Audience Experiences of a Volumetric Virtual Reality Music Video.

Gareth W. Young*
V-SENSE
Trinity College Dublin
Dublin, Ireland.

Néill O'Dwyer
V-SENSE
Trinity College Dublin
Dublin, Ireland.

Matthew Moynihan
V-SENSE
Trinity College Dublin
Dublin, Ireland.

Aljosa Smolic
V-SENSE
Trinity College Dublin
Dublin, Ireland.

ABSTRACT

Music videos are short films that integrate songs and imagery and are produced for artistic and promotional purposes. Modern music videos apply various media capture techniques and creative postproduction technologies to provide a myriad of stimulating and artistic approaches to audience entertainment and engagement for viewing across multiple devices. Within this domain, volumetric technologies are becoming a popular means of recording and reproducing musical performances for new audiences to access via traditional 2D screens and emergent virtual reality platforms. However, the precise impact of volumetric video in virtual reality music video entertainment has yet to be fully explored from a user's perspective. Here we show how users responded to volumetric representations of music performance in virtual reality. Our results preliminarily demonstrate how audiences are likely to respond to music videos and offer insight into how future music videos may be developed for different user types. We anticipate our essay as a formative starting point for more sophisticated, interactive music videos that can be accessed and presented via extended-reality technologies.

Index Terms: Human-centered computing—Human computer interaction (HCI)—Empirical studies in HCI

1 INTRODUCTION

What is your favorite music video? When we think about this question and the music video genre in general, most (older) people, think back to the heyday of music television in the 80s and 90s. Since the 1980s, both the rise and fall of music television have been well documented into the early 21st century [39]. By the mid to late 00s, fewer music videos appeared on television, and attention was shifting instead to the growing popularity of social media, and media-sharing websites [28]. With the inevitable growth and acceptance of online, peer-to-peer music and video sharing platforms, musicians were able to engage with their fans directly, giving both artists and audiences more presidency and control over their viewing practices and listening experiences [8]. With the recent resurgence of extended-reality (XR) technology, more affordable ambisonics, and the rising pattern of volumography in modern film making, the next generation of music video fans may have different perspectives to add to this pioneering area of creative media research.

In the new 21st century post-televisual era, the current market offers several accessible and innovative modes of creation and dissemination via XR technology, such as volumetric video (VV) capture, as well as playback via head-mounted displays (HMDs) for augmented reality (AR) and virtual reality (VR) experiences. On the one hand, in the examination of current computer-generated imagery (CGI) production tools and techniques, it can be observed that many new music videos are exploring the role of emergent 3D capture systems as instruments for new audiovisual production techniques. On the other hand, new paradigms for home media consumption present audiences with state-of-the-art XR platforms that engage the viewers' imagination, provide them with interactive experiences and immerse the audience within virtual environments.

*Corresponding Author: YoungGa@tcd.ie

New music videos can be accessed via emergent XR technology that provides numerous accessibility and interactivity advantages over standard television viewings. Within this domain, VR can be broadly defined as "a computer-generated digital environment that can be experienced and interacted with as if that environment were real" [20]. Research on VR technologies often concerns the human-computer interface [20], focusing on presenting and quantifying realistic, immersive, interactive, simulated worlds. Unfortunately, this approach neglects to acknowledge the creative potential of VR content as an art form in and of itself [32].

Volumography — the practice of creating digital 3D objects by calculating volumetric geometries from an image or video of the original — has been successfully used to capture 3D musical performances (see Figure 1) and is quickly becoming a popular technique for capturing live-action performances and playing them back for audiences at different times and locations [9, 13]. Volumography can be reconstructed dynamically as a moving image, and these 3D outputs are known as VV [19]. Of commercial note was the 2019 release of Björk Digital (a VR exhibition) and the Vulnicura album, available as a downloadable application for multiple VR platforms. These experiences allow the audience to explore CGI characters and provide other interactions within the IVE.

In this paper, we explore the application of VR for presenting new music video content – explicitly focusing on the application of VV. Using a formative pilot study approach, we seek first to identify future audience needs. Our experiment focuses on the central dimensions relevant to audiences. We then highlight potential challenges with VV content creation and VR experiences. Therefore, our research aims to inform XR music VV practices by providing information on audience experiences.

