A Focal Press Book

ARTIFICIAL INTELLIGENCE AND MUSIC ECOSYSTEM

Edited by Martin Clancy

Routledge

7 Extended Reality

Music in Immersive XR Environments: The Possibilities (and Approaches) for (AI) Music in Immersive XR Environments

Gareth W. Young and Aljosa Smolic

Introduction

Cross reality (XR) is defined as:

> The union between ubiquitous sensor/actuator networks and shared online virtual worlds – a place where collective human perception meets the machines' view of pervasive computing.
>
> (Paradiso & Landay, 2009: p. 1)

XR encompasses multiple software and hardware technologies that enable content creation for augmented reality (AR), cinematic reality (CR), augmented virtuality (AV), and virtual reality (VR). Research on XR technology often concerns itself with the human-computer interface (Jerald, 2015), focusing on presenting and quantifying realistic, immersive, interactive, simulated worlds. However, for XR technology to achieve higher impactful prominence as a platform for artistic expressions, it is also necessary to explore the fundamentals of interface technology design for content creation and performance-style issues. With many industry-focused advances being made in information and communication technology, computing, big data analytics, machine vision, and artificial intelligence (AI), immersive technologies will be enhanced for cyber-physical and social experiences (Bastug et al., 2017).

AI has drastically changed how we view art and creativity in the twenty-first century, fundamentally augmenting the very nature of contemporary creative practices (López de Mántaras, 2016). Although computers have been used as a tool for artistic expression for many decades (AIArtists, 2019), with the advent of real-time machine intelligence, high-speed computers, and computational thinking, artists have developed numerous creative applications for high-profile audiovisual art exhibitions (MoMA, 2017). AI applications that create both abstract and representational art have advanced considerably with the discovery of generative

DOI: 10.4324/9780429356797-8

adversarial networks (GANs), giving rise to the field of computational creativity (Veale & Cardoso, 2019). In the music domain, computational creativity has focused on generating musical scores for human musicians, musical improvisation and accompaniment, and software that creates music in an artist's particular style or musical genre. This attention also includes the further development of specific musical applications for algorithmic music composition (Zaripov, 1960; Herremans, 2016), musical pattern recognition (Kane, 2016), automated accompaniment systems (Vercoe, 1984; Dannenberg, 1984), solo performances (Vercoe & Puckette, 1985), interactive scores (Robertson & Plumbley, 2006), and the ultimate recognition as a stand-alone creative AI artist (Zulić, 2019).

The possibilities and approaches for AI music in immersive XR environments for media entertainment can be much more than immersive gaming, just as film entertainment is more than blockbuster cinema. Contemporary XR productions offer an expansive array of entertainment value to existing and emergent three-dimensional (3D) media practices and beyond. However, the current technological unification of AI and XR arguably adds value over more traditional multimedia practices to supplement the main creative focus by making full use of the spatiality afforded with six degrees of freedom (6-DoF) experiences and implementing more engaging interactions and player agency for music, cinema, and gaming. By exploring the combined use of these technologies in creative practice, objects of attention within an immersive virtual environment (IVE) can be presented so that audiences can move around, interact, and engage with materials, making the overall experience fundamentally unique. Although potentially highly immersive and engaging, this also brings to light issues concerning conventional performance practice in IVEs and how artists can engage audiences remotely with artistic expressions remaining intact and as intended.

Within the XR domain, VR technology is currently experiencing a resurgence in commercial interest. Therefore, the barriers of VR for use in the entertainment industry are changing (Evans, 2018). *VR* can be broadly defined as 'a computer-generated digital environment that can be experienced and interacted with as if that environment were real' (Jerald, 2015: p. 9). Jaron Lanier first coined the technical term *VR* in the 1980s (Kelly et al., 1989). Today, Lanier proposes several alternative definitions, including a reference to art practice: 'a twenty-first-century art form that will weave together the three great twentieth-century arts: cinema, jazz, programming' (Lanier, 2017: p. 3).

Via XR technologies like VR, musicians can engage audiences by presenting spatialised sonic materials employing multimodal stimulation, which the viewer can interpret through their unique background and lived experiences. Thus producing a qualitatively different knowledge of traditional performance practices within an unaccompanied or shared space (Parker & Saker, 2020). Potentially, through the use of AI and XR combined, audiences and performers can actively participate in an immersive performance experience, contributing their unique subjective views within specific societal contexts (Cappello, 2019). This level of engagement can be delivered in VR via the creative use of immersion, interaction, and imagination (Burdea & Coiffet, 2003) in a way that provides new insight into creative practice and how audiences can be engaged by artists (O'Dwyer et al., 2020). AI will no doubt play some role in future XR music practice. Nevertheless, as was highlighted by Lanier (1988):

> In order for computer art or music to work, you have to be extra careful to put people and human contact at the centre of attention.

