and provides the audience with the imagery of the film. This film body is thus simultaneously located inside the world of the film and is also the window into this world (Crick 2011; Sobchack 1994). As such, the film body simultaneously observes the fictitious reality and conveys it to the viewer (Sobchack 1994).

This argument has also been extended to the realm of 3D games towards the conceptualization of a 'game body' that the player controls, in addition to the avatar, that takes the form of the game camera that creates both the point of perception as well as an implied 'corporeal' or bodily experience within the gameworld (Crick 2011). This is especially apparent in third-person perspective (3pp) games, since there is a separation between the camera and the avatar and the two can often be controlled separately, to an extent. The implication of the finite framing discussed above is that the player is always perceptually located 'outside' the game with a clear 'distance' between themselves and the gameworld. This is also described in the theory of the 'surrogate body' which refers to the viewer 'lending' their own bodily presence to the screen by experiencing their response to the media experience, among other ways, in terms of their bodily experiences thereof (Spiel and Gerling 2019). The theory highlights the physical distance between the player's body and the gaming hardware, such as the screen and inputs, and the effect of this on experiential

aspects, such as enjoyment. We can thus broadly distinguish between a conventional definition of perspective and perceptual framing in games in the following way: conventional perspective describes the distance between the avatar and the camera, whereas perceptual framing describes the distance between the camera (through the gaming hardware) and the player's physical body.

Once again, however, this corporeal conceptualization focuses on the camera, which serves as the ocular window into the film world via the screen, at the expense of its auditory equivalent. One might think of this 'point-of-audition' as the 'microphone' of this entity, which perceives a complex mixture of sound that is created from many different sources and which may or may not align with what is seen. Similar to the visual perspective created by the camera position/angle, the use of audio is also used to convey perspective. For example, director Darren Aronofsky has described his decision not to include any music for the film *Mother!* in the following way (Aronofsky 2017):

> I've always toyed with making a fully subjective movie … I think this is the first time that I've fully submitted to the rules of trying to make every single shot in it from Jennifer Lawrence's character as Mother's point of view … Music is actually an objective tool, it's a tool that a director uses to tell an audience how to feel … The music wasn't coming out of Jennifer's character, it was coming out of my mind, and we wanted to make the film purely from her point of view.

This distinction is commonly applied in terms of the 'diegesis' which refers to a hypothetical/fictional world presented through some narrative medium, such as the 'world' of a film, and situates the characters within this fictional/diegetic frame (Garner 2017, 98). All perceived information associated with this narrative would thus have a situational relationship with this diegetic frame which, traditionally, either places it inside or outside this frame. Information within the diegesis, such as dialogue, ambience and other sound which, narratively speaking, is 'perceptible' by diegetically situated characters, would be referred to as diegetic, while all information outside this frame that is not perceptible to characters, such as music and voice-overs, would be referred to as non-diegetic/extradiegetic (Redecker 2022).

Aronofsky's reasoning, however, emphasizes not only this diegetic aspect, but also the use of audio as a way to present a perspective in a cinematic context (Sobchack 1994); this is related to the concept of 'focalization' which conceptualizes all diegetic information as being conveyed through a character's perception and experiences (Jørgensen 2007). Furthermore, unlike the use of the camera which always presents a singular perspective at a single point in time, this example highlights the ability of the soundscape to combine the 'objective' and 'subjective' perspectives simultaneously, such as mixing the use of Foley and music. This also has implications for the concept of the game body, which states that the player 'passes through the screen' by occupying the diegetically located game body (Crick 2011). The focus here, however, is placed entirely on visual perception and neglects the fact that the game body also conveys non-diegetic audio cues to the player.

VR, however, aims to completely eradicate the perceptual distance between the player and the display technology in order to place the user 'inside' the virtual space. For example, in their discussion of the surrogate body in games, Spiel and Gerling emphasize the fact that VR platforms are able to remove the physical distance between the bodies of the player and that of the display and state that in VR games the "physical body of the player merges with

the surrogate body to the game" (Spiel and Gerling 2019). One might thus argue that VR, in this sense, does not generally allow for a perspective that is detached from the avatar in the virtual environment, since the predominant design approach is to also place the player in the body of the avatar performing game actions, i.e., a first-person perspective (Hoppe et al. 2022). While this approach more clearly aligns with the goal of providing congruence between the player's sense of self and the avatar through which they enact their performance in a game (Isbister 2016, 11–13), it also deprives designers and researchers of the opportunity to explore perceptual and cognitive edge-cases pertaining to the diminishing of the boundaries between the physical world and the gameworld.