2 BACKGROUND AND RELATED WORK

Music videos exploded into mainstream home entertainment and popular culture during the 1980s with MTV's genre-defining television channel. Since then, the synonymous use of the term "music video" has been ingrained into the television medium; however, music videos have a much longer and varied history. Musical short films first appeared in the late 1920s [3], and audiovisual entertainment platforms and computational epochs tangentially link music with visual media throughout the 20th century [12, 14, 18]. If maintaining a central focus on the popularity of music television, these histories can be unpacked across distinct epochs that relate to television, as pre-televisual, televisual, and the current post-televisual era [21]. As such, the post-televisual age for music videos represents some significant shifts away from traditional broadcasting techniques [10].

VR technology has been developed over many years and has seen several peaks and troughs in its popularity [22]. Since 2010 there has been an increase in the commercial successes of XR, and VR technology is experiencing yet another renaissance [11]. The VR industry is increasing, and the market size of consumer-grade equipment was projected to increase from 6.2 billion US dollars in 2019 to more than 16 billion US dollars by 2022 [38]. Moreover, despite the manufacturing challenges faced in 2020, the VR market grew. With an increase in spending during COVID-19 to just over

(a) Twelve camera volumetric capture

(b) 3D reconstruction: mesh, object, texture

Figure 1: Volumetric capture and reconstruction, as described by Pagés et al. (2018)

US$1.8bn, VR has managed to generate US$615mn in revenue in 2020 and is expected to grow to US$1.4bn by 2025 [37].

As VV continues to become more affordable and commonplace in VR [19, 34], it can be used to combine the creative and interactive freedoms afforded by CGI and IVEs, the realism and the postproduction of 360° videos, and the overall advantages of 6DoF. Furthermore, VR is multimodal and can enhance the qualitative experiences of volumetric music videos through spatial audio [45] and haptics [43]. As such, volumetric music videos present a unique opportunity to provide artists and audiences with realistic reproductions and immersive multimodal musical experiences.

Evaluating human-computer interaction (HCI) in musical contexts can be challenging [42]. Although user experience evaluation methods can be applied to improve the overall design of interactive digital systems, nontraditional arts technology disciplines may still reject them [17]. Nevertheless, many commonalities exist, and music interaction studies actively contribute to applied computer science [16, 24]. User experience studies in this context follow HCI guidelines, frequently informed by pilot studies and mixed methods approaches [5, 6, 15, 17, 43]. Therefore, our purpose was to capture audience experiences when viewing VV content and provide content creators with a better understanding of the people they are designing for. Thus, our formative research questions are:

RQ1 Are users concerned about the attractiveness of music VV content, as well as pragmatic and hedonistic qualities?

RQ2 What are the latent needs and emotions of audiences, and what are the problems they face?

3 METHODOLOGY

An A/B experiment is presented that exposed users to representative musical VV content designed for traditional screen-based media and VR. The task given to the participants was to view a music video, report subjective opinions, and provide data to investigate what components are required to create quality VV content. This process involved observing, engaging, and empathizing with music video audiences to gather data on their experiences and motivations to engage with new music materials and understand their concerns, requirements, and challenges.

3.1 Participants, Apparatus, and Materials

Recruitment took place in the Republic of Ireland during the latter half of 2021. The project and the university network made a general call for participation. Volunteers were enrolled across a broad spectrum of potential user types and were invited to the laboratory to experience a volumetric music video on a one-to-one basis. The aims of the research and experiment procedures were communicated via a research information sheet before arrival, and any questions were addressed via email. The institutes' Research Ethics Committee gave ethical approval for the following methodology.

Participants were asked to report demographic information via an online questionnaire in advance of the study. This data included age, gender, education, and employment status. The participants were also asked to report on 7-point Likert scales their competencies with digital technologies (1 = Unskilled to 7 = Excellent); their familiarity of music videos (1 = Unfamiliar to 7 = Extremely Familiar); and their expertise using VR and AR technologies (1 = Novice to 7 = Expert). Participants were then scaled as "Novices," "End-users," and "Advanced Users," as described by Nielsen [26].