This chapter explores XR technology as an active stage for the spatial presentation of musical performance and artistic imagery, balancing user interaction, immersion, and imaginative storytelling via unique 3D audiovisual musical performance practices. We document the history of the combined audiovisual approach, highlighting the significant technological techniques as they were developed, and discuss the implementation of new strategies for the potential combined use of XR and AI in music, live performance, and immersive scenographies.

3D, XR, and Music

Technological evolution in modern engineering and its application in artistic practice have grown hand in hand for many years, with several unique audiovisual entertainment platforms and computational epochs linking music with visual media and thus XR. While music performance has a long and varied history, we focus on the advent of contemporary audiovisual technology and its application in capturing and reproducing music performance and the music video as a standalone modern art form.

In the 1950s, advent XR technologies were proposed to present audiences with a multimodal 'experience theater' (Robinett, 1994). This work would later create much smaller viewing screens that could be placed closer to the viewer's eyes and accompanied by crudely spatialised stereo sound. This technology was also capable of stimulating other sensory modalities, such as touch, taste, and smell. By 1962, experience theatre concepts were fully matured, and the available multimodal media technologies were also stable enough for the application.

Research and development in the field flourished. As a result, the Sensorama was created, which is still lauded today as an early prototype XR platform (Heilig, 1962; Robinett, 1994). The Sensorama was a patented platform for solo experiences and was mechanical (see Figure 7.1a). This multimodal immersive media viewer included a stereoscopic colour display, fans, odour emitters, stereo sound, and a moving chair. The Sensorama films were visceral, aptly titled *Motorcycle*, *Belly Dancer*, *Dune Buggy*, *Helicopter*, *A Date with Sabina*, and *I'm a Coca-Cola Bottle*, and could be viewed in especially purposed cinema for communal viewing.

The first head-mounted visual display system in immersive, computer-generated simulation applications can be seen in Figure 7.1(b) (Sutherland, 1968). This *Ultimate Display* was a device that could simulate wireframe objects and rooms, ultimately simulating reality to the point of perceived reality. The device presented individual users with a virtual world, viewed through a head-mounted display

(a)

(b)

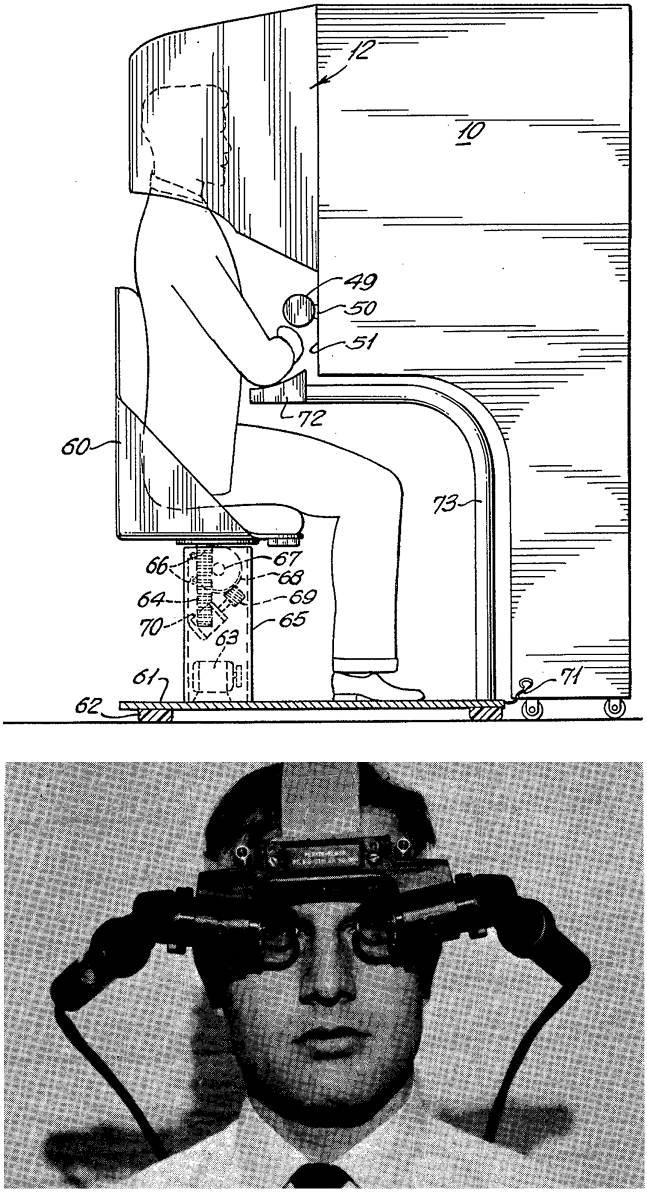

Figure 7.1 (a) The Sensorama (Heilig, 1962) and (b) early head-mounted display optics (Sutherland, 1968).