At this point we can summarize some implications regarding VR audio and the perceptual frame. Firstly, the concepts of the perceptual frame, game body, and surrogate body, which emphasize the distance between one's own body and that of the in-game camera/microphone, provide an additional dimension in understanding a user's total sense of auditory perspective in VR. Combined with the conventional notion of perspective in games, which focuses on the distance between the in-game camera/microphone and the avatar, a more holistic approach might be to consider the total auditory perspective as the combination of these two 'distances.' Understanding the auditory perspective in VR thus requires us to consider both the finite framing which still exists in the audio domain and the effect of conventional perspective, which is an underexplored concept in VR. This is also impeded by a lack of terminology for discussing auditory perspective or an ocularcentric tendency in this regard, e.g., the 'in-game camera' and the 'point of view.'

Secondly, along with the physical distance that may be overcome through the surrogate body, Spiel and Gerling also highlight an 'aesthetic distancing' which results from the perceived distance between the audiovisual stimuli and the spectator's embodied experience. This view regards aesthetic distancing not merely as an obstacle to be overcome, but as a necessary ingredient for an enjoyable aesthetic experience which could be undermined by the physical proximity between the physical and virtual bodies. This adds another dimension to designers' consideration of the use of auditory perspective, since this might contribute to maintaining this aesthetic distance. Finally, the soundscape of VR has the ability to encompass both the aforementioned 'objective' and 'subjective' perspectives simultaneously, which has raised questions for the designers of VR audio, such as whether to directly apply existing non-diegetic approaches of music for film and games or whether designers need to rethink the role of music for VR games entirely (Phillips 2018). Understanding these implications requires an examination of how a VR user experiences their own embodied experience in VR, specifically as it pertains to audio perception, and how we might rethink the often-misunderstood concept of realism.

1.4 Realism and cine-realism

Virtual reality (VR) systems are often created with realism as the primary design goal, including perceptual realism and interaction fidelity, to the point where the terms are sometimes equated with one another (Rogers et al. 2022). This could be considered self-evident to an extent by drawing parallels between the terms in the sense that both strive toward the 'real' (real-ism and real-ity). As a result, some aspects of the design of VR applications default toward certain decisions, one of which is the choice of viewer perspective. VR games are predominantly created with the assumption of a first-person perspective in the sense that the player perceives the virtual environment from the perspective of a virtual

character and interacts with this environment through this character (Black 2017; Cui and Mousas 2023). Since this egocentric perspective mimics one's interaction with physical reality through one's own body, it is the most obvious choice to create a realistic way of interacting with a digital reality (Hoppe et al. 2022).

On the one hand, the concept of realism, which is nearly pervasive throughout the discourse on VR both in research and commerce (Rogers et al. 2022), is useful for reasons that include the emotional response from feeling as if portrayed events are really happening, e.g., for phobia therapy or training purposes (Slater et al. 2020). On the other hand, our understanding of what realism entails is limited and often incoherent, since unrealistic/impossible phenomena are often perceived as 'realistic' within different contexts, such as games and simulations (Bosman et al. 2021; Slater et al. 2022). A narrow focus on realism as a design goal also limits what can be achieved with technologies such as VR to create interesting and engaging experiences. Not only can realism itself be conceptualized along different typologies, but experiences within the general domains of art and entertainment often aim for a degree of unrealism and infidelity for the sake of their experiential goals (McArthur et al. 2017; Rogers et al. 2022). In fact, what consumers of media experiences, such as film and games, perceive as realistic portrayal of events often sharply deviates from any counterpart in physical reality and is rather based on media-specific conventions (Grimshaw 2007). The realism that audiences might expect is thus not based on a direct relationship with physical reality, but rather on a 'cinematic realism' or cine-realism that replaces one's reference point on which to base expectations (Collins 2008, 134).