In total, 13 volunteers contributed to the study, identifying as 10 Males and 3 Females with an average age of 30.46 (SD = 3.41). According to the Irish National Framework of Qualifications (NFQ), the education profile of the cohort consisted of levels 10 (n = 8), 9 (n = 3), and 8 (n = 2). The group's employment according to the professional Nomenclature of Economic Activities (NACE) was Scientific and Technical Activities (n = 8) and Education (n = 7). All participants self-reported normal or corrected-to-normal vision and hearing. The experiment was designed with two stimulation factors:

- **Baseline Scenario (B_0)** – Participants view a music video via a traditional TV screen
- **Experiment Scenario 1 (S_1)** – Participants view a music video via a VR device

The two experiment scenarios were deemed suitable stimuli platforms for analysis as a formative pilot study. The first scenario was a classical TV-baseline measure. The second variable was selected as VR for viewing volumetric music videos. Each stimulus was presented in random order and followed by a post-task report on the subjective experiences of both.

A Samsung UN55B7000 55-Inch 1080p 120 Hz LED HDTV was used for viewing a music video and as an external monitor for the VR experience. A Valve Index VR system was used, with dual 1440 x 1600 RGB LCDs running at 120 Hz and a specified field of view (FoV) of 130°. The Valve Index wireless controllers were used as the input device. The total space available was defined within a 2x2m interaction area. The VR experiment was conducted on a Dell Arora PC with an Intel Core i7 processor and a dedicated NVIDIA GTX 2080 Ti graphics card. A Logitech K400+ was used to access and control the music video on the PC. For both scenarios, the participants wore headphones (Ollo HPS S4X).

Participants were invited to attend the project laboratory at a time and date that suited their schedule and asked to adhere to COVID-19 protocols. Due to the current heightened risk of COVID-19 infection, the HMD was treated with a hydrophobic nanotech coating and hygienically sterilized using a UVC HMD cleaning device (CleanBox X1). All touch surfaces and hand-held devices were cleaned with antiseptic wipes before and after each session. Social distancing and mask-wearing were also enforced.

Representative VV stimuli for S_1 were selected from the Björk *Vulnicura* virtual reality album that contains a combination of live-action video (360° and 2D) and real-time rendered 3D IVEs as well as spatial audio. From this collection, the track *Notget* [6:38] was selected for this study as it contains a 3D IVE and VV character (see Figure 2 (a)). The same song was chosen for scenario B_0, as a

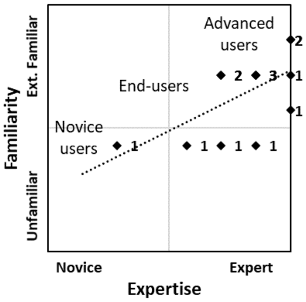


Figure 2: User-cube with number of user types (the dotted line represents the linear average)

traditional 2D music video directed by Warren Du Preez and Nick Thornton Jones. This media was accessed in HD via the YouTube website (`https://youtu.be/aWrV8NQnbqE`).

The stimuli were delivered in a counterbalanced order. Participants were introduced and made familiar with each scenario and input device with no additional training required. For S_1, navigation between tracks was handled by a laser-pointer interface that extended from each Index controller within the IVE. For B_0, a web browser was used to access YouTube. Participants adjusted the volume of the experiences themselves for comfort. A seat was provided for participants who chose to sit. Participants were left by themselves for each scenario duration.

After the experiment and by themselves, the participants completed an online questionnaire via a separate PC to capture their immediate post-task responses to the VR stimuli. The first questionnaire was the User Experience Questionnaire (UEQ) that consists of 26 pairs of contrasting attributes [23]. On these scales, ”Attractiveness” is a pure valence dimension. ”Perspicuity,” ”Efficiency,” and ”Dependability” are pragmatic quality aspects (goal-directed), and ”Stimulation” and ”Novelty” are hedonic quality aspects (not goal-directed). The circles between these attributes represented gradations between the opposites across a seven-stage scale. The participants were then asked via a 5-point Likert scale how familiar they were with ”*Notget*” before the experiment (1 = Not at all Familiar to 5 = Extremely Familiar) and which version of the music video they preferred (the music video, the VR experience, or neither). This question was expanded with a simple ”why?” explanation.

The final four questions were open-ended and explored previous VV content experiences, latent problems solved by using VV, the potential impact on new music videos, and personal feelings towards the technology; this line of inquiry was operationalized by the following questions — ”What previous knowledge or experience have you had with this technology?”; ”What problems could they potentially solve with regards to access to live performances?”; ”How do you think volumetric music videos will affect musicians and audiences?; How do you personally feel about using new technologies in this way?”. Following this, participants were debriefed and allowed to ask further questions. The study took 30 minutes to complete.

