

Pop Music Diegesis and the 360° Video

Zack Bresler

To cite this article: Zack Bresler (08 Nov 2023): Pop Music Diegesis and the 360° Video, Popular Music and Society, DOI: [10.1080/03007766.2023.2272680](https://doi.org/10.1080/03007766.2023.2272680)

To link to this article: <https://doi.org/10.1080/03007766.2023.2272680>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 08 Nov 2023.



Submit your article to this journal [↗](#)



Article views: 132



View related articles [↗](#)



View Crossmark data [↗](#)

Pop Music Diegesis and the 360° Video

Zack Bresler 

Department of Popular Music, University of Agder, Kristiansand, Norway

ABSTRACT

One approach to studying music videos is through the framework of diegesis, which considers the relation of sounds to narrative structure in film. Unlike most visual media, music videos flip the diegetic picture as image functions to support music, allowing for new narrative interpretations. Narratological possibilities of several 360° pop music videos are examined to demonstrate pop music diegesis, which operates through navigational agency and diegetic immersion. The viewer of an immersive music video is a staged element of compositional design, implied by agency afforded through interaction and envelopment. Moreover, the essay expounds discourses on popular music in immersive media.

KEYWORDS

360° video; diegesis; immersion; music video; popular musicology

Introduction

This essay builds on existing research into music video and immersive media¹ by asking how immersive pop music video productions can shape the narratives that audiovisual pop texts attempt to illustrate through technologically enabled agency and immersion. The main argument is that so-called immersive media, in this case 360° pop music videos, situate the viewer on various levels within the narrative structure of music video, thus allowing for different modes of narratology and meaning in the agential space. This raises further questions: What are the audiovisual features that enable immersive experience in immersive media, and how do these forms of immersive media elicit subject positions differently from traditional films, recorded tracks, and music videos?

Creators of pop music productions often operate within narrative structures, conveying ideas through audiovisual storytelling. Part of the unfolding of a music video occurs in the “aesthetic space,” where sound and image synthesize hermeneutic positions that are unique to their confluence (Bresler and Hawkins). In addition to source-bonding, the phenomena whereby sounds are associated with their supposed causes as they either appear on-screen or in the memory of the listener (Smalley, “Space-Form” 38), the aesthetic space is formed in the viewer’s interpretation, within which sound and image are connected to abstract feelings, intertextual sources, and deep personal meanings. For this essay, it is important to additionally consider the agential space, as it is through interactivity that the viewer is granted a role in the diegesis of a music video.

CONTACT Zack Bresler  zackbresler@gmail.com  Department of Popular Music, University of Agder, Kristiansand, Norway

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

The present study focuses on 360° music videos, which are a form of virtual reality (VR) videos that are available to stream via platforms such as YouTube and Facebook. A 360° video is a panoramic video format that can be captured using a combination of ultrawide-angle cameras that can see in all directions. This means that the viewing space wraps all the way around the viewer, requiring them to navigate the space to see parts of the complete image. In the most basic level, a 360° video can be navigated on a 2D screen such as a computer and navigated through “click-and-drag” gestures on a mouse or with keyboard controls. Viewing a 360° video on a mobile device is slightly more intuitive as one can simply move their phone around to navigate as if their body represents the position of the 360° camera and the phone is a narrow viewscreen. While this is not the same as Augmented Reality (AR), it functions using similar gestures. In the best case, a 360° video is viewed using a virtual reality headset such as an Oculus, wherein the viewer simply navigates by moving their head around in the virtual space.²

This format is chosen for analysis because 3D and 360° media offer an easily demonstrable case for the viewer’s role in the diegesis of music video. However, the findings of this essay are applicable to “traditional” music videos and even acousmatic music recordings. Furthermore, diegesis and narratology are understudied aspects of popular music and music video. Developing on existing research within the field of popular musicology, this essay proposes a two-fold hermeneutic framework called *pop music diegesis*, which relies on two aspects of engagement with interactive media: agency and immersion. A particular mode of each of these concepts is operational in the majority of 360° pop music videos, termed *navigational agency* and *diegetic immersion*.

Navigational agency refers to the degree to which the viewer has control over their movements, while diegetic immersion refers to the degree to which the viewer has a defined and participatory role within the narrative structure. This framework is demonstrated through the inclusion of various examples from four 360° pop music videos:³ *Life Support* (2018) by Taryn Southern, *Revolt* (2016) by Muse, *The Hills remix* (2015) by The Weeknd feat. Eminem, and *Stor Eiglass* (2015) by Squarepusher. These videos are all freely available on YouTube and combine to represent a wide array of production techniques and narrative structures that demonstrate the various degrees of pop music diegesis. They demonstrate the efficacy of interpreting music video through this framework.

Pop Music Diegesis

In media and film scholarship, it has been common to reference sound and music with respect to a film’s diegesis—that is, the internal, logical space of the film’s story world. The terminology is rooted in narratological literature studies, having been brought into film scholarship by Claudia Gorbman. In Gorbman’s application, diegetic sound is that which emanates from the story world itself (i.e. dialogue, sound effects), while nondiegetic sound is supporting the narrative for the viewer but is not “heard” as such from “within the scene” (e.g. background music) (197). This dichotomy between diegetic and nondiegetic has been problematized by most film scholars since. For example, Ben Winters has argued for the essentiality of much nondiegetic music and sound to the “identity of the fictional narrative space presented in film” (230). Winters ultimately argues for the usefulness of the terminology, suggesting the term “intradiegetic” as the

broad category for sounds that are fundamental to narrative structure—and thus central to the film’s diegetic frame—but which are not implied to exist in the fictional story-world as such (237–38).