(HMD), and stimulated the user via 3D sound and tactile feedback. Powering this HMD was the *Sword of Damocles*, a prototype computer that used advanced hardware to create this virtual world in real time, allowing users to interact with objects in the digital world. This multimodal media platform and HMD arrangement have become a core design for contemporary XR today.

Stereoscopic 3D (S3D), a display technology that creates a depth impression by projecting two different views into the users' eyes, was also being used around this time to expand the viewer's level of immersion beyond that of the flat-screen 2D experience of classical cinema and TV described thus far. Although this principle has been known since the 1850s, the technology used has seen several trends ebb and flow in the media entertainment industry over the years. One such wave occurred in the 1950s when projection systems based upon anaglyph (red-green) glasses were popularised in public cinemas for exclusively produced 3D movies. However, in the past century, S3D has continuously appeared and disappeared from mainstream entertainment after short-lived spurts of popularity, primarily due to the technical limitations of the time providing only poor visual quality and leading to health issues, such as fatigue or motion sickness. Nevertheless, S3D has remained a notable, if relatively a niche, 3D entertainment technology seen in the home, theme parks, and other media-specific venues.

In the eighties, the term *virtual reality* was once again on the vanguard of technological innovation with the creation and rise of VPL Research (Lanier, 2017). At this point, television had become ubiquitous, and most households own at least one TV set today. However, VR hardware was more expensive and mainly provided simulation devices for research into medical training, flight simulation, CAD design, and military purposes, primarily losing interest in home entertainment markets. However, by 1988 the Cyberspace Project by Autodesk had achieved the real-time rendering of 3D environments on smaller personal computers, and interactive VR media, via console gaming, was on the rise.

In the nineties, home console gaming platforms were becoming more popular, and interest in the commercial release of VR for gaming slowly grew. At this time, Sega developed the Sega VR system, and Nintendo released the Virtual Boy, ensuring that VR remained well regarded in the high-end gaming scene. Notably, at the 1992 SIGGRAPH conference, VR and music were presented together in public via a musical performance from Jaron Lanier, demonstrating how music could be played in VR using a DataGlove device and performed in front of a live audience (Lanier, 2017). Unfortunately, many commercial VR projects failed to grow much past this date due to the incompatibility of computer processing and graphics rendering. Media technology research shifted focus to developing the World Wide Web. This gentle demise was then followed by a long period of indifference to commercial VR technologies in general.

During this time, a notable application of S3D for music entertainment was released – the 3D science fiction experience *Captain EO*, which featured Michael Jackson. *Captain EO* was shown at Disney theme parks from 1986 through 1998 (with revivals from 2010 onwards), written by George Lucas and directed by Francis Ford Coppola (Jackson et al., 1986). In these popular 'immersive rides,' audiences sat in traditional theatrical configurations and wore 3D glasses to experience filmed events collectively. In addition to providing 3D depth perception, the experience also included in-theatre effects, such as lasers and smoke, that were synchronised to the filmed events. With the audience experiencing a sense of embodiment within the narrative, they became a more integral component inside the overall story design (Parker-Starbuck, 2011). As such, Disney productions like *Captain EO* and *Honey, I Shrunk the Audience* were explicitly created to provide theme park audiences with collective immersive experiences. However, S3D could not support 360-degree look-around (3DoF) or move-around (6DoF) features of later XR-technology-mediated experiences.

Music video practices were also thrust into prominence with the launch of the television channel MTV, which based much of its content around the medium for over twenty years (Tannenbaum & Marks, 2011). The format of MTV's musical content changed considerably over its lifetime. However, it mainly included introductions and voice-overs from video jockeys (VJs), music news, interviews, concerts, promotions, and a particular focus on creative visual representations of musical narratives. From 2000 onwards, conventional outlets for promoting and viewing music videos dropped, playing fewer music videos due to the rise of social media and video-sharing websites (NPR, 2008). With the success of peer-to-peer music-sharing services and online video-sharing platforms in the late 2000s, artists could directly engage with fans, controlling the viewing and listening experience of the consumer (Burgess & Green, 2018). Moreover, the music video format was also changing; vertical videos were being made to view in mobile phone platforms (Jaekel, 2015). This shift shows us today that as new technology presented itself for the consumption of audiovisual materials, musicians and creative technologists were historically adept at adapting, and the industry continues to innovate and explore emergent platform-specific outputs to this day.