4 Results

Empirical data was collected and analyzed. Quantitative data descriptively reported attitudes towards volumetric music video content. Qualitative data were coded and used to enrich and add depth of knowledge to these findings. Due to the distribution of variables, we used non-parametric methods and chose 0.05 for the significance level. The qualitative analyses was guided by the frequency and fundamentality of the issues raised by the participants [1, 27]. Two researchers completed the thematic analysis and had an inter-coder agreement of Kappa (k) = 0.8.

4.1 Population Variables

Data relating to the cohorts' ability to use digital technology (M = 6.38; SD = 0.65), their familiarity or knowledge of music videos (M = 5.62; SD = 1.45), and their expertise or experience in using XR technologies (M = 4.46; SD = 1.13) were captured to identify specific user-types (Nielsen, 1994). All users were self-reported as having ”Very Good” to ”Excellent” ability to use digital technology. The distribution of user types was heavily weighted towards advanced users, see Figure 2 (b). A Kruskal-Wallis test revealed no statistically significant difference ($p > 0.05$) in questionnaire scores across all three different user groups (Novice Users, n = 1, End-users, n = 3, Advanced Users, n = 9), see Figure 2 (b). Therefore, with no significant differences between population variables in the quantitative data, the outcome variables data for user-type were combined for analyses. Before the experiment, the participants considered themselves ”Not at all Familiar” (M = 1.15, SD = 0.55) with Björk's song *Notget*. Participants stood for S_1 and sat for B_0

4.2 Personal Preference

The participants were divided on their preferred experience of B_0 (46.15%), S_1 (46.15%), or no preference for B_0 or S_1 (7.69%), see Figure 3(a). When queried why they preferred one platform, the cohort identified unique features and shortcomings. The qualitative data were subject to thematic analysis, and preferences were acknowledged as platform-specific characteristics.

On the one hand, preference for B_0 could be categorized as visual understanding and familiarity with the format. B_0 was reported to be visually more attractive than S_1, particularly the environment and costumes. Likewise, B_0 was more vividly pleasing as it delivered a more detailed image, and the visual quality was considered ”crisper” than that of S_1. Moreover, the participants who preferred B_0 enjoyed the amalgamation of music and video; for example, the music tempo was supported by cinematic cuts and camera angles. As one participant expressed, ”The music video was able to display the creative intentions of the artist more clearly.” Therefore, B_0 appeared to have a higher production value associated with it than S_1. For example:

> ”There are a lot of visual effects in the music video. I can see the details, like the jewelry mask she wears, her facial expression, hair, and jellyfish-like mask, and there are different scenes in the music video. I can understand the story from the changing scenes.”

Overall, the cinematography of B_0 was thought to be better than that of S_1. As a passive audiovisual experience, the format was considered more accessible. Although the 3D spatial interaction of the audio in VR was novel, it diminished some users' overall understanding of archetypal stereo representations of music. Preferences were also influenced by familiarity with traditional music videos and pre-formed user expectations; this was expressed as ”I feel that I understood the song more from watching the video.”

On the other hand, preferences for S_1 related to immersion and the unique affordances of the VR platform. S_1 provided an immersive, novel, and authentic experience, and the content was considered more personal, visceral, and surreal. The ”visual experience” was supposedly boosted through the use of VR. Furthermore, a sense of presence was afforded within the imaginary IVE, enhancing unique understanding and personal experiences of the content; for example, ”There was a sense of ownership/freedom as if I could interact with the surroundings.” Therefore, the immersive nature of S_1 positively enhanced the experience for these participants.

Additionally, the spatial effect of VR music led to innate feelings of ”exploration” and firsthand experiences of music. The impact of being placed ”within” the song's visual narration was considered more impactful and effective. B_0 was deemed predictable, that ”we see a lot of music videos portrayed that way.” For S_1, users felt fearful, especially users with phobias, although tensions eased up as

Table 1: User Experience Measures for S_1

Scale	Mean	Std. Dev	Conf Int.	λ^2	x^2 (2, n = 13)	p
Attractiveness	1.60	1.03	0.56	0.94	3.36	0.19
Perspicuity	1.42	0.71	0.39	0.69	3.32	0.19
Efficiency	0.81	0.84	0.46	0.68	1.48	0.48
Dependability	0.29	1.09	0.59	0.73	2.80	0.25
Stimulation	1.77	0.98	0.53	0.84	2.05	0.36
Novelty	1.98	0.94	0.51	0.91	1.77	0.41

Note: A Guttman's lambda-2 (λ^2) $\geq$0.70 was sufficiently consistent for our evaluation

the video moved to a lighter pace, further highlighting the effects of presence. This reflection was considered attractive, as S_1 invoked more emotions than B_0. For example, "I was impressed with the visual effects only in the music video, but as an audience member, I didn't feel connected to it, unlike the VR experience." The participant who expressed no preference recounted similar sentiments but could not resolve a particular choice.