There are several reasons for adopting strategic unrealism in different forms of media. One of these pertains to the so-called uncanny valley, where near-perfect representations of humans and human-like qualities are generally found to be much more unsettling than blatantly unrealistic representations (Mori 1970). While this is a well-known problem for visual areas such as character design (Schwind et al. 2018), its analog in the audio domain has received limited attention (e.g. Grimshaw 2009). Discussions of the uncanny valley in the audio domain tend to come in the form of a warning and/or an invitation: aiming for a perfect recreation of reality risks creating unpleasant reactions from listeners (Rumsey 2014), which could be utilized if this is the goal of the experience, e.g., horror games (Grimshaw 2009). Along this vein, it is interesting to note that a disproportionate amount of research in game audio focuses on horror games and, by extension, the creation of what are generally considered 'negative emotions' through strategic use of audio (Grimshaw 2009; Rogers and Weber 2019). Outside this relatively niche genre, however, the threat of the uncanny valley is often met with the advice to steer away from perfect representation toward stylization and abstraction since, as it is hypothesized, listeners tend to adopt a more critical stance when encountering source materials that have clear references toward real-world origins, e.g., a piano versus a synthesizer (Rumsey 2014). One might thus extend this logic toward the visual and aural perspectives created by VR games and explore the possibility of strategic unrealism through exploration of different viewer/listener perspectives. For example, adopting a third-person perspective in VR might disarm a user's critical expectations of experiencing all audiovisual stimuli in a way that strictly resembles their physical reality, thus creating more creative freedom for game and sound designers to explore different approaches. There is also untapped potential in exploring levels of incongruence between visual and audio stimuli, which might serve as a valuable tool to sustain interest over time (McArthur et al. 2017). A valuable concept in understanding what realism and cinematic realism entails in VR is that of embodiment. While this concept has

been implicitly discussed in the previous sections regarding the film/game/surrogate bodies, it is worth explicitly discussing this concept and how it relates to listening.

1.5 Embodied listening

The term embodiment is used to refer to the general conceptualization of body ownership, including the conscious/unconscious, affective and phenomenological framing thereof and the ontological, psychological, epistemological, legal and ethical issues this entails (de Vignemont 2011). Embodied cognition emphasizes the link between using one's body to interact with the world and the influence this has on how we are able to make sense of it (Stingel-Voigt 2020). One aspect of embodiment that is of particular relevance in the context of audio perception is that of self-location. While interacting with physical reality, we use the observable characteristics of objects to infer their spatial properties, such as viewing an object to determine its size and distance. However, inverse to this, we also use the physical characteristics of the environment to infer our own sense of self-location, i.e., where we are in 3D space. In this regard, self-location can be considered the inverse of localization, as it is the localization of oneself relative to the sounding environment as opposed to localizing elements in the environment relative to oneself. Under normal conditions, our sense of embodiment takes place from an egocentric frame of reference, i.e., we perceive reality from the reference point of our own bodies and perceptual organs. As such, the process of self-locating involves the perception of sensory input with a perspectival structure which, as discussed as part of spatial audio, includes our perception of sound in space (Millière et al. 2018).

Furthermore, referring back to ITDs, ILDs and HRTFs, the fact that the ability to hear spatially is tied to one's own anatomy makes the process of self-locating an individualistically embodied process, which is enabled by the aural technology provided by VR hardware and software. One can also extend the notions provided by Heidegger and Merleau-Ponty of embodied cognition whereby we are able to incorporate the tools we use into our bodily schema (Gallagher 2013). For example, by virtue of wearing headphones while using VR and using head rotation in the process of localizing external sounds and ourselves in a virtual 3D space, we are extending our bodily schema to use the headphones and the implied virtual ears of the avatar as tools to perceive this space (Murphy 2021, 183). Once again, however, the fidelity of this implied body's hearing organs are limited by the ability of spatial audio reproduction techniques to match their characteristics to that of the listener (Zhao et al. 2022) as well as the tracking fidelity of the hardware which can cause breakdowns in this perceived sense of body ownership (Murphy 2021, 184).