In the late 2000s, the industry saw another boom in S3D technology for cinema and home entertainment applications. The digital production and projection technologies of the time were matured enough, with many of the limitations of earlier systems alleviated. Such was the sustained popularity of S3D; it remains popular in 4D cinema entertainment today, while 3DTV, on the other hand, did not survive these trials. One notable musical example in this new era of S3D was the movie *U2 3D*, released in 2007. *U2 3D* was a 3D concert film featuring the band U2 performing during their Vertigo Tour in 2006 (U2 et al., 2007). Spectators of such footage were given a bodily, up-close presence onstage with the band, along with the added novelty provided by stereoscopic 3D vision.

Although commercial VR technology has been around since the mid-eighties and has endured many setbacks (Lanier, 2017), VR is currently undergoing yet another renaissance (Evans, 2018). However, VR has taken a long time to become affordable and provide a positive user experience. The slow rise back into the limelight arguably started in 2010, when the prototype of the Oculus Rift was first developed. Facebook's investment into VR in 2016 led to a resurgence in the technology, culture, and consumption of VR, and the industry has been growing ever since. The VR market is increasing, with the size of consumer-grade equipment projected to increase from 6.2 billion US dollars in 2019 to more than 16 billion US dollars by 2022. Additionally, unit shipments of VR devices are expected to reach 12.5 million in 2020 (Statista, 2020). Contemporary applications of VR encompass computer-generated environments that simulate the physical presence of people and objects and provide realistic sensory experiences (Jerald, 2015).

As such, animated, 360-degree video productions have become a popular way to allow the viewer to experience 'fly-on-the-wall' musical performances within immersive, imaginary worlds with three degrees of freedom (3DoF) – referring to the tracking of rotational motion as pitch, yaw, and roll. For example, the band Gorillaz (2017) openly released *'Saturnz Barz (Spirit House)'* as both a traditional 2D music video and a visually explorable, animated, 360-degree 3DoF video experience on YouTube (see Figure 7.2). Similarly, Mac Demarco (2017) released a 2D and 3DoF music video for '*This Old Dog*.' Additionally, Demarco released a VR experience for separate download that afforded the viewing audience further freedom of movement within the virtual world beyond the 2D viewing portal provided by traditional PC monitors. The release of stand-alone VR experiences is becoming more familiar with studios releasing animated immersive music videos to launch new albums and capture the imagination of new audiences. For example, Squarepusher's *'Stor Eiglass'* (2015) was used to celebrate the *Damogen Furies* album.

For more explorative experiences, immersive music videos can also facilitate the viewer in the exploration of computer-generated imagery (CGI), from within the scene of the IVE with six degrees of freedom (6DoF), where the viewer can control their position as forwards/backwards, up/down, left/right along the same 3DoF perpendicular axes described earlier. While many such animated experiences play along the linear timeline of the music, they allow the viewer to explore CGIs and provide control over other 3D content fully. Unfortunately,

Figure 7.2 Gorillaz's *Saturnz Barz (Spirit House)* presented as a 360° animated video (Gorillaz, 2017).

computer-generated characters and environments can feel 'uncanny' and 'unnatural' to viewers (Mori et al., 2012; Zell et al., 2015) or require high-budget productions to achieve adequate realism (Perry, 2014). Although motion capture offers some compromises, the requirement for actors to wear markers creates other hurdles for content creators to overcome.

Several advanced cameras have been developed and adapted to capture live performances and bridge realistic performance representations via 360-degree videos and immersion of 3D content (see Figure 7.3). One such camera, the scalable, mirror-based, multicamera system OmniCam-360, developed at Fraunhofer HHI, allows live video recording in a 360-degree panoramic format (Fraunhofer HHI, 2020a). Such recordings have been made during live shows; with real-time camera stitching, the 360-degree video panorama can be made viewable in VR HMDs (Fraunhofer HHI, 2020b). Moreover, it is possible to produce S3D VR content with the OmniCam-360 by equipping two cameras per mirror segment, providing parallax-free, 360-degree panoramic recording for distances larger than two meters within a 60-degree vertical capture space. This system has been used successfully to capture musical performances, providing real-time, remote viewing in VR for backstage areas and bringing classical orchestras to more casual living room viewers (Fraunhofer HHI, 2020b).