There are numerous theoretical and disciplinary challenges to discussing diegesis at all in relation to music video. In film studies, narrative is often pitted against spectacle as its opposite. For example, Andrew Darley claims that:

in critical studies of the dominant cinema institution, centred upon analysis of classical narrative films, attention has most frequently focused on the “tension” between the narrative dimension and the visual dimension, that is, between identifying with characters, being absorbed in a fictional world and following the plot on the one hand, and the pleasures involved in looking at images on the other. (104)

Darley continues that “spectacle is, in many respects, the antithesis of narrative [It] halts motivated movement” (104). This is a concern for this study, in particular since the aesthetics of music videos are often described *primarily* in terms of spectacle (Ålvik; Auslander, *Liveness*; Auslander, *In Concert*; Burns; Hawkins, *Settling*; Hawkins, *British*; Hawkins, *Queerness*; Korsgaard, “Music Video Transformed”; Korsgaard, “SOPHIE’s”; Burns and Lafrance). Korsgaard claims that “rather than comprising a unified field, music video is actually defined by its very heterogeneity, its wide range of different audiovisual expressions” (*Music Video After MTV* 37). In its narrativity, “music video presents a range all the way from extremely abstract videos emphasizing color and movement to those that convey a story” (Vernallis, *Experiencing* 3). Thus, one might surmise that there are as many genres of music videos as there are of music. While music videos are often spectacular, many have varying degrees of traditional narrative structure that make diegetic perspectives relevant. Spectacle in film has also been linked to a Freudian instinct toward voyeurism, scopophilia, and spectator positioning and to the pleasure of viewing a film that arises from the “conditions of screening and narrative conventions [that] give the spectator an illusion of looking in on a private world” (Mulvey 9). Following this, the 360° video presents the most extreme version of the subjective camera, wherein the viewer has not only a first-person view but a fully controllable one, and wherein part of the format lies in the pleasure of viewing from the inside.

Any perceived conflict between narrative and spectacle is ultimately semantic, rooted in definitions of narrativity and diegesis that include primarily classical narrative elements like characters and plots. In film studies, a related debate has taken place regarding the role of special effects in films—in particular, the grand, digitally constructed visual spectacles in movies like *Jurassic Park*, *Avatar*, and *Titanic*. Aylish Wood has pointed out that visual effects such as the detailed digital reconstruction of the *Titanic* “operate at another dimension of the narrative . . . that places a particular emphasis on the story of the fall of this technological giant” (372) and that it is the overlooking of this other dimension that “leads commentators to argue that spectacle interrupts narrative” (372). Similarly, music videos make use of the “musicalization of vision,” whereby images are “shaped according to and respond to different musical parameters” (Korsgaard, *Music Video After MTV* 65).

While the images of most music videos are undoubtedly spectacular in the sense that they exist primarily for the pleasure of viewing them, this does not mean they are not

diegetic. To the contrary, the spectacular elements in the musical presentation of image in music video, like special effects in contemporary films, can be seen as diegetic because they operate on the level of world-*building*, as opposed to world-explaining or world-developing. Moreover, abstract as they may be, stories with low levels of classical narrativity are still stories, and the open-ended audiovisual design of music videos encourages viewers to read a multitude of narratives that explain the meanings of pop songs.

Considering diegesis from within the discipline of critical musicology, Walther-Hansen has theorized about the “phonographic diegesis” of pop music recordings, including a typology of recorded music staging centered around the idea of diegetic, metadiegetic, and extradiegetic sounds (34–35). Derived from the work of Gérard Genette, “metadiegetic” and “extradiegetic” refer to sounds that exist outside the story-world (i.e. diegetic sounds) but that nonetheless contribute to narrative. Metadiegetic sounds are those emerging from within characters—for example, the sounds of internal monologues or the sounds of imagination—while extradiegetic sounds are those external to the story-world completely—for example, background music. Walther-Hansen approaches analysis by focusing on the edge-cases, wherein the diegetic framing changes through the course of the track, thus exposing diegetic boundaries. While he is concerned primarily with sound recordings, there is a further distinction to be made when considering the diegesis of a music video, and furthermore an immersive music video, as new narrative interpretations will surface in the aesthetic space as the music is made audiovisual, and even more in the agential space as the viewer is staged within the video using immersive technologies. 360° music videos can serve as easily accessible edge-cases for understanding the motivations, technologies, and interpretations that make up the creation and reception of music videos in general.

The diegetic frame of a pop music video is a complicated matter. Walther-Hansen’s typology may be useful for acousmatic recordings, but it is difficult to label any sounds at all in a music video as meta- or extradiegetic, since the video itself acts to clarify the diegetic role of the sound events whose narrative framing may be in question in the sound recording. Thus, music video confounds the normal conceptualization of sonic diegesis, since, in contrast with other forms of video entertainment, the sound in a music video is arguably the main text while the image serves a supporting role. The diegetic dichotomy fails to capture the complexities of the sound-image relationship in pop music video. In an essay that explains the use of music video aesthetics in films in general, Vernallis reiterates that music video is a fundamentally *musical* form:

Free-ranging camera movements like dollying, handheld, reframing, and crane shots reflect music’s flowing, processual nature; blocks of image highlight song structure, intense colourization illuminates features like a song’s harmony, sectional divisions and timbre; visual motifs speak to musical ones (“Music Video” 277)

If the diegetic dichotomy is, in Kassabian’s words, “not sufficient to cover the various examples of music that cross over, through, around, and under that boundary” (91), then it is even less adequate for music videos, whose entire point seems to be to illuminate, demonstrate, exaggerate, and complicate the stories told by music recordings. The viewer comes to the music video knowing, in most cases, that it is an extension of an already existing recorded track, and this intertextual duality of the music video implies

a multiplicity of entry points to the song's interpretation. Part of what can constitute a song's meaning is in the story it tells, and the viewer's role as the interpreter of musical meaning cannot be ignored, since, contrary to being a passive and external entity, the viewer of a music video is the switch that completes the narrative circuit.

Navigational Agency

The experience of the viewer is important in understanding narrative structures in music videos. When used for music production, immersive and interactive media technologies such as virtual reality, surround sound and 3D audio, and 360° videos create a situation in which the viewer can be seen as a staged part of the composition (Bresler and Hawkins). This is because the viewer's experiences are centered: The viewer is placed on the audiovisual stage and thus thrust into an active and participatory role within the performance. Movement by the viewer implies the possibility for immersion and interactivity, and *navigational agency* describes the immersive pleasures of interacting with the narrative of a music video through spatial movement and control. Navigational agency is a spectrum, and thus any video will have varying degrees and modes of it through the types and qualities of movement afforded.