The concept of embodiment and egocentric perception provides a useful lens for discussing a wide variety of techniques in auditory storytelling, from a fictional sense of self-localization to the selective presentation of visible sources that would normally emit sound. Starting with the former, the creators of audio for media experiences generally design for the properties of a fictional space and attempt to recreate the properties of sound within this space from a specific point of audition, which is usually the same as the visual point of view (Collins 2013b). While the construction of detailed acoustic spaces is done for the sake of authenticity and credibility, the ability to control auditory perspective through embodied self-location also allows for a degree of identification with a character and their emotions, goals etc. Notably, beyond the representation of known physical spaces, auditory

perspectives can also present unknown perspectives, such as that of embodying non-human species (Collins 2017). On the other hand, a technologically simpler but no less powerful approach for emphasizing the auditory perspective is that of selective presentation of audio sources. For example, during an action sequence, a film might provide detailed sound for each event occurring on screen, but if the film transitions toward a more emotional scene in the same context, most of the aforementioned sounds might be blocked out to give focus to dialogue. The embodied listening experience in this case leans into the cine-real to convey a character's focus and internal state rather than merely the sum total of their 'raw' heard experience. Through this change in the use of focalization, the intention thus shifts from visceral realism toward emotional identification with characters, which can be a powerful technique to reveal the internal subjectivity of characters in a narrative (Collins 2013b). Adapting this mindset for VR by shifting away from a physical realism-driven mindset toward that of the cine-real, designers of VR experiences can use the flexibility of virtual and interactive experiences such as VR, which affords interesting possibilities for the creative use of auditory perspectives, towards ends such as reducing the psychological distance between a user and a character/avatar.

The implications of these concepts can once again be summarized at this point. Firstly, embodied listening as an activity situated within one's own anatomical markers toward self-location is a useful construct in understanding one of the tenets of a 'realistic' listening experience within virtual space. However, a narrow aim towards achieving 'true realism' in VR audio is fraught with misapprehension as audiences of media experiences in general tend to rely more on convention or cine-realism on which to base their expectations. Coming close to high-fidelity reproductions, but not close enough also introduces the risk of creating the so-called uncanny valley through sound design. An approach for intentionally deviating from this approach toward strategic unrealism would thus be to refer to sound design strategies from existing media, such as film and traditional games, similar to how games already borrow conventions from film (Jørgensen 2013, 6). By applying conventional approaches in designing Foley, music, dialogue, etc. in other media, VR users might apply the same willing suspension of disbelief that they apply in other forms of media.

Secondly, however, the reliance on a cinematic realism does not imply a fixed cine-reality on which audiences base their expectations, since conventions change over time for different media (Collins 2017). Since VR as a medium is still relatively novel in terms of mainstream adoption, it is crucial to not only investigate how different aspects of audio perception function in VR, but also to challenge the notion of universally correct/appropriate design decisions. Purposefully deviating from narrow conceptualizations of VR as a platform could provide a way to diversify the range of experiences and create more opportunities for exploring its experiential capabilities. For example, the narrow focus on first-person perspective makes the concept of perspective itself an underexplored concept in VR games (Monteiro et al. 2018), especially as it pertains to audio feedback. Lastly, while the focus of this chapter has been on the perceptual qualities of VR technology, the role of decision-making and agency plays a significant role on the creation and perception in interactive systems. However, we might already speculate about ways in which perspective and agency might be combined, such as VR gameplay that provides game mechanics which allow the player to alter their auditory perspective toward accomplishing their goals.

1.6 Conclusion

The point of departure for many discussions on audio is that of the 'traditional' medium which is being deviated from. Game audio is often compared to film audio and VR, in turn, to the screen and keyboard/controller of desktops and consoles. We rely on an established vocabulary of concepts from which to draw parallels and contrasts, for the benefit and, some would argue, detriment of communication and ideation. This benefit, however, does not only extend toward the emergence of new technologies and techniques, but also provides a new lens with which to look back upon the foundation onto which is being built. In their examination of musical production technologies, Ragnhild Brøvig-Hanssen and Anne Danielsen state: "Just as the new is always heard in light of the old, the old will always be revisited in light of the new" (2016, 68). Examining emerging technologies against more established foundations, such as narratology, film theory and embodiment, has allowed us to point toward ways in which listening in VR needs to be reconceptualized and provided directions for doing so in reference to perspective, realism and embodied listening. Insights from theory and practice in this chapter point towards considerations and future directions for researchers and designers in this fast-moving discipline. Our perceptions of the 'natural' and 'believable' are also fluid and contingent on convention from various areas of practice (Schafer 1969, 3; Grimshaw 2007). We do not consider the topics discussed in this chapter to be an exception, and therefore we argue it would be a mistake to regard them as static and unyielding in the ever-changing landscape of sound design for games and virtual environments.