4.3 UEQ

The study aimed to capture formative data relating to experiences of a volumetric music video; therefore, participants were asked to focus their user experience evaluations on S_1 only. The UEQ captured the attractiveness and hedonic (stimulation and novelty) and pragmatic (efficiency, perspicuity, and dependability) experiences of S_1 (see Table 1 and Figure 3(b)). Hedonic qualities defined non-task-related quality items, and the pragmatic quality scales describe the task-related quality ratings.

4.4 Qualitative Feedback

A thematic analysis was applied to the open-ended reports of the participants. This approach emphasized occurring problems and their importance during the music video.

4.4.1 Previous Knowledge and Experiences of Volumetric Music Videos

A thematic analysis of previous knowledge and experiences presented variable involvements with VV before the experiment; some prior knowledge focused on interactive theatrical works, such as *Virtual Play* [29, 31] and *Augmented Play* [40] (XR reinterpretations of Samuel Beckett's *Play* (1963)) and *XR Ulysses* (XR reinterpretations of James Joyce's *Ulysses*), collaborations between V-SENSE (an academic research group) and the Irish start-up company Volograms. VV was also observed in several video game experiences for characters with facial and body features based on actual people, such as *Awake: Episode One* (2018), etc. Participants deemed that experiences of VV work that incorporated interactive elements were more favorable. However, it was noted that it was unfair to compare these types of VV with a music video, as their modalities were markedly different. Moreover, environment-specific experiences highlighted situations where VV could be used best, depending on the application, for example, whether the user should experience VV outside or in the home. VV was also used in other mixed reality applications to overlay engaging visuals on real-life scenes such as AR cultural heritage tourism [32].

Additionally, different platforms for VV media consumption had been experienced, highlighting, for example, "Smartphones are better suited for casual creative applications with VV" and "For movies, sports, etc., an HMD would be better suited." While the experiment focused on VR, comments were also directed towards previous VV experiences with AR technology. For example, "I probably would not prefer to watch volumetric music videos on PCs and laptops. Tablets or smartphones might also work for AR applications." Compared to other experiences with VV, the visual saliency within VR was criticized for being too narrow, as the action was missed during the single viewing experience because the user's attention was pulled in multiple directions. From these previous experiences, HMDs were generally considered the best devices to view VV content; however, the quality of the content was described as not yet photorealistic enough or physically engaging within the IVE. As noted:

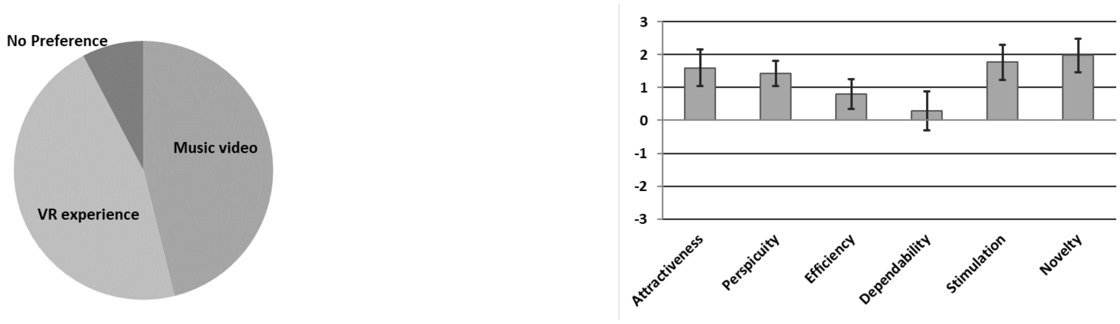
> "I believe that volumetric music videos can be used best to give a sense of live performance, as was seen in the Björk video; however, a greater engagement with the viewer could have been made by incorporating the viewer's position into the dance."