Although often limited by 3DoF and linear timelines, many immersive videos can be seen in everyday practice. For instance, OneRepublic's *'Kids'* transports viewers back and forth between two bedrooms (OneRepublic, 2016), immersing the viewer in the music and the characters and narrative devices presented in the song lyrics and videography. Similarly, EDM artist Avicii released a live-action video for the song *'Waiting for Love,'* where the viewer could control the camera position with 3DoF across a 180-degree performance space, tracking the

Figure 7.3 Björk's *Stonemilker*, presented in VR as a 360° music video (Björk, 2015).

performance of multiple dancers within the mise en scène (Avicii, 2015). Other artists, such as the Weeknd (2015), Muse (2016), and Sampha (2017), have each used live-action, 360-degree music videos paired with CGI and Hollywood movie-grade special effects to accompany their music performances with varying degrees of success. Although 3DoF technology has helped to re-establish affordable VR, with the announcement to discontinue Samsung Gear VR and Oculus Go, 3DoF technologies are quickly being outmoded for more innovative and immersive 6DoF platforms (Oculus Developer Success Team, Oculus Developer communications, June 23, 2020).

In response, volumography – the practice of creating digital 3D objects by calculating volumetric geometries from an image or video of the original – has grown in popularity for capturing 3D musical performances. In this context, the volumography can be static or dynamic and reconstructed as a 3D moving image in applications like VR or AR. The final 3D outputs are known as volumetric video (VV). Early high-profile examples of VV include the use of LiDAR – an acronym of 'light detection and ranging' or 'laser imaging, detection, and ranging' – for *'House of Cards'* by Radiohead (2008). More creative examples of VV include the project *NEBULA* by Marcin Nowrotek (2018), generating abstract visualisations in response to music and live performance.

Recently, VV capture was used for Billy Corgan's *'Aeronaut'* and Tino Kamal's *'VIP'* (see Figure 7.4b); both were captured at the Microsoft Mixed Reality Capture Studio (Corgan & Rubin, 2017; Kamal & Lane, 2019). VV content can also be integrated into live choreography and music performances to provide 3D holograms that accompany artists onstage. The most recent notable example is the use of VV at the 2019 Billboard Music Awards (Madonna & Maluma, 2019); this VV was captured at Dimension Studios and used Unreal Engine to execute the experience in real time (see Figure 7.4a). Although several notable studios worldwide can capture, control, and generate VV content, many of them are highly expensive professional set-ups that include large numbers of high-definition cameras and other types of sensors.

At present, VV is becoming a popular technique for capturing live action in 3D space for reproduction on emergent XR platforms (Huang et al., 2018). Combining the freedom of creativity afforded by CGI, the realism of live-action video, and the 6DoF provided by contemporary XR platforms, VV presents a natural progression for new music videos to retain the depth, parallax, and other visual properties of the original scene. These properties can be used to further enhance the qualitative experience of music videos through the application of spatial audio practices (Zhang et al., 2017) via ambisonic recording and binaural playback technologies and haptics, such as audio-tactile haptic feedback (Young et al., 2016, 2017). Therefore, VV presents itself as a disruptive technology for providing artists and audiences with real-time capabilities and immersive, multimodal experiences of musical practice and performance. With the capture of live music within 3D space and its delivery via XR technologies, audiences can explore a VV capture, providing augmented engagements with pre-recorded real-time performances via new multimedia technologies. To date, the amalgamation of these

(a)
(b)

Figure 7.4 (a) Madonna and Maluma's live performance at the 2019 Billboard Music Awards (Madonna & Maluma, 2019), and (b) Tino Kamal's *'VIP'* volumetric music video (Kamal & Lane, 2019).

technologies has proven to be one of the most critical developments in the capture of live-action musical performances since the creation of musical short films of the late 1920s.

The Possibilities (and Approaches) for (AI) Music in Immersive XR Environments

Multimedia has traditionally been applied to capture and reproduce performance and then use technology creatively as a postproduction tool to substitute human-to-human performance. With the volumetric capture of a live act, where a performer is physically present, and the performance is captured 'at the moment,' a more humanistic 3D representation can be reproduced. In XR with full 6DoF, this has only been seen previously in mono- or stereoscopic 3DoF 360-degree video and 6DoF CGI music videos separately. Where performances and audiences are remotely located, time and space can be linked using media-sharing platforms, where the sensation of commonality between human-to-human communication is shared via the use of modern telepresence technology. This effect allows artists to communicate and share with their fans more freely when a physical performance is impossible.