References

Altman, Rick. 1992. *Sound Theory, Sound Practice. AFI Film Readers*. New York, London: Routledge.

Apple. 2020. 'AirPods Pro (2nd Generation)'. Apple. 2020. www.apple.com/airpods-pro/.

Aronofsky, Darren. 2017. *Director Darren Aronofsky & Sound Supervisor Craig Henighan – Conversations with Sound Artists – Season 3*. https://soundworkscollection.com/post/director-darren-aronofsky-sound-supervisor-craig-henighan-conversations-with-sound-artists-season-3.

Black, Daniel. 2017. 'Why Can I See My Avatar? Embodied Visual Engagement in the Third-Person Video Game'. *Games and Culture* 12 (2): 179–99. https://doi.org/10.1177/1555412015589175.

Bosman, Isak de Villiers, Koos de Beer, and Theo J. D. Bothma. 2021. 'Creating Pseudo-Tactile Feedback in Virtual Reality Using Shared Crossmodal Properties of Audio and Tactile Feedback'. *South African Computer Journal* 33 (1). https://doi.org/10.18489/sacj.v33i1.883.

Brøvig-Hanssen, Ragnhild, and Anne Danielsen. 2016. *Digital Signatures: The Impact of Digitization on Popular Music Sound*. Cambridge: MIT Press.

Carvalho, Davi Rocha, William D'Andrea Fonseca, Jacob Hollebon, Paulo Henrique Mareze, and Filippo Fazi. 2021. 'Head Tracker Using Webcam for Auralization'. In Proceedings of the Inter-Noise 2021, 1–5. Congress, WA. https://eprints.soton.ac.uk/450040/.

Collins, Karen. 2008. *Game Sound: An Introduction to the History, Theory, and Practice of Video Game Music and Sound Design*. Cambridge: MIT Press.

———. 2013a. *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. Cambridge: MIT Press.

———. 2013b. 'Sonic Subjectivity and Auditory Perspective in Ratatouille'. *Animation* 8 (3): 283–99. https://doi.org/10.1177/1746847713507164.

———. 2017. 'Calls of the Wild? "Fake" Sound Effects and Cinematic Realism in BBC David Attenborough Nature Documentaries'. *Soundtrack*, 10 (1): 59–77. https://doi.org/10.1386/ts.10.1.59_1.

Crick, Timothy. 2011. 'The Game Body: Toward a Phenomenology of Contemporary Video Gaming'. *Games and Culture* 6 (3): 259–69. https://doi.org/10.1177/1555412010364980.

Cui, Dixuan, and Christos Mousas. 2023. 'Evaluating the Sense of Embodiment through Out-of-Body Experience and Tactile Feedback'. In *Proceedings of the 18th ACM SIGGRAPH International*

Conference on Virtual-Reality Continuum and Its Applications in Industry, 1–7. VRCAI '22. New York: Association for Computing Machinery. https://doi.org/10.1145/3574131.3574456.

Gallagher, Shaun. 2013. 'Phenomenology'. In *The Encyclopedia of Human-Computer Interaction*, 2nd ed. www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/phenomenology.

Garner, Tom A. 2017. *Echoes of Other Worlds: Sound in Virtual Reality: Past, Present and Future*. Palgrave Studies in Sound. Cham: Springer.

Grimshaw, Mark Nicholas. 2007. 'Sound and Immersion in the First-Person Shooter'. *The 11th International Computer Games Conference*, November. https://vbn.aau.dk/en/publications/sound-and-immersion-in-the-first-person-shooter.