4.4.2 Improvements to VR Music Videos

The first potential improvement was related to the current accessibility of live performances during COVID-19 lockdowns. The broader appeal of VR was that it could reach wider audiences compared to the current limitations of live concert venues. However, the main reason our users would prefer a live experience over VR was for the engagement — "the artist being there, and the public all around." Participants believed that one fundamental aspect of live performances with VR was missing, vis-á-vis — "being there together and experiencing the same spectacle together." It was also noted that audiences should see fellow fans/spectators during a performance to experience a shared concert, a vital facet that brings a sense of closeness and belongingness to the people around. This togetherness could be lost in VR unless there was a way of simultaneously sharing the same VV experience (e.g., social VR). These experiences would be similar to watching a film at a cinema or streaming a movie.

It was believed that VV could make a digital performance more intimate than a live performance as the viewer is alone in an environment with the performer. However, what was lost was the energy associated with watching a concert in a group. Compared to a physical show, the proximity and social components of being in a crowd, which can be unpredictable in terms of the behavior of others, affected the experience of *Notget*. These types of physical interactions can be an integral "experience" in and of themselves — "The sense of being in a group can only be simulated (weakly so far) in such VR experiences." Most people cannot get near the stage during a live concert, especially at large venues. This situation is perhaps more valid for people "people who can't travel, or are not able-bodied" [32] — "To experience a performance from any position of choice, e.g., center stage or being part of a crowd in a small club or a vast stadium by choice." volumetric music videos presented in VR can potentially solve this problem by getting audiences closer to the performance space.

Concerning visual realism (or naturalism), the uncanny valley of the VV experience was also a formidable hurdle that needed to be resolved. CGI and IVEs can feel "uncanny" and "unnatural" [25, 44] and require high budgets to achieve realism [36]. According to our participants, naturalism was a critical factor to improve volumetric music videos, such as image quality, higher resolution textures, absence of artifacts in the 3D models, and smooth animation. However, more realistic (or high fidelity) representation should not curb the creative potential of the content — "I would love to watch a VR concert with an HMD, especially in unreal, mythical, historical places." While the overall quality of the presented VV was "good but far from realistic," the overall experience was very different than an actual concert. With IVEs presented in VR, the creators/directors could weave the music into a narrative and tell a different story for each part of a song. It would also be beneficial to customize the



(a) User preference.

(b) UEQ.

Figure 3: Results.

audio/video for the best experience for the individual. Unfortunately, the VV quality would depend on the quality of the user's hardware."

4.4.3 Advantages and Disadvantages of VR Music Videos

A thematic analysis of responses to the potential problems and benefits of using VV in VR questions revealed several pros and cons regarding volumetric music video consumption, see Table 2. These areas of concern were centered around five core themes: Experience, Audience, Industry, Performer, and Technology.

5 Discussion

Concerning the attractiveness of music VV content, as well as pragmatic and hedonistic qualities (RQ1), while considered "attractive" (M = 1.60, SD = 1.03) and positive "hedonic" factors were measured (M = 1.88), the "pragmatic" elements of the platform were rated lower (M = 0.84), see 1. Our cohort rated the *Notget* VR experience as easy to become familiar with (M = 1.42, SD = 0.71), stimulating to use (M = 1.77, SD = 0.98), and novel (M = 1.98, SD = 0.94). However, the efficiency (M = 0.81, SD = 0.84) and dependability (M = 0.29, SD = 1.09) of the experience was rated lower. These measures indicated that users applied unnecessary effort during the VR experience and did not feel securely or predictably in control of the interaction. The cohort was divided on their preference for musical content presented in VR and via traditional media. Priority for B_0 could be categorized concerning visual understanding and familiarity with the format. Choices for the VR experience were related to immersion and the unique affordances of the VR platform. Both platform versions of *Notget* brought individual factors of enjoyment that influenced preference for one platform over the other. The VR experience was considered more immersive, but it struggles to achieve equivalent fidelity to B_0 due to numerous digital constraints, e.g., texture resolution. Users also commented that although they perceived lip-sync issues in both, this was more prominent in VR.