In some 360° music videos, the viewer is granted an easily discernible and defined perspective, wherein the viewer's placement in the diegetic frame is explicit enough that the viewer may be considered a character within the story. Other times, the viewer is placed into the scene as an outside observer. In any case, the viewer is invited to participate through interactions with the stage. It is in these interactions that 360° videos and other immersive media formats make explicit an implicit feature of music and music videos in general: that for a song to mean something is an active process that includes the experiences of the viewer. In pop music, and especially in pop music videos, the construction of diegesis includes the viewing experience itself. As the user interacts with the music video through various means, they participate in creating the very narrative they consume.

This notion is supported through the concept of ecological perception, first introduced by James Gibson in psychology (*Ecological; "Theory"*) and brought into musicology and music psychology through Clarke's theory of an ecological approach to the perception of musical meaning. Clarke posits that meaning comes forth from the confluence of the listening "environment" (a technical term that encompasses not only the space and place of the listening but also the background, taste, and experience of the listener) and the musical performance, be it recorded or live, replete with its various structural affordances. While the term "affordance" implies a kind of structuralism where particular structures in the pop score demand particular responses from listeners, Gibson maintains that affordances have a dialectical quality that implies "the complementarity of the animal and the environment" (*Ecological* 119).

Writing about digital hypertext narratives, Murray asserts that "activity alone is not agency" (124). Agency is more than the sum of the interactive participations of the viewer. It is the "satisfying power to take meaningful action and see the results of our decisions" (123). This definition is useful but unclear as to what constitutes a "meaningful action." Is it necessary that the user can do whatever they want without restriction? Or can the medium place constraints, even large ones, on the viewer's possible actions while

still yielding degrees of agency? A central argument of this essay is that the freedom to navigate space is meaningful on its own.

In general, music videos are not hypertexts—viewers do not make decisions that constitute direction for the narrative, and, regardless of the viewer’s actions, the plot will unfold in the same way. However, VR and 360° music videos offer the viewer interactivity in the form of spatial navigation, where the narrative unfolds around the viewer and she must use her body to actively engage with the text to experience it fully. In this way, while immersive music video is not hypertextual, it can nonetheless be considered a form of ergodic cybertext, where “nontrivial effort is required to allow the reader to traverse the text” (Aarseth 1). Although music videos are, by definition, linear in the sense that they follow musical form, interaction with the virtual space allows the reader to be “constantly reminded of inaccessible strategies and paths not taken, voices not heard” (Aarseth 3). Indeed, spatial navigation is itself a highly pleasurable form of interactivity: “[C]onstruing space and moving through it in an exploratory way . . . is a satisfying activity regardless of whether the space is real or virtual” (Murray 125).

In dealing with navigational agency in VR and 360° music videos, there are several spatial and navigational aspects to consider:

- (1) *Stage configuration*—what are the size, shape, and depth of the virtual environment?
- (2) *Degrees of freedom*—how much movement is the viewer afforded?
- (3) *Range of motion*—what are the limits of this movement?

Stage Configuration

The ideal implementation of virtual reality is normally imagined as something like the holodeck—the fictional room from *Star Trek* in which the user enters and tells the computer the parameters of the environment and story they would like to experience, and the room transforms to the specifications, creating for the user a completely accurate sensory experience that is indistinguishable from reality. While the ideal of the holodeck has thus far been impossible to deliver, this imagination for what VR could someday become has driven much of the research and interest in VR since its inception in the 1990s (Murray). Spatially, the holodeck metaphor demonstrates what is ultimately necessary to create a virtual spatial environment. Marie-Laure Ryan claims that “being inside a computer-generated world involves three distinct components: a sense of being surrounded, a sense of depth, and the possession of a roving point of view” (53).

Through degrees of freedom and range of motion, concepts addressed in the next sections, immersive music videos feature this “roving point of view.” Perceived dimensions of the environment and their morphologies are a useful starting point for discussion. The model Virtual Audiovisual Space (VAVS) is useful for describing and interpreting the spatial configuration of the VR audiovisual stage (Bresler and Hawkins). VAVS uses as its basis Camilleri’s model of sonic space to describe the position, disposition, and temporal unfolding of sound and visual objects. In addition, it builds on Denis Smalley’s notion of source-bonding to describe the connection of sounds within a scene to supposed causes (“Space-Form”) and on the notion of the

aesthetic space, which comprises the interpreted meanings that are synthesized in the viewer's audiovisual experiences. Any interpretation of space within this context necessitates a description of its apparent size, shape, and quality.

Sound and image are not always aligned in their spatial construction. In music videos, while the video may be shot in a real space (an outdoor stage, a small room, a warehouse) the sounds of the pop recording are not normally altered to match the expected sonic properties of the video's scenic space. Unlike many kinds of VR experience, 360° music videos often feature 3D visuals with static, stereophonic audio. In other words, while the viewer is invited to interact with and move through the visual space, the sonic space often remains a fixed stereo image that, depending on whether viewing in a head-mounted display or on a screen at a distance, either follows the user or seems to be an unmoving element. Moreover, just as in 2D music videos, 3D videos commonly feature animated visual scenes without analog in the physical world and for which the viewer would have no auditory reference for the acoustic properties the sound in such a space should have.

For example, in Muse's *Revolt* video, the scene opens with the sounds of sirens and driving cars while text overlays the screen explaining banishment of "freedom" in 2025 as "government drones fill the sky." A moment later, police vehicles appear in frame—black, tanklike SUVs that stop to release masked officers. Panning around the view in the 360° video, it is noticeable that the sounds do not move to reflect the viewer's movements—the police car sirens do not appear to come from the direction of the cars themselves; rather, they are in stereo as in the acousmatic recording. Shortly thereafter the band begins performing outside in the open, positioned as if on stage at a rock concert while their audience is the clashing of military police and rebellious protesters. What is heard over the video is in fact the stereo release of "Revolt," without any spatial change to reflect the outdoor scene or any spatialization of sound to position the band's members where they are with respect to the viewer's gaze in the 3D virtual space (Figure 1).

These contradictory spatialities are not surprising, given that they are part and parcel of the music video paradigm. Brøvig-Hanssen and Danielsen confirm that surreal spatial



Figure 1. *Revolt* by Muse: view of bassist Chris Wolstenholme with the riot happening in the background; visible in frame is a "government drone."

configuration has “a tendency to point the listener toward a real-world physical phenomenon even as it acts to undermine that reality” (27). In the above example, the two physical phenomena are those of the nighttime, outdoor protest clash and the indoor, studio-polished recording of the rock band Muse. As these two realities fade into each other they do not cause conflict for the viewer. On the contrary, they work together abstractly to communicate an effective narrative statement on civil conflict. Smalley has referred to “spatial simultaneity” as the phenomenon whereby the listener can be, without conflict, “aware of simultaneous spaces” that are either implicit or explicit, or whereby “the listener remains aware of the existence of a space in its absence” (“Spectromorphology” 124). While this refers to the multiplicity of spatialities within audio recordings, it is apt to consider the contradictions in spatiality between sound and image (or, in many cases, both within the recording and within the image, and between the sound and image) in music video.

Degrees of Freedom

In the field of immersive and interactive media technologies, the level of designed spatial control is often described according to degrees of freedom (DOF), where each degree represents a possible axis for free movement in the virtual space. The first three axes, which comprise the total movement possibilities for so-called 3DOF media, are the rotational axes, wherein the user can move within the parameters of yaw (swivel the head as in the “no” gesture), pitch (as in nodding “yes”), and roll (tilting the head toward the shoulders). These dimensions combine to represent all the possible movements of the head at the level of the neck without moving the shoulders, and technologically they are simple to implement since when recording 360° videos, the 360° camera is a stationary object that records a 3D visual field. Thus, the 3DOF video allows the viewer to rotate their relatively narrower frame of view within this stationary spherical or cylindrical image.

In animated or digitally manipulated videos, three more axes can be added to create the 6DOF environment, where the viewer can additionally move along the translational axes: forward-backward, left-right, and up-down. This is impossible with live-action video recordings since one cannot place a camera at all the possible points in a scene where a viewer may want to position themselves. Regardless, much research has been done on producing 6DOF audio recordings—for example, using combined third-order ambisonic microphones (Rivas Méndez et al.). While these kinds of techniques for recording make 6DOF sound possible, mono and stereo recordings mixed in virtual 3D sound formats such as ambisonics or Dolby Atmos seem to be the preferred method for recordists attempting to produce content for 3D systems. Björk’s 2019 album of VR music videos, *Vulnicura VR*, utilizes 6DOF movement with 3DOF audio—the user is free to move in all six degrees in the VR video, while the audio is locked to the motions of the head at the neck (Bresler and Hawkins).

In many cases, producers of live-action 360° videos move the camera during capture, thus creating translational movement on behalf of the viewer, who only has rotational control. Currently, most immersive music videos, and all the music videos discussed in this essay, are 3DOF live-action or animated videos.⁴ For example, in The Weeknd’s *The Hills remix*, the viewer is granted a perspective of a hovering camera that follows the



Figure 2. *The Hills* remix by The Weeknd feat. Eminem: front and back views as meteors destroy the city.

artist's slow walking out into a dystopian urban street at night (Figure 2). Initially, the view faces The Weeknd directly, although he never gazes to the camera, instead staring straight ahead with a look of complete absence. As the viewer rotates, they can see what look like asteroids crashing down on the city.

Looking back at the artist for reaction, none is found, and instead the viewer continues the slow progression through the scene. This clip shows how rotational movement combines with predefined lateral movement to create a forward sense of motion. The video moves with the artist who is walking in a slow procession, but the camera movements are not steady. Instead, they bob up and down with each step, creating the impression that the viewer too is walking with the artist, or perhaps they are his own out-of-body experience in this jarring scene of destruction. Regardless of the amount of movement afforded to the viewer, the existence of any degrees of freedom provides the user with navigational agency, especially when there exists visual or auditory content that is only accessible by reorienting one's focus in a 3D visual field.

Range of Motion

Even within the freedom of motion granted in an immersive video, there exists variation in the real and perceived range of that motion. For example, although the videos discussed to now have been shot in 360°, the producer may also choose to limit the possible field to a 180° frame for various reasons. For one, cutting the field of view in half, effectively into a single hemisphere, can allow for a higher fidelity image since a smaller frame can provide higher resolutions with the same camera. Additionally, this choice can set limits on navigational agency by design, since the producer may feel they need to restrict the movements of the viewer into this single hemisphere. A definition of agency such as Murray's (124) can imply that part of what enables it in cybertext is a one-to-one relation between a taken action and an intended result. Given that such relation can be difficult to define, Mason argues that movement alone does not constitute diegetic agency, but rather affect, which "is a necessary path to agency . . . and we must be fluent with our means of affect to experience immersion" (31).

The experience of viewing an immersive video for the first time supports Mason's argument. A natural starting point is to search for the boundaries of the experience, asking: Is the video a full 360° or 180°? Is there an avatar or a body? Is there freedom to move in space, or simply to rotate the perspective? In practice, the first experience of a VR or 360° production is one of determining the range of motion possible, and this activity of finding the boundaries serves to increase the chances of having an immersive experience, since what is required is that one feels able to make free and meaningful choices. By knowing the limitations of these choices, one can more easily make the kinds of choices that are possible. Without taking this step, a viewer might, in the middle of a video, suddenly decide to turn around but find they cannot, or might try to move their hand but find they do not have one. Any of these types of experiences only serve to remind the viewer of the nonreality of their experience, ultimately taking them out of it.

Diegetic Immersion

Immersion, in general, is the experience of losing oneself within an activity, and it is often likened to experiences such as the notion of the flow state (Csikszentmihalyi). Thus, immersion can be described in terms of the pleasure of a repetitive action—the feeling of losing time when engaged in enjoyable, repetitive, and comprehensible activities. But it can also occur, as in the flow state, in activities that are the right amount of both challenging and engaging, such as the cognitive task of reading and comprehending a difficult text, since this experience can be to the reader both profound and empowering.