———. 2009. 'The Audio Uncanny Valley: Sound, Fear and the Horror Game'. In *Audio Mostly 2009*, 21–26, Glasgow. https://vbn.aau.dk/files/61573698/audioUncannyValley_MG.pdf.

Grimshaw, Mark Nicholas, and Gareth Schott. 2007. 'Situating Gaming as a Sonic Experience: The Acoustic Ecology of First Person Shooters'. In *Situated Play*. Tokyo: Digital Games Research Association (DiGRA). www.digra.org/.

Hoppe, Matthias, Andrea Baumann, Patrick Chofor Tamunjoh, Tonja-Katrin Machulla, Paweł W. Woźniak, Albrecht Schmidt, and Robin Welsch. 2022. 'There Is No First- or Third-Person View in Virtual Reality: Understanding the Perspective Continuum'. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–13. CHI '22*. New York: Association for Computing Machinery. https://doi.org/10.1145/3491102.3517447.

Isbister, Katherine. 2016. *How Games Move Us: Emotion by Design*. Cambridge: MIT Press.

Jenny, Claudia, and Christoph Reuter. 2021. 'Can I Trust My Ears in VR? Head-Related Transfer Functions and Valuation Methods with Descriptive Attributes in Virtual Reality'. *International Journal of Virtual Reality* 21 (2): 29–43. https://doi.org/10.20870/IJVR.2021.21.2.4831.

Jørgensen, Kristine. 2007. 'On Transdiegetic Sounds in Computer Games'. *Northern Lights: Film & Media Studies Yearbook* 5 (1): 105–17. https://doi.org/10.1386/nl.5.1.105_1.

———. 2013. *Gameworld Interfaces*. London: MIT Press.

Kurabayashi, Hiroaki, Makoto Otani, Kazunori Itoh, Masami Hashimoto, and Mizue Kayama. 2014. 'Sound Image Localization Using Dynamic Transaural Reproduction with Non-Contact Head Tracking'. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences E97-A* (9): 1849–58.

Liaquat, Muhammad Usman, Hafiz Suliman Munawar, Amna Rahman, Zakria Qadir, Abbas Z. Kouzani, and M. A. Parvez Mahmud. 2021. 'Localization of Sound Sources: A Systematic Review'. *Energies* 14 (13): 3910. https://doi.org/10.3390/en14133910.

McArthur, Angela, Rebecca Stewart, and Mark Sandler. 2017. 'Sounds Too True to Be Good: Diegetic Infidelity – The Case for Sound in Virtual Reality'. *Journal of Media Practice* 18 (1): 26–40. https://doi.org/10.1080/14682753.2017.1305840.

Millière, Raphaël, Robin L. Carhart-Harris, Leor Roseman, Fynn-Mathis Trautwein, and Aviva Berkovich-Ohana. 2018. 'Psychedelics, Meditation, and Self-Consciousness'. *Frontiers in Psychology* 9. www.frontiersin.org/articles/10.3389/fpsyg.2018.01475.

Monteiro, Diego, Hai-Ning Liang, Andrew Abel, Nilufar Bahaei, and Rita de Cassia Monteiro. 2018. 'Evaluating Engagement of Virtual Reality Games Based on First and Third Person Perspective Using EEG and Subjective Metrics'. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 53–60. https://doi.org/10.1109/AIVR.2018.00015.

Mori, Masahiro. 1970. 'The Uncanny Valley'. *Energy* 7 (4): 33–35.

Murphy, Dooley. 2021. '*Through the Screen Door: Virtual Reality Experiences*'. PhD thesis, Faculty of Humanities, Department of Communication: University of Copenhagen. https://curis.ku.dk/portal/da/publications/through-the-screen-door(8ce7cbfc-fe3c-454b-be4b-5ae1f137beaf).html.

Phillips, Winifred. 2018. '*Composing Video Game Music for Virtual Reality: The Role of Music in VR*'. Blog post April 15. https://winifredphillips.wpcomstaging.com/2018/04/15/composing-video-game-music-for-virtual-reality-the-role-of-music-in-vr/.

Redecker, Björn. 2022. 'Game Music as a Boundary Object: Applying a Thought Pattern to the Transgressive Audio of Playdead's *Inside* (2016).' *Journal of Sound and Music in Games* 3 (1): 36–49. https://doi.org/10.1525/jsmg.2022.3.1.36.

Rieger, Martin, and Dominik Zingler. 2023. '3D Audio: The Next Step in Realism for Gaming: Benefits and Pitfalls of 3D Audio in Games from a Creative Content and Technical

Implementation Point of View'. *Games: Research and Practice* 1 (1): 12: 1–12:3. https://doi.org/10.1145/3583059.

Rogers, Katja, Sukran Karaosmanoglu, Maximilian Altmeyer, Ally Suarez, and Lennart E. Nacke. 2022. 'Much Realistic, Such Wow! A Systematic Literature Review of Realism in Digital Games'. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, 1–21. CHI '22*. New York: Association for Computing Machinery. https://doi.org/10.1145/3491102.3501875.

Rogers, Katja, and Michael Weber. 2019. 'Audio Habits and Motivations in Video Game Players'. In *Proceedings of the 14th International Audio Mostly Conference: A Journey in Sound, 45–52. AM'19*. New York: Association for Computing Machinery. https://doi.org/10.1145/3356590.3356599.

Rumsey, Francis. 2014. 'Spatial Audio — Reconstructing Reality or Creating Illusion?' Niles, IL, May 21. www.youtube.com/watch?v=y82nth2Pnwk.

Schafer, R. Murray. 1969. *The New Soundscape: A Handbook for the Modern Music Teacher*. Don Mills: BMI Canada.

Schwind, Valentin, Katrin Wolf, and Niels Henze. 2018. 'Avoiding the Uncanny Valley in Virtual Character Design'. *Interactions* 25 (5): 45–49. https://doi.org/10.1145/3236673.

Serafin, Stefania, Michele Geronazzo, Cumhur Erkut, Niels C. Nilsson, and Rolf Nordahl. 2018. 'Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions'. *IEEE Computer Graphics and Applications* 38 (2): 31–43. https://doi.org/10.1109/MCG.2018.193142628.

Siu, Alexa F., Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 2020. 'Virtual Reality Without Vision: A Haptic and Auditory White Cane to Navigate Complex Virtual Worlds'. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13. Honolulu: ACM. https://doi.org/10.1145/3313831.3376353.

Slater, Mel. 2009. 'Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments'. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1535): 3549–57. https://doi.org/10.1098/rstb.2009.0138.

Slater, Mel, Domna Banakou, Alejandro Beacco, Jaime Gallego, Francisco Macia-Varela, and Ramon Oliva. 2022. 'A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality'. *Frontiers in Virtual Reality* 3. www.frontiersin.org/article/10.3389/frvir.2022.914392.

Slater, Mel, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, et al. 2020. 'The Ethics of Realism in Virtual and Augmented Reality'. *Frontiers in Virtual Reality* 1. https://doi.org/10.3389/frvir.2020.00001.

Sobchack, Vivian. 1994. 'The Scene of the Screen. Envisioning Cinematic and Electronic "Presence" '. In *Materialities of Communication*, 1st ed., edited by Hans Ulrich Gumbrecht and K. Ludwig Pfeiffer, 83–106. Writing Science. Stanford: Stanford University Press.

Spiel, Katta, and Kathrin Gerling. 2019. 'The Surrogate Body in Play'. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, 397–411. CHI PLAY '19. New York: Association for Computing Machinery. https://doi.org/10.1145/3311350.3347189.

Stingel-Voigt, Yvonne. 2020. 'Functions and Meanings of Vocal Sound in Video Games'. *Journal of Sound and Music in Games* 1 (2): 25–43. https://doi.org/10.1525/jsmg.2020.1.2.25.

Vignemont, Frédérique de. 2011. 'Embodiment, Ownership and Disownership'. *Consciousness and Cognition, Brain and Self: Bridging the Gap* 20 (1): 82–93. https://doi.org/10.1016/j.concog.2010.09.004.

Zhao, Manlin, Zhichao Sheng, and Yong Fang. 2022. 'Magnitude Modeling of Personalized HRTF Based on Ear Images and Anthropometric Measurements'. *Applied Sciences* 12 (16): 8155. https://doi.org/10.3390/app12168155.