The qualitative feedback identified specific effectors that impacted our understanding of user experiences and described the audiences' latent needs, emotions, and the problems they faced (RQ2). Our cohort was somewhat familiar with the VV format and expressed prior engagement with VV via different platforms, and the combined use of VV and music content stimulated many comparisons to live concert performances, the embodiment of performers, and audience interactivity. VR created a more vivid experience for our cohort; however, they believed increased interaction within the IVE would improve user agency and presence. It was suggested that these issues could be solved by the VV looking at the audience (as reported in earlier pilot studies by O'Dwyer et al. [33]), becoming more physical by occupying personal space and keeping the avatar pointing in the viewers' direction when addressing the audience. However, significant adverse reactions to interpersonal space violations have been observed in stereoscopic 3D displays [41]. Other research has also shown that, compared to stylization, photorealism increases place illusion, changes the emotional response, decreases the perceived movement realism, and does not affect the character's comfort levels [46]. Furthermore, "disturbing environments" have been shown to impact user experiences [4] negatively.

Feedback on human experiences emphasized the many advantages of immersive and imaginative worlds for VV musical content. Increased feelings of presence exemplified this effect within the IVE as the experience progressed and became more abstract. The poten-

Table 2: Advantages and disadvantages of S_1 – VR music videos

Core Themes	Advantages	Disadvantages
Experience	· Immerisve virtual worlds · Interactive 3D content · Creative freedom · Presence	· Dehumanizing · Unsocial · Lack of performer interaction · No person-to-person reciprocal actions
Audience	· Outreach · Enhanced live experiences · Unique performances · Intimate engagement	· Access to XR technology · Technology adoption · The pervasiveness of the platform · Financial costs
Performer	· Virtual musical experiences · Enhanced live performances · Access to new genres of music · Non-intrusive for performers. · Live and can be captured for future viewing	· Lack of spontaneity · Absence of ambiance · Soloistic · Lack of natural performance or "reality"
Industry	· Cost-efficient concerts · New jobs and skills in the music industry · Telepresent performances · Market accessibility	· Increases in performance costs · Current skill deficits · Resistance to change
Technology	· Novel · Innovative	· Analogue Vs. digital experiences · Comfort

tial to open up alternative creative approaches for content providers was also advantageous. However, these IVE/VV experiences were also criticized for not providing human-to-human or shared experiences. It has been shown that sharing an affair with another person without communicating amplifies one's experience [7]. This effect impacted the audiences' expectations of a live performance by dehumanizing the performer as they could not resolve the absence of performer interaction. Therefore, future design iterations should consider the impact of visual acuity, immersion, and presence on audiences' expectations as established by traditional performance.

From an audience perspective, the cohort commented on the uniqueness of the performance, stating that it enhanced the concept of live music video performances. The use of VV gave the audience an impression of intimacy as they could be physically closer to the performer and within the performance space. This intimacy was advantageous for reaching new audiences and engaging with emergent media platforms. Although expanding [37, 38], VR music video markets were considered much smaller than traditional platforms and would incur additional financial costs for the consumer. The impact of novelty is something that content creators can design into their experiences, ramping up the innate functionalities of VR and VV as a musical experience develops over its duration.

Virtual musical experiences in VR were considered to enhance live music performances but should not attempt to replace them; video did not kill the radio star, and VR music videos will not kill televisual music videos. The performer would benefit from using VV in VR as they could potentially be viewed live in a non-invasive way because the audience would be virtually closer, not physically. As this type of (offline) VV is pre-recorded, performances are available for repetitive viewing. The performer would be provided with a new medium to explore progressive performance approaches that would benefit the expansion of their music performance genera. However, the fixed-format would inherently lack spontaneity.

The six degrees of freedom (6DoF) presented in VR gave our users a unique way to view and interact with the performance media. The VV studio capture was not shared with an audience, and therefore it was thought that it lacked the energy and ambiance that a live entertainer would experience when performing on stage in front of an audience. This soloistic and absence of natural performance expectations was thought to impact the performer and reduce their experience of audience interactions. Within this space, shared or social VR technologies are already highly successful at facilitating live interactions, albeit with lo-res animated avatars; there is a long way to go before photorealistic content can be displayed in real-time.

6 Future Work

The advantages and disadvantages of the experience highlighted several factors that can be addressed for quality volumetric music video experiences. These user-identified factors will be used to develop new research approaches to music video content creation, highlighting a veritable number of do's and don'ts when designing this type of content. For example, conventional narrative experiences are often delivered via a 2D screen (using the 4th wall), and VR technology can disrupt this paradigm to make the audience present within an IVE rather than a remote viewer. That said, HMD displays are often used to deliver first-person perspectives, and given that other devices are limited to visualizing through a 2D screen, they are often reliant on and created for third-person narratives.

For user-centered designs, this opens a tension between the conventional paradigm, where artists present audiences linear narrative journeys, and the ambