14

Enacting Virtual Boundaries: Music Video and the Changing Technological Landscape

Eugy Han and Saul Ivan Quintero

Virtual reality (VR) has been regarded as a promising tool across multiple disciplines due to the breadth of its applicability. From conducting scalable studies in psychology, to taking students on field trips to impossible-to-reach locations, to training employees on how to handle dangerous scenarios without putting them in risk, VR has the potential to be used in exceptionally diverse ways. The unique affordances of VR allow creators to build experiences that make users feel immersed, present, and embodied in a mediated environment. Beyond translating activities that are viable in the physical world, VR allows us to upend rigid definitions of what is possible in the physical world and affords us the ability to expressively curate imaginative experiences conceptually limited only by our own inventiveness.

One of the most prominent cases of this includes the extraordinary ways in which short-duration audiovisual media such as the music video has qualitatively transformed in how it is engaged with and how the media itself is structured with the advent of consumer accessibility to VR media. While lengthy media such as gaming or feature-length motion films have exploited the unique spatiotemporal affordances of VR, little attention has been granted to how immersive media such as VR or Augmented Reality (AR) are influencing short media production. The spatial expanse of VR effectively transmutes the perceptual dimensions of music videos, for instance, by positioning viewers within the video itself taking up the entirety of their visual and auditory receptive fields. VR can also engender new possible audiovisual experiences altogether for these types of short media forms. In this chapter, we argue that the unique spatiotemporal affordances of VR have shifted the way media artists, in particular those creating shorter productions such as music videos, conceptualize audiovisual relations, and multisensory

perceptual storytelling in general. We first start with an abridged history of shifts in audio-visual relations in music videos throughout the advent of volumetric capture and head-mounted-display commercialization. Following this, we discuss the ways in which the popularity of VR and its multisensory channels suggest a cultural penchant for increasingly participatory and immersive experiences. Finally, we consider how VR categorically alters the nature and structure of the traditional music video to form a novel arena for perceptual play.

Music Video as Experienced through VR: A Brief History

It is too early to conceptualize music videos presented on a traditional planar screen as a thing of the past—nevertheless, the conventional architecture of a music video has been silently transforming along with new media technology in recent history (for examples of new types of music videos and discussions of early VR in music videos, see Korsgaard 2013). In the late 1990s, the alternative rock band Ten Speed Racer created one of the first "360 degree" music videos. Although the video is "immersive" on its surface, it structurally feels analogous to panoramic photography in video format which lacks the binocular depth cues, such as stereopsis, that contemporary VR media includes to mimic visual depth perception. Commercial head-mounted displays (HMDs) had not emerged as a consumer staple at the time of the video's release. Because streaming platforms such as YouTube and Vimeo were also not yet optimized for immersive video streaming, early reproductions of this music video and others like it could be accessed mainly in a still, flat image screen restricted format where the full panorama is visible albeit not in a way that captures the intended immersiveness of the expanded user field of view. Volumetric capture (a technique used to capture 3D space) grew in popularity in the late 2000s with the advancement of light field cameras (a camera that captures information about light in a space), structured light scanning (a camera that uses light to capture information about a 3D object), LiDAR scanning (a scanner that uses light to determine the distance between two objects; ultimately used to create 3D maps), and photogrammetry (the process of extracting information about objects and space through 2D images). With the advancement of being able to increase the degrees of freedom of the user and how their body is able to engage with media artifacts, demand for such media increased.

Early on, YouTube facilitated the availability of these kinds of immersive forms of pre-figurative VR media engagement. 360-degree videos, although intended to be experienced using HMDs, can be experienced by users even in the traditional 2D format by simply clicking and dragging the screen in order to move around the expanded perceptual environment (Figure 14.1). However, such viewing of videos falls short of the immersiveness of the HMD experience, rendering it more of a restricted interactive music video. Moreover, in the case of audiovisual content like music videos, because the responsibility of exploration is prescribed to the viewer, the typical alignment between temporal cuts in the visuals and shifts in sounds as a tool for guiding momentum in the video is ultimately restricted. A viewer must consider the explore/exploit tradeoff

Figure 14.1 Still images from the "Stonemilker" (2015) music video by Björk. The two images are of the same moment observed from two different points of view manipulated by dragging the screen with the mouse.

between maintaining a familiar but fixed viewpoint or shifting to oblique, unfamiliar vantage points. Consequently, the viewer takes on a greater cognitive load and attention is siphoned away from the content itself.

Although these avenues of media engagement are restricted relative to currently available devices or formats, they nonetheless exploded in popularity in the early 2010s with popular musicians creating their own videos, including Avicii, Björk, Fort Minor, and Foals, to name a few.

HMD-facilitated VR represents a categorical shift in multisensory media and availability of HMDs has only increased over time, invoking fervent discourse on the new frontiers of audiovisual media (see Nordahl and Nilsson 2014, who discuss sound perception and production and how they manifest in presence within VR, or Findlay-Walsh 2021, who describes listening as a process of environmental "inhabiting" and considers the "auditory experience as the real-time construction of reality" (72)). HMDs allow for the viewer to be "inside" the video-content itself. Beyond passively viewing or engaging in a somewhat limited manner with the video, the viewer can navigate around a 3D environment, embody the character in a narrative, or interact with objects and their surroundings.

HMD music videos released in early years tend to neglect audiovisual temporal dynamics in service of exploiting the unique spatial affordances of HMDs. Many consist of fixed voyeuristic vantage points for audience members which expand the field of view giving the audience some degree of agency over where they chose to rest their gaze. These videos have few visual cuts or shifts and these visual shifts are only rarely aligned with some musical shift. Take the music video for the song "Mountain at my Gates" (2015) by Foals, for example. There are no actual visual cuts in the entire duration of the video—something that is rather uncommon in traditional music videos. Another example of this penchant is visible in early fully animated VR music videos such as the video created by visual artist TAS for the dubstep track "Odyssea (Grouch Remix)" by Olie Bassweight (2017). Here, there are also no visual cuts as the generative-like psychedelic visuals move on their own with the observer vantage point being forced to train along, merely *witnessing* the changing environment. Ostensibly, one reason for this interest in spatial dynamics is a result of the enhanced visual affordances of VR. Whereas 2D content is bound to the edges of a phone or computer screen, an HMD quite literally takes up the viewer's entire visual field. In 2D-based experiences, there are a plethora of exteroceptive distractions from the physical world, such as a notification popping up on a screen, or the light

flickering above, or a person walking past the window. Such visual distractions are shut out once an HMD is on. This allows the artist to surgically control the visual experience from start to finish without allowing other visual artifacts to distract the viewer. 2D music videos, while having great potential for eliciting intense emotional participation, are restricted in how perceptually interactive they can be. It makes sense that, as VR first expanded the visual spatial field, VR music videos simply represented a superstimulus of traditional music video conventions.

Despite being 360-degrees, presented through an HMD, and incorporating some degree of interactivity, these videos have several limitations relative to contemporary VR immersion techniques. Because of the fixed vantage point, they lack representation of head motion visual parallax (Thatte and Girod 2021). As a result, there is a visual-vestibular sensory conflict since our heads typically move congruent with changing visual images. The sensory conflict is visually noxious and can elicit nausea, and generally degrades the sense of presence that a viewer experiences in the virtual space.

Reshaping Sight and Sound: VR Redefining the Senses

In VR, the artist has immense control over the rules governing the senses. Because the viewer is completely immersed in the artist's world, the artist can set their own rules on what "seeing" entails. "Notes on Blindness" (2016, dir. Middleton) is one such example in which "seeing" is redefined. This VR experience takes the viewer through a sensory and psychological experience of a man who has lost his sight. By combining binaural audio and playing with light, the creators take the viewer to a "visual" world beyond sight. Here, the conventional rules of "seeing" are redefined by the artist (i.e., concrete colors and shapes are replaced by light). Similarly, "6x9" (2016) by *The Guardian* (UK) is a VR experience in which a viewer is placed into a US solitary confinement prison cell. The experience guides the viewer through what kind of psychological damage can be caused by being in extreme isolation for a prolonged period of time and induces feelings of claustrophobia. In both examples, the artist fixes the perspective from which the viewer experiences not only the narrative, but also how they literally see and feel about the world. In VR, not only can the artist guide the viewer to *what* they want the viewer to see, they can also guide the viewer to *how* they want the viewer to see. It is curious to imagine how

classic music videos would translate into the context of this surgical control over an expanded and mediated visual narrative. In the music video for the song "One" by American metal band Metallica, clips from the film *Johnny Got His Gun* (1971) vacillate with cold imagery of the band performing their morose but gritty anti-exclamations toward war and the psychological effects of war on veterans. Certainly the aesthetics of this video can be at least somewhat attributed to the technological constraints. Thereby, how would a video such as this translate to a VR experience? From the editing techniques to how the image and sounds intermix, how would the topic of intimate psychological health be explored in the context of metal and pre-recorded visual sampling techniques?

Sound in VR experiences has received considerably less attention, at least in the realm of research. However, it plays just as important of a role on the viewer's experience. Typical music engagement takes place using headphones which offer unique spatial conditions for listening. As Charles Stankievich discusses in his essay on interiority and headphone listening: "A modern technological prosthetic, headphones are quite literally a bracketing of the world for a precise analysis of sound" (Stankievech 2007: 5). Jacob Downs also entertains this conception of a unique spatiality to headphone listening in his essay exploring how headphones "territorialize" space for listeners, referring to listening as "homely," even "wombic" (Downs 2021: n.p.). In traditional music video watching, the spatiality offered by the restricted visual screen is incongruent with the expanded, intimate spatiality offered by headphones. This incongruence, while not necessarily interfering with the narrative exploration through the alignment of visual temporal changes with auditory changes, communicates a spatial distance between the music space and the visuals. Visual information may be especially preferred by our brains to inform our conception of space and our body's relationship to space, or our positionality (McGovern 2016). VR-based music videos turn the shortcomings of 2D screen media access on their head by expanding the visual field and making ambisonic and visual fields more spatially congruent, superimposing and conjoining them to one another.

While typical HMDs do not include headphones, they employ techniques such as head-related transfer function (HRTF) filtering that are designed to spatialize audio in a methodologically sophisticated manner. HRTF filtering involves modifying sounds to mimic how the ears receive and perceive sounds from particular locations in space. The brain relies on binaural cues such as the inter-aural time and intensity differences of perceived sounds to estimate the location of sounds. In other words, by tracking how long it took for a sound to

arrive at a particular ear compared to the other or how loud or intense the sound arrived at one ear versus the other, this allows the inner ear bone, along with the pinna and the brain to detect the angular distance of the sound and where along the azimuth and altitude axes it originates from. The brain tracks what are called Head-Related Impulse Responses (HRIR) which constitute the auditory signals being processed by the auditory system as a function of time. The HRTF is a Fourier transform of the HRIR response curves essentially determining the perceived location by the impulse frequency patterns over time. Both generic and individualized HRTFs exist and can be implemented for realistic spatialized audio. Because factors including the size, shape, and density of the head, ear, ear canal, etc. alter how sounds are sensed, due to some frequencies being consequently attenuated while others are amplified, each individual has optimal individualized filtering that can be conducted to enhance the auditory experience (Berger et al. 2018). In VR, generic and/or individualized HRTFs are based on the perceived visual architecture of the virtual space. Extending Jacob Down's metaphor, HMDs transpose imagined audiovisual spaces into enlarged, visual fields and territorialized auditory topographies.

Although long-form music-video adjacent VR experiences exist, such as the VR concert, the short VR music video faces entirely new possibilities and potential challenges. If a music video director aims to create a music video experience as immersive as long-form media, then they may have to convince the brain that the observer is indeed immersed in the audiovisual experience with a restricted amount of time. Consequently, we see a proclivity toward maximalist aesthetics or hyper-stimulation in some VR music videos with the high number of sensory events allowing for sensory evidence to accumulate and inform the brain's estimation of the causal relationship between the sensory signals in the music video and the observer's body. Some examples of the various VR music videos that incorporate similar aesthetics include Mac Demarco's "This Old Dog" (2018), Muse's "Revolt" (2017), and Marshmello & Crankdat's "Falling to Pieces" (2020). Certainly not all VR music videos necessarily incorporate an inflated amount of sensory events. Some VR music videos incorporate minimal aesthetics that nonetheless can offer sensory evidence of observer presence in other ways. For instance, in The Poppy VR180 experience by pop-rock singer Poppy, the singer simply sits in front of the observer and calmly gazes back "acknowledging" the viewer's presence by gazing toward the camera.

Music video producers do not necessarily only seek a glorified sense of presence and it would certainly be an oversimplification to suggest that. In fact,

traditional music videos often employ incongruence between imagery and sound content through attentional lures, spatiotemporal onset discrepancies, etc. It is important to reflect on and consider how these classic techniques ubiquitous in non-VR music videos may translate or be expelled in the new technological landscape for media engagement. Furthermore, it is important to highlight the prominence of the spatial dimension of VR experiences in altering short-form media and MV engagement behaviors.

Haptics, Interactivity, and Virtual Presence

Beyond audiovisual elements, VR-based music videos can incorporate other forms of sensory feedback that amplify the ability of a music video to affect the user's body. Touch, or haptic feedback, is one powerful example of this novel approach to production. Many HMDs come with hand controllers. Not only do the hand controllers and their various buttons and triggers act as fingers, taking in users' input to navigate the virtual world, they also serve as sources of output. Several applications emit vibrations to enhance the user experience—such as making the user feel as though they are using a real, vibrating chainsaw when cutting down a tree (Gorlick 2011)—or alerting the user that something is happening—such as sending off a vibration as a warning that a user's avatar is violating another avatar's space. Though quite simple, such haptic feedback can serve unique purposes in music videos that are not afforded via their traditional 2D form. There are a range of types and patterns of vibrations—from soft to harsh, or constant to pulses, or relaxing to startling—that can be combined to add to the visual and audio experience. The music video can now, beyond delivering light to the eyes and vibrations to the inner ear bones, make mechanical impressions directly onto the body and the skin. This effectively expands the space in which a music video interacts with viewers from their head to the rest of the body. Imagine listening to a punk singer's rampageous screams while feeling the tremor of their voice through intense vibrations in your hands.

Hand controller vibrations are only the start to what haptic feedback can look like in VR-based music videos. As new hand-tracking technologies—which may allow users to be controller-free—and new accessories—such as haptic feedback gloves or suits—are beginning to emerge, the near future of VR may involve drastic changes. With advances in technology, the types of

feedback that artists can incorporate become more sophisticated. The physical experience can be textured with sensations of pressure, weight, heat, or even pain. Viewers can feel their body floating through the sky, or a weight pressing down on their chest. Sensations of heat or coolness can be stimulated throughout parts of the body—perhaps even both at the same time. Lastly, the haptic sensors can deliver punches or caresses, or maybe something more extreme, like gunshots.

Some evidence suggests that VR-based media can more strongly induce affective states than 2D screen-based renditions (Pallavicini and Bouchard 2019). This increased potential for eliciting affective responses can ostensibly be attributed to an enhanced sense of virtual presence (Diemer et al. 2015). Aside from the expanded sensorium that virtual reality supplies, this is a major point of interest for music video directors aiming to potentiate their creations. What kinds of factors influence virtual presence? Scholars have extensively discussed this topic, with a recent paper suggesting that former work by presence researchers has been biased toward the feeling of the physical body being located in virtual space, as "being there" (Weber et al. 2021). While "presence" is an all-encompassing term that describes the subjective experience of "being there," it can be further divided into subdimensions, such as social, telepresence, ambient, spatial, and self-presence. Particularly relevant here is spatial presence, that is, the extent to which one feels present in the mediated environment, rather than the immediate physical environment where the body is located (Biocca 1997). Telepresence is also relevant here, referring to the extent to which one feels that they are co-existing with another agent or object at the same *time* and not necessarily in the same space. A simple example of this includes speaking on the phone with a loved one that lives across the other side of the world. A voicemail does not imply the same degree of presence as the spontaneous, accessible voice of a loved one. An increased perceptual accessibility suggests greater potential for affecting and being affected by the immediate environment. Moreover, if an observer appraises that they are being *emotionally* affected by virtual events, they will report increasingly higher degrees of presence. A study found that when offering fake heart rate variability feedback to observers in order to deceive them into thinking they were being emotionally aroused by virtual events, observers felt increasingly present in the virtual environment (Weibel et al. 2011).

Others have argued that the sense of "being there" is only a fragment of the sense of presence. For instance, recent VR music videos now include a

Figure 14.2 Still images from Tyler Hurd's VR experience for "Old Friends" (2016), a track by Future Islands.

component of interactivity whereby observers have agency over what occurs in the video. Bonobo's 2017 music video for their song "Outlier" allows observers to reach out and bend a visual stream of beams that flows to the rhythm of the music, creating a tri-modal chiasm of engagement. Other music videos, such as VR creator Tyler Hurd's creation for the song "Old Friend" (2016) by the band Future Islands (Figure 14.2), have upped the interactive factor to where observers embody and control not only perceptual information but also a dancing avatar that interacts with virtual characters in the music video narrative.

The observer is able to not only choose actions for the avatar, but also engage and enact those actions themselves by moving their own body and simulating the desired gestures. The actions of this user-controlled avatar consequently affect the progression of the narrative itself, rendering the observer an active participant in the curation of their own VR music video experience.

As Weber and colleagues discuss in their theoretical review on perceived realism and presence in VR, it is not *necessary* that the depicted narrative actually displays a string of events possible outside of virtual space for VR to feel immersive and present. Instead, artists and VR music video directors can strategically select what perceptual or agential cues they want to embed within their environments in order to invoke feelings of presence to a preferred degree. The mediated virtual environment, along with the perceptual and agential sensitivities of an observer, can be thought of as co-constructing virtual presence.

VR MV Ethical Considerations

As a final commentary, it is also critical to note that these infinite possibilities raise ethical concerns. From the user's perspective, as Laura Marks writes in her work on haptic visuality, a newcomer into a space controlled and curated by another must decide whether to keep their own "bodily habits" and knowledge about their body at the risk of feeling out of place in a new environment or shed their habits in order to acquiesce to the available modes of operating (Marks 2000: 209). Theoretical frameworks such as the Transformed Social Interaction offer researchers and practitioners an understanding of how changes, both drastic and subtle, can influence the way people perceive themselves and others around them, and ultimately social behaviors (Bailenson et al. 2004). While VR systems most often strive to render a veridical representation of the physical world, they can also systematically track and filter and transform signals, amplifying and suppressing certain features. Such perceptual and social transformations can have a lasting impact that extend beyond the virtual environment. Future VR work must contend with the consequences of strategic curation of perceptual and agential cues as they may imply normative forms of presence.

From the creators' point of view, the increasingly high-tech nature of VR produces roadblocks in accessibility to certain modes of production. Classic VR techniques including volumetric capture are exorbitantly expensive and incur sizable production costs that can be significantly greater than non-VR productions. The complex multisensory VR interfaces discussed exacerbate these costs, generating accessibility barriers for music video directors and especially independent artists. It is important not to overemphasize technological aesthetics and to consider the potential consequences of calibrating normative standards of evaluating media based on "technological advancements" alone. Ultimately, we must be privy to how VR influences what we value in our cultural artifacts and how that may disproportionately harm or benefit particular members of our communities.

Literature

Bailenson, Jeremy N., Andrew C. Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. "Transformed Social Interaction: Decoupling Representation from Behavior and Form in Collaborative Virtual Environments." *Presence: Teleoperators & Virtual Environments* 13.4 (2004): 428–41.

Berger, Christopher C., Mar Gonzalez-Franco, Ana Tajadura-Jiménez, Dinei Florencio, and Zhengyou Zhang. "Generic HRTFs May Be Good Enough in Virtual Reality. Improving Source Localization through Cross-Modal Plasticity." *Frontiers in Neuroscience* 12 (2018): 1–9.

Biocca, F. "The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments." *Journal of Computer-Mediated Communication* 3.2 (1997): n.p.

Diemer, Julia, Georg W. Alpers, Henrik M. Peperkorn, Youssef Shiban, and Andreas Mühlberger. "The Impact of Perception and Presence on Emotional Reactions: A Review of Research in Virtual Reality." *Frontiers in psychology* 6 (2015): 26.

Downs, Jacob Kingsbury. "Acoustic Territories of the Body: Headphone Listening, Embodied Space, and the Phenomenology of Sonic Homeliness." *Journal of Sonic Studies* 21 (2021): n.p.

Findlay-Walsh, Iain. "Virtual Auditory Reality: Inhabiting Digital Pop Music as Simulated Space." *SoundEffects* 10.1 (2021): 71–90.

Gorlick, Adam. "New Virtual Reality Research—and a New Lab—at Stanford." Stanford Report, April 8, 2011, at https://www.stanfordvr.com/mm/2011/stanford-new-vr-research.pdf (accessed January 16, 2023).

Korsgaard, Mathias Bonde. "Music Video Transformed." In *The Oxford Handbook of New Audiovisual Aesthetics*. Ed. John Richardson, Claudia Gorbman, and Carol Vernallis, 501–21. Oxford: Oxford University Press, 2013.

Marks, Laura U. *The Skin of the Film: Intercultural Cinema, Embodiment, and the Senses*. Durham and London: Duke University Press, 2000.

McGovern, David P., Andrew T. Astle, Sarah L. Clavin, and Fiona N. Newell. "Task-Specific Transfer of Perceptual Learning Across Sensory Modalities." *Current Biology* 26.1 (2016): R20–21.

Nordahl, Rolf, and Niels C. Nilsson. "The Sound of Being There: Presence and Interactive Audio in Immersive Virtual Reality." In *The Oxford Handbook of Interactive Audio*. Ed. Kasey Collins, Bill Kapralos, and Holly Tessler, 213–33. Oxford: Oxford University Press, 2014.

Pallavicini, Federica, and Stéphane Bouchard. "Assessing the Therapeutic Uses and Effectiveness of Virtual Reality, Augmented Reality and Video Games for Emotion Regulation and Stress Management." *Frontiers in Psychology* 10 (2019): 2763.

Stankievech, Charles. "Headphones, Epoche, and L'extimité: A Phenomenology of Interiority." *Offscreen* 11.8/9 (2007): 1–11.

Thatte, Jayant, and Bernd Girod. "Real-World Virtual Reality With Head-Motion Parallax." *IEEE Computer Graphics and Applications*, 41.4 (2021): 29–39.

Weber, Stefab, David Weibel, and Fred W. Mast. "How to Get There When You Are There Already? Defining Presence in Virtual Reality and the Importance of Perceived Realism." *Frontiers in Psychology* 12 (2021): 628298.

Weibel, David, Bartholomäus Wissmath, and Fred W. Mast. "The Role of Cognitive Appraisal in Media-Induced Presence and Emotions." *Cognition & Emotion* 25.7 (2011): 1291–8.