By building upon such approaches to sharing creative musical practice and incorporating both AI and XR, collective performances will be potentially enhanced by providing audiences with an immersive 6DoF platform for sharing the intentions of the artist, enriching and augmenting artist/audience relationships, and facilitating interactive practices that can potentially occur between human-machine and machine-human performances. In such applications, where physical human-human proximity is impossible, the 'live' experience of the musical performance within a virtual environment becomes more intimate, bridging artistic practice, audience engagement, and theoretically providing interactive computational music experiences.

Music videos have traditionally been viewed as passive media; however, AI and XR can potentially be combined to turn them into an interactive medium. Currently, audience interactions with VV are limited to a passive observer role, as is seen in traditional videos as TV or cinema. VV expands interaction in the spatial sense, allowing the viewer to enter a scene with 6DoF, but it does not support real-time interaction. VV is still a recording and, therefore, not as fully interactive as a 3D CGI experience. This limitation is because VV content cannot be easily changed, modified, or manipulated once it has been captured and processed, all of which would be essential for a VV to be interactive. However, in games and other computer graphics applications, autonomous agents are driven by AI and digital humans, as game characters, avatars, etc. have reached a high level of sophistication that can learn from their in-game experiences and build empathy with the audience (Young et al., 2021). Future interactive VV content will require such AI-powered autonomous agents and applying AI in other relevant scene elements.

Currently, these are two separate worlds, with VV on one side, 3D and created from real-world footage but not intelligent, and AI agents as avatars on the other, enabling content interaction but lacking affordable humanistic realism. Therefore, bridging these two worlds would allow for the ultimate, immersive, and interactive musical experience. AI must be integrated into VV content to achieve this connection, ranging from simple 3D geometry features, such as animated 3D rigging, to behavioural rules and high-level semantic awareness. These are still challenging topics that are subject to ongoing research. Once readily possible, it will remain a challenge to balance the recorded nature versus the intelligent nature that a VV character should exhibit. This balancing act will probably be a design choice, and corresponding authoring tools will be needed for VV content creation. Once we bridge the gap between VV and AI, we can design fully interactive XR-AI experiences, enabling 6DoF-spatial exploration for the audience and allowing creative, interactive artistic content. We may then enter an era of immersive music performance in a truly interactive environment that will enable audiences to play along as members of the band or learn new playing techniques from the greatest artists of the time.

In the datafication of a physical, creative practice, such as a musical performance, future XR and AI integration will allow for interactive musical experiences. In this way, XR and AI will serve as a unified communication tool for the capture and reproduction of a musical performance, as well as providing an interactive experience for audiences. Performance, in its many forms, can present composition in a state of flux, as although a piece of music ought to be performed the way the composer intended it to be performed, the performance can also be subject to interaction and interpretation as a performer deviates, learns, and plays. Beyond this, AI and music creativity algorithms can produce unique, interactive performances, as is often seen on the physical stage. However, the inference of motives, meanings, and concepts derived from artistic intentions is likely to be heightened through the use of XR, and the correct comprehension of these factors is not only delicate but also relies on a multitude of innately human-to-human communication channels.

Conclusions

Although Lanier (2017) expressed that 'VR = -AI (VR is the inverse of AI),' the possibilities and approaches for AI music and XR are currently expansive and subject to ongoing research. Digital media allows artists presently to engage audiences in new ways. However, future advances in the expanding metaverse cannot be easily achieved without AI and XR integration. Whether engaging with musical performance via cinema, TV, or XR, the primary function of the platform is the prevalence of musical performances that can occur remotely at multiple times and locations. With current information and communications technology (ICT), the capabilities for the efficient capture and transmission of performance over time and space are readily achievable in multiple media formats, with data rates and compression algorithms being advanced all the time.

Nevertheless, for the communication of artistic creativity or musical performance, it is not enough for XR to be used to reconvey a 6DoF version; it should also consider the creation or contribution of unique content via an AI actor, as simulation or simulacra. In this way, high-speed ICTs may be used to facilitate the recording and reproduction elements of a musical performance by presenting it to a telepresent audience, with XR-AI platforms allowing for immersive and creative interactions to occur between artists and audiences, as might be experienced in a live or collaborative performance space.

Bibliography

3IT (2018) 360° Figaro [360 video]. *Innovation Center for Immersive Imaging Technologies: Berlin*. www.3it-berlin.de/360-figaro-first-opera-recording-realized-for virtual-reality-with-the-omnicam-360/.

AIArtists (2019) *Timeline of AI art*. https://aiartists.org/ai-timeline-art

Avicii (2015) *Waiting for love [360 video]*. YouTube. https://youtu.be/edcJ_JNeyhg.

Bastug, E., Bennis, M., Médard, M. & Debbah, M. (2017) Toward interconnected virtual reality: Opportunities, challenges, and enablers. *IEEE Communications Magazine*. 55 (6), 110–117.

Björk (2015) *Stonemilker [360 video]*. YouTube. https://youtu.be/gQEyezu7G20.

Burgess, J. & Green, J. (2018) *YouTube: Online video and participatory culture*. New York: John Wiley & Sons.

Burdea, G.C. & Coiffet, P. (2003) *Virtual reality technology*. New York: John Wiley & Sons.

Cappello, G. (2019) Active audiences. In: *The international encyclopedia of media literacy*. https://onlinelibrary.wiley.com/doi/10.1002/9781118978238.ieml0003.

Corgan, W.P. & Rubin, R. (2017) *Aeronaut [Volumetric music video and VR experience]*. YouTube. www.youtube.com/watch?v=36_4dx_0Qdk.

Dannenberg, R.B. (1984) An on-line algorithm for real-time accompaniment. *International Computer Music Conference (ICMC)*. 84, 193–198.

DeMarco, M. (2017) This old dog [vr/360 video]. *Inception*. https://inceptionvr.com/experience/mac-demarco-old-dog-vr-music-video-experience/.

Evans, L. (2018) *The re-emergence of virtual reality*. London: Routledge.

Fraunhofer HHI (2020a) *OmniCam-360*. www.hhi.fraunhofer.de/en/departments/vit/technologies-and-solutions/capture/panoramic uhd-video/omnicam-360.html.

Fraunhofer HHI (2020b) *Shootings*. www.hhi.fraunhofer.de/en/departments/vit/technologies-and-solutions/capture/panoramic uhd-video/shootings.html.

Gorillaz (2017) *Saturnz Barz (spirit house) 360 [360 video]*. YouTube. https://youtu.be/lVaBvyzuypw.

Heilig, M.L. (1962) *US patent no. 3050870*. Washington, DC: US Patent and Trademark Office.

Herremans, D. (2016) *MorpheuS: Automatic music generation with recurrent pattern constraints and tension profiles*. https://hal.univ-lorraine.fr/IRCAM/hal-03278537v1.

Huang, Z., Li, T., Chen, W., Zhao, Y., Xing, J. et al. (2018) Deep volumetric video from very sparse multi-view performance capture. *Proceedings of the European Conference on Computer Vision (ECCV)*. 336–354.

Jackson, M., Coppola, F.F. & Lucas, G. (1986) *Captain EO [3-D film with special effects]*. New York: Lucasfilm/Walt Disney Pictures.

Jaekel, B. (2015) *Snapchat and Spotify challenge YouTube as a premiere music video source*. www.mobilemarketer.com/ex/mobilemarketer/cms/news/social networks/20489.html.

Jerald, J. (2015) *The VR book: Human-centered design for virtual reality*. San Rafael, CA: Morgan & Claypool.

Kane, S. (2016) How a pianist became the world's most famous futurist. *Business Insider – Tech Insider*. www.kurzweilai.net/business-insider-tech-insider-how-a-pianist-became-the-worlds-most-famous-futurist.

Kamal, T. & Lane, R. (2019) V.I.P. *Vimeo*. https://vimeo.com/353783055.

Kelly, K., Heilbrun, A. & Stacks, B. (1989) Virtual reality: An interview with Jaron Lanier. *Whole Earth Review*. 64 (2), 108–120.

Lanier, J. (1988) Virtual instrumentation. *The Whole Earth Review*. www.jaronlanier.com/instruments.html.

Lanier, J. (2017) *Dawn of the new everything: A journey through virtual reality*. New York: Random House.

López de Mántaras, R. (2016) Artificial intelligence and the arts: Toward computational creativity. In: *The next step: Exponential life*. Madrid: BBVA Open Mind.

Luhmann, T., Robson, S., Kyle, S. & Boehm, J. (2013) *Close-range photogrammetry and 3D imaging*. Berlin: Walter de Gruyter.

Madonna & Maluma (2019) *Medellín*. YouTube. https://youtu.be/9Z1GdMuC9E8.

MoMA (2017) *Thinking machines art and design in the computer age, 1959–1989 [Exhibition]*. www.moma.org/calendar/exhibitions/3863.

Mori, M., MacDorman, K.F. & Kageki, N. (2012) The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*. 19 (2), 98–100.

Muse (2016) *Revolt [360 video]*. YouTube. https://youtu.be/91fQTXrSRZE.