Notions of immersion and "letting go" have also been a central part of studies of dance club music and club culture, and this is fundamentally tied to notions of temporality. For example, Frith states that "dance is not just to experience music as time, it is also to experience time as music . . . more intense, more interesting, more pleasurable than 'real' time" (156). Similarly, Hawkins shows how the dance floor can enable "the sensation of being 'loved up' (an expression often used by DJs and clubbers) [which] suggests a state where the body of the individual or the crowd is immersed in sound" ("Temporal" 122). Importantly, these immersive club experiences are enabled by musical features such as

the beat and the groove, as well as by environmental and social factors such as lighting, sound volume, and involvement of the crowd.

Immersion is often pitted as the antecedent to agency in multimedia like games and hypertext narratives, since high degrees of agency are seen as breaking the story into small, difficult-to-synthesize parts, while higher degrees of immersion need more complex narratives that require consistency and reduce the possibilities for agency. So, what constitutes the elements that form diegetic immersion within 360° videos? There are two main factors that dictate the propensity for diegetic immersion. The first is *visual saturation*—the construction of the relevant visual field. Second is the *perceived role*, which is related to the viewer's narrative embeddedness and designed experiences of embodiment.

Visual Saturation

Although the viewer may be free to move within a visual scene, it is not necessarily the case that there is something happening in every part of the scene. The amount of space containing engaging visual material needs to be considered—albeit that what may be considered “engaging” or “interesting” in this context is certainly a matter for individual interpretation, since the absence of image can be just as engaging as the presence of one. Still, it is true that in music videos, directors guide the viewer's gaze through camera movements, framing, color, and lighting in order to invite them into diegetic immersion. *Visual saturation* refers to the amount of utilized space within the visual field and how the producers of the video have used visual features to suggest and guide the viewer. Visual saturation is different from the apparent physical dimensions of the 360° or 180° video—one can talk about the perceived size and shape of the stage in terms of what is possible. This is something more qualitative and hermeneutic, which is the amount of the visual field that the viewer finds relevant to explore in their viewing experience.

An example from the opening of Taryn Southern's *Life Support* is illustrative. The scene begins in a spooky wood as the viewer is moved toward a lone, run-down house with a nighttime city skyline visible in the distance. Once the viewer arrives at the house, however, they are transported into a vast, dark space with nothing surrounding them except a large, strange machine flanked on both sides by rectangular screens showing images of brain scans (Figure 3). Turning around, the viewer will see that this machine apparatus is, for some time, the only visible object in the entire video, which is just as well because it is visually captivating with its moving arm waving around a human body like a rag doll. After some time, lights and images begin to emanate from the machine, moving past the viewer to the rear of the scene, drawing their attention there to notice that there are now things happening behind them—flashing and moving light patterns that reveal parts of a seemingly infinite darkness.

Here, the producers of the video have carefully crafted the viewer's visual attention by first revealing a space they can freely explore (the woods) before drawing their attention to a narrow frame (the machine), which encourages them to remain still for a moment. Finally, the producers slowly open up the visual scene with moving lights, reminding the viewer of the immersive qualities of their visual experience. These strategies invite the viewer to explore the space while guiding them toward the most relevant visual aspects of the scene.



Figure 3. *Life Support* by Taryn Southern: a large and mysterious machine interfaces with a lifeless (for now) humanoid body.

Perceived Role and Viewer Subjectivity

Considering how immersion functions within so-called immersive audiovisual media, it is important to ask: How embedded is the viewer into the story? Because the video is 360°, there will be audiovisual material that surrounds the listener, who will have some degree of navigational agency within the space. However, their narrative embeddedness is a question of their role within the story. In short, the answer to the question will lie somewhere on a line between an outside observer and independent narrative agent.

Helpful in this context is the notion of the subject position in film studies, which Johnston has defined as “the way in which a film solicits, demands even, a certain closely circumscribed reading from a viewer by means of its own formal operations” (333). In other words, its use is an attempt to allow for analyses of meaning that are constructed both in the formalization of the film and in its reception, thus skirting the failures of both structuralism and postmodernist relativism. Subject position has been used to describe the listener’s role in popular-music meaning. For example, Clarke suggests that in music the narrative content and the framing of subject position occur “not through the semiotic language of ‘codification,’ but through the perceptual principle of ‘specification’” (125). In other words, although the viewer is ultimately the arbiter of meaning, their interpretations are nonetheless shaped in part by structural elements of the music, which specify relationships and correlate to particular responses. The semiotic language of audiovisual codes is but one way of explaining the structures that lend themselves to perceptual specification. In music analysis, one can only accurately explain such textual elements, one’s own interpretation of them, and perhaps some alternative interpretations they can imagine.

How then do immersive media elicit subject positions differently from traditional films, recorded tracks, and music videos? Extending the concept of subject position, immersive media engage directly in subject positioning. That is, through the placement of

the viewer directly on the stage, and in particular through their freedom of movement on the stage, the 360° video has become a platform for the viewer to participate in positioning their own subjectivity in the audiovisual scene. Granted, this embeddedness can be aided through the implication of a character role.

Scenes from Muse's *Revolt* can be helpful in unpacking subject positioning. At first the viewer seems to be simply an observer—they are moved through the scene, which was presumably recorded on a moving 360° video camera, going back and forth between the protest and the performing band. Embedded within the visual field are futuristic, digital, circular overlays that seem to identify objects within the scene, such as a person's face or a vehicle in the background, displaying illegible data about the various identified objects, reminiscent of the first-person views from films like *The Terminator* and *RoboCop* (Figure 4). In the opening on-screen text, we are told of "government drones," which are visible floating around the scene, and it quickly becomes clear to the viewer that their perspective is that of one of these autonomous, robotic surveillance cameras. Through clever manipulation of the video, the producers have placed the viewer firmly within the diegesis, giving them the privileged view of the imagined government overseers. The band also stage themselves as sympathetic to such causes—as the cameras hover over the musicians they identify their faces in the same way as those of the revolting citizens. While the viewer has no control over their lateral movements, they nonetheless have rotational control, and looking around at various people and objects as they are automatically scanned and identified, one cannot help but feel a sense of complicity.

Assisting in the development of the user-character's role is the presence of a body or an avatar. The above example illustrates this to an extent—the embodiment in first-person of a surveillance drone is confirmed through the overlaid surveillance data, and the erratic movements of the camera, which mimic the movements of the other drones that are visible in the video, encourage the user to take the role of the camera by rotating their own view in erratic ways. Going further, one can be granted a human body or bipedal avatar, which can serve as a stand-in for one's own body and heighten the



Figure 4. *Revolt* by Muse: a protester has been identified by the viewer's drone with the text "Target Armed".

embodied experience. For instance, in Squarepusher's *Stor Eiglass*, the viewer finds themselves in a neon, psychedelic dreamscape, moving steadily through a barrage of imagery that conjures up memories of video games, 1980s shopping centers, and the imagined sci-fi city of the future (but in high-contrast). When the viewer looks around in the scene, they will notice upon looking down that they have been given a body (Figure 5)—naked and cartoonishly shaped, and with a cleanly severed neck (complete with a visible bone) just below the point of view, as if the viewer's head is floating above.

The body appears to be sitting down at first, with arms extended and gripping a set of joysticks, but then, throughout, the vehicle on which we travel changes, becoming at one point a bicycle, and eventually it goes away, and we see our character walking. As the song progresses, the body changes, at one point suddenly becoming a woman, with large, naked breasts now obfuscating some of the view below. Later in the video, as the song gets more energetic and the imagery becomes more and more psychedelic and fractal, the body disappears entirely as the viewer finds themselves in an overwhelming, symmetrical, spinning scene of changing shapes and colors.

The above example with its realistic naked body demonstrates very clearly how the gendered body is always part of the design of embodied experiences. Whenever media imply an embodied experience or subject position, it is critical to ask the question: Whose body is it that is being implied? A major part of the utopian ideology of the digital-virtual environment is the freedom that the digital world grants us in transforming our “creative thoughts and imagination” into “reality and actuality through digital means” (Rambarran 1). *Stor Eiglass* depicts a parodic spin on this, as the nude body on display is at first coded male, and later (and without warning) female—at times it is completely motionless, and other times it moves in an autonomic fashion. Always visible is the decapitated neck upon which the viewer's lens resides. Ultimately, Squarepusher offers a humorous critique of utopian virtual ideology through this imagery, illustrating that in media that purport to transform a person into their ideal digital selves, the best they can offer is a new set of interchangeable avatar categorizations.

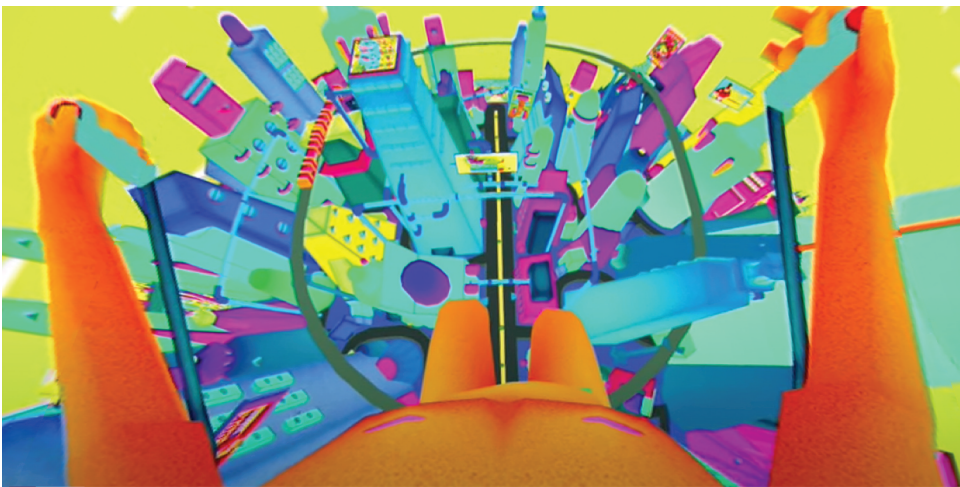


Figure 5. *Stor Eiglass* by Squarepusher: Looking down at the psychedelic city, the viewer's “body” is visible.

Conclusion

Being immersed in a story is a fundamentally human experience, and thus it is no surprise that the multitude of technologies for multimedia storytelling are so concerned with assisting us to more easily find such experiences. While the discourses around immersion in film, music, video games, and other forms of media often focus on the distinction between agency and immersion, agency and immersion are in fact allies in storytelling. In different forms of media, they function in different ways. For example, in video games, players have much higher degrees of agency than in more structured media such as film, but there still exists a wide range of agency from the auto-scrolling, single-control interactivity of mobile games like *Flappy Bird* to the total open-world possibilities in games like *The Legend of Zelda: Breath of the Wild* (Collins). Within this range there are many levels of complexity within the stories that are told, or are able to be told. The same is true for other media—the introduction of expanded modes of access and interaction creates a different range of possibilities for storytelling.

Music videos are a special form of media. Kelly has insisted that they are “always already a hybrid medium, comprising audio and visual forms and structures that intersect and interrelate in ways that can be described as intermedial” (219). Unlike other forms of film, television, or video, where music extends the interpretive possibilities of the visual and dialogic narrative, music videos do the opposite, using visuality to extend the hermeneutic position of the musical text. Considering popular music in new, immersive, and interactive forms of media, including 360° videos, gives analysts recourse to ponder anew the ways that subject positioning can occur in pop multimedia. The formation of pop music video diegesis is not only a structural and musical phenomenon but is itself dialogical. In other words, viewers of music videos are the cocreators of narrative structure. 360° music videos offer an easy-to-demonstrate case for this, since the way they stage listeners within the story world is obvious. However, these processes are not exclusive to music presented in these technologically innovative ways. Viewers of music videos have always been a nexus of audiovisual meaning, and while the story is told by the creators of a music video, the diegetic frame is only complete when we acknowledge the role of the viewer in its formation.

Notes

1. See Bresler; Bresler and Hawkins; Burns and Hawkins; Burns and Woods; Dibben; Hansen; Jirsa and Korsgaard; Kelly; Korsgaard, “Changing Dynamics”; Liljedahl; Morie; Rambarran; Ryan; Vernallis, Herzog, and Richardson; Walther-Hansen; Winters.
2. One very affordable way to experience 360° video in virtual reality is by using a smartphone headset viewer, such as a Google Cardboard. This is a cheap, stereoscopic, handheld set of lenses that a smartphone slots into. At the time of publication, a Google Cardboard sells for less than \$10, and plastic headsets from other manufacturers can be purchased at electronics stores very affordably. Additionally, Google has instructions on building one DIY using everyday materials (<https://arvr.google.com/cardboard/get-cardboard/>).
3. *Life Support* by Taryn Southern: <https://www.youtube.com/watch?v=LW19Oi2NHps>; *Revolt* by Muse: <https://www.youtube.com/watch?v=91fQTXrSRZE>; *The Hills remix* by The Weeknd feat. Eminem: <https://www.youtube.com/watch?v=2fhjdtQDcOo>; *Stor Eiglass* by Squarepusher: https://www.youtube.com/watch?v=6Olt-ZtV_CE.

- In general, 6DOF experiences currently require a complete virtual reality implementation such as an Oculus Rift headset, since the CPU processing required for playback is significant and cannot be effectively streamed or played on most mobile devices.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Notes on Contributor

Zack Bresler is a musicologist and sound artist based in Stavanger, Norway. Motivated by an obsession with immersive and interactive media such as surround and 3D sound, virtual reality, and interactive art, he focuses his research primarily on pop music as it is presented in audiovisual new media, with an emphasis on the aesthetics of music productions, compositional design, listener experience, and staging. The research in this essay was funded through a doctoral fellowship at the University of Agder, Norway, where Bresler successfully defended his dissertation in 2022.

ORCID

Zack Bresler  <http://orcid.org/0000-0003-0861-7323>

Works Cited

- Aarseth, Espen J. *Cybertext: Perspectives on Ergodic Literature*. Johns Hopkins UP, 1997.
- Ålvik, Jon Mikkel Broch. “‘Armed with the Faith of a Child’: Marit Larsen and Strategies of Faking.” *The Routledge Research Companion to Popular Music and Gender*, edited by Stan Hawkins, Routledge, 2017, pp. 253–66.
- Auslander, Philip. *In Concert: Performing Musical Persona*. U of Michigan P, 2021.
- . *Liveness: Performance in a Mediatized Culture*. 2nd ed., Routledge, 2008.
- Bresler, Zack. “Immersed in Pop: 3D Music, Subject Positioning, and Compositional Design in the Weeknd’s ‘Blinding Lights’ in Dolby Atmos.” *Journal of Popular Music Studies*, vol. 33, no. 3, 2021, pp. 84–103.
- Bresler, Zack, and Stan Hawkins. “‘A Swarm of Sound’: VR Immersion in Björk’s Video ‘Family’.” *Music, Sound, and the Moving Image*, vol. 16, no. 1, 2022, pp. 29–52.
- Brøvig-Hanssen, Ragnhild, and Anne Danielsen. *Digital Signatures: The Impact of Digitization on Popular Music Sound*. MIT P, 2016.
- Burns, Lori Anne. “The Concept Album as Visual-Sonic-Textual Spectacle: The Transmedial Storyworld of Coldplay’s Mylo Xyloto.” *IASPM Journal*, vol. 6, no. 2, 2016, pp. 91–116.
- Burns, Lori A., and Stan Hawkins, editors. *The Bloomsbury Handbook of Popular Music Video Analysis*. Bloomsbury Academic, 2019.
- Burns, Lori, and Marc Lafrance. “Celebrity, Spectacle, and Surveillance: Understanding Lady Gaga’s ‘Paparazzi’ and ‘Telephone’ Through Music, Image, and Movement.” *Lady Gaga and Popular Music: Performing Gender, Fashion, and Culture*, edited by Martin Iddon and Melanie L. Marshall, Routledge, 2014, pp. 117–47.
- Burns, Lori, and Alyssa Woods. “Humor in the ‘Booty Video’: Female Artists Talk Back Through the Hip-Hop Intertext.” *The Routledge Companion to Popular Music and Humor*, edited by Thomas M. Kitts and Nick Baxter-Moore, Routledge, 2019, pp. 310–20.
- Camilleri, Lelio. “Shaping Sounds, Shaping Spaces.” *Popular Music*, vol. 29, no. 2, May 2010, pp. 199–211.

- Clarke, Eric F. *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford UP, 2005.
- Collins, Karen. *Playing with Sound: A Theory of Interacting with Sound and Music in Video Games*. MIT P, 2013.
- Csikszentmihalyi, Mihaly. *Flow: The Psychology of Optimal Experience*. Harper & Row, 1990.
- Darley, Andrew. *Visual Digital Culture: Surface Play and Spectacle in New Media Genres*. 2000. Routledge, 2002.
- Dibben, Nicola. "Visualizing the App Album with Björk's Biophilia." *The Oxford Handbook of Sound and Image in Digital Media*, edited by Carol Vernallis, Amy Herzog, and John Richardson, Oxford UP, 2013, pp. 682–706.
- Frith, Simon. *Performing Rites: On the Value of Popular Music*. Harvard UP, 1996.
- Genette, Gérard. *Narrative Discourse Revisited*. Translated by Jane E. Lewin. Cornell UP, 1988.
- Gibson, James J. *The Ecological Approach to Visual Perception*. 1979. Psychology Press, 2015.
- . "The Theory of Affordances" *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by Robert Shaw and John Bransford, 1977, pp. 67–82.
- Gorbman, Claudia. "Narrative Film Music." *Yale French Studies*, vol. 60, 1980, pp. 183–203.
- Hansen, Kai Arne. "(Re)reading Pop Personae: A Transmedial Approach to Studying the Multiple Construction of Artist Identities." *Twentieth-Century Music*, vol. 16, no. 3, Oct 2019, pp. 501–29. doi:10.1017/S1478572219000276.
- Hawkins, Stan. *The British Pop Dandy: Masculinity, Popular Music and Culture*. Ashgate, 2009.
- . *Queerness in Pop Music: Aesthetics, Gender Norms, and Temporality*. Routledge, 2016.
- . *Settling the Pop Score: Pop Texts and Identity Politics*. Ashgate, 2002.
- . "Temporal Turntables: On Temporality and Corporeality in Dance Culture." *Musicological Identities: Essays in Honor of Susan McClary*, edited by Steven Baur, Raymond Knapp, and Jacqueline Warwick, Ashgate, 2008, pp. 121–34.
- Jirsa, Tomáš, and Mathias Bonde Korsgaard. "The Music Video in Transformation: Notes on a Hybrid Audiovisual Configuration." *Music, Sound, and the Moving Image*, vol. 13, no. 2, Autumn 2019, pp. 111–22.
- Johnston, Claire. "Structuralism and Its Aftermath." *The Cinema Book*, edited by Pam Cook and Mieke Bernink, 2nd ed., BFI Publishing, 1999, pp. 323–421.
- Kassabian, Anahid. "The End of Diegesis as We Know It?" *The Oxford Handbook of New Audiovisual Aesthetics*, edited by John Richardson, Claudia Gorbman, and Carol Vernallis, Oxford UP, 2013, pp. 89–106.
- Kelly, Jem. "The Palimpsestic Pop Music Video: Intermediality and Hypermedia." *The Bloomsbury Handbook of Popular Music Video Analysis*, edited by Lori A. Burns and Stan Hawkins, Bloomsbury Academic, 2019, pp. 219–33.
- Korsgaard, Mathias Bonde. "Changing Dynamics and Diversity in Music Video Production and Distribution." *The Bloomsbury Handbook of Popular Music Video Analysis*, edited by Lori A. Burns and Stan Hawkins, Bloomsbury Academic, 2019, pp. 13–26.
- . *Music Video After MTV: Audiovisual Studies, New Media, and Popular Music*. Routledge, 2017.
- . "Music Video Transformed." *The Oxford Handbook of New Audiovisual Aesthetics*, edited by John Richardson, Claudia Gorbman, and Carol Vernallis, Oxford UP, 2013, pp. 501–24.
- . "SOPHIE's 'Faceshopping' as (Anti-)Lyric Video." *Music, Sound, and the Moving Image*, vol. 13, no. 2, Autumn 2019, pp. 209–30.
- Liljedahl, Anders Aktor. "Musical Pathfinding; or How to Listen to Interactive Music Video." *Music, Sound, and the Moving Image*, vol. 13, no. 2, Autumn 2019, pp. 165–85.
- Mason, Stacey. "On Games and Links: Extending the Vocabulary of Agency and Immersion in Interactive Narratives." *Interactive Storytelling: 6th International Conference, ICIDS 2013, Istanbul, Turkey, November 6–9, 2013: Proceedings*, edited by Hartmut Koenitz, Tongus Ibrahim Sezen, Gabriele Ferri, Mads Haahr, Digdem Sezen, and Güven Çatak, Springer, 2013, pp. 25–34. (International Conference on Interactive Digital Storytelling.)

- Morie, Jacquelyn Ford. "Performing in (Virtual) Spaces: Embodiment and Being in Virtual Environments." *International Journal of Performance Arts and Digital Media*, vol. 3, no. 2–3, 2007, pp. 123–38.
- Mulvey, Laura. "Visual Pleasure and Narrative Cinema." *Screen*, vol. 16, no. 3, Autumn 1975, pp. 6–18.
- Murray, Janet H. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Updated ed., MIT P, 2017.
- Rambarran, Shara. *Virtual Music: Sound, Music, and Image in the Digital Era*. Bloomsbury Academic, 2021.
- Rivas Méndez, David, Calum Armstrong, Jessica Stubbs, Mirek Stiles, and Gavin Kearney. "Practical Recording Techniques for Music Production with Six-Degrees of Freedom Virtual Reality," Conference paper, Audio Engineering Society, New York, 7 Oct. 2018. AES.org, <https://www.aes.org/e-lib/browse.cfm?elib=19729>. Accessed 17 Aug. 2023.
- Ryan, Marie-Laure. *Narrative as Virtual Reality: Immersion and Interactivity in Literature and Electronic Media*. Johns Hopkins UP, 2001.
- Smalley, Denis. "Space-Form and the Acousmatic Image." *Organised Sound*, vol. 12, no. 1, Apr. 2007, pp. 35–58.
- . "Spectromorphology: Explaining Sound-Shapes." *Organised Sound*, vol. 2, no. 2, Aug. 1997, pp. 107–26.
- Vernallis, Carol. *Experiencing Music Video: Aesthetics and Cultural Context*. Columbia UP, 2004.
- . "Music Video, Songs, Sound: Experience, Technique and Emotion in *Eternal Sunshine of the Spotless Mind*." *Screen*, vol. 49, no. 3, Autumn 2008, pp. 277–97.
- Vernallis, Carol, Amy Herzog, and John Richardson, editors. *The Oxford Handbook of Sound and Image in Digital Media*. Oxford UP, 2013.
- Walther-Hansen, Mads. "Sound Events, Spatiality and Diegesis—The Creation of Sonic Narratives in Music Productions." *Danish Musicology Online*, Special Edition—Sound and Music Production, 2015, pp. 29–46, https://www.danishmusicologyonline.dk/arkiv/arkiv_dmo/dmo_saernummer_2015/dmo_saernummer_2015_lyd_musikproduktion_02.pdf. Accessed 17 Aug. 2023.
- Winters, Ben. "The Non-Diegetic Fallacy: Film, Music, and Narrative Space." *Music & Letters*, vol. 91, no. 2, May 2010, pp. 224–44.
- Wood, Aylish. "Timespaces in Spectacular Cinema: Crossing the Great Divide of Spectacle versus Narrative." *Screen*, vol. 43, no. 4, Winter 2002, pp. 370–86.