Nowrotek, M. (2018) NEBULA [music video]. *Vimeo*. https://vimeo.com/256668685.

NPR (2008) *The fall of TRL and the rise of internet video*. www.npr.org/templates/story/story.php?storyId=96869060&t=1591716501367.

O'Dwyer, N., Young, G.W., Johnson, N., Zerman, E. & Smolic, A. (2020) Mixed reality and volumetric video in cultural heritage: Expert opinions on augmented and virtual reality. *Culture and Computing*. 1–20, LNCS 12215.

OneRepublic (2016) *Kids [360 video]*. YouTube. https://youtu.be/eppTvwQNgro.

Paradiso, J.A. & Landay, J.A. (2009) Guest editors' introduction: Cross-reality environments. *IEEE Pervasive Computing*. 8 (3), 14–15.

Parker, E. & Saker, M. (2020) Art museums and the incorporation of virtual reality: Examining the impact of VR on spatial and social norms. *Convergence*. 26 (5). doi:10.1177/1354856519897251.

Parker-Starbuck, J. (2011) *Cyborg theatre: Corporeal/technological intersections in multimedia performance*. Cham: Springer.

Perry, T.S. (2014) Leaving the uncanny valley behind. *IEEE Spectrum*. 51 (6), 48–53.

Radiohead (2008) *House of cards [music video]*. Youtube. www.youtube.com/watch?v=8nTFjVm9sTQ.

Robertson, A. & Plumbley, M.D. (2006) *Real-time interactive musical systems: An overview*. Proceedings of the Digital Music Research Network Doctoral Research Conference, London.

Robinett, W. (1994) Interactivity and individual viewpoint in shared virtual worlds: The big screen vs networked personal displays. *ACM SIGGRAPH Computer Graphics*. 28 (2), 127–130.

Sampha (2017) *(No one knows me) like the piano [360 video]*. YouTube. https://youtu.be/V-ncE-yR8mI.

Squarepusher (2015) Stor Eiglass [VR experience]. *Warp*. https://warp.net/updates/squarepusher-stor-eiglass-vr-experience.

Statista (2020) *Virtual reality (VR) – statistics facts*. Statista Research Department. www.statista.com/topics/2532/virtual-reality-vr/.

Sutherland, I.E. (1968) A head-mounted three-dimensional display. *Proceedings of the 9–11 December, Fall Joint Computer Conference, Part I*. 757–764.

Tannenbaum, R. & Marks, C. (2011) *I want my MTV: The uncensored story of the music video revolution*. New York: Penguin.

U2, Owens, C. & Pellington, M. (2007) *U2 3D*. Washington, DC: National Geographic Entertainment.

Veale, T. & Cardoso, F.A. (2019) *Computational creativity: The philosophy and engineering of autonomously creative systems*. Cham: Springer.

Vercoe, B. (1984) The synthetic performer in the context of live performance. *International Computer Music Conference (ICMC)*. 84, 199–200.

Vercoe, B. & Puckette, M. (1985) Synthetic rehearsal: Training the synthetic performer. *International Computer Music Conference (ICMC)*. 85, 275–289.

The Weeknd (2015) *The hills remix feat: Eminem [360 video]*. YouTube. https://youtu.be/2fhjdtQDcOo.

Young, G.W., Murphy, D. & Weeter, J. (2016) Haptics in music: The effects of vibrotactile stimulus in low-frequency auditory difference detection tasks. *IEEE Transactions on Haptics*. 10 (1), 135–139.

Young, G.W., Murphy, D. & Weeter, J. (2017) *A qualitative analysis of haptic feedback in music focused exercises*. New Interfaces for Musical Expression (NIME) Conference, Aalborg University, Copenhagen.

Young, G.W., O'Dwyer, N. & Smolic, A. (2021) Exploring virtual reality for quality immersive empathy building experiences. *Behaviour & Information Technology*. 1–17.

Zaripov, R.Kh. (1960) An algorithmic description of a process of musical composition. *Doklady Akademii Nauk SSSR*. 132 (6), 1283–1286.

Zell, E., Aliaga, C., Jarabo, A., Zibrek, K., Gutierrez, D. et al. (2015) To stylize or not to stylize? The effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)*. 34 (6), 1–12.

Zhang, W., Samarasinghe, P.N., Chen, H. & Abhayapala, T.D. (2017) Surround by sound: A review of spatial audio recording and reproduction. *Applied Sciences*. 7 (5), 532.

Zulić, H. (2019) How AI can change/improve/influence music composition. Performance and education: Three case studies. *INSAM Journal of Contemporary Music, Art and Technology*. 1 (2), 100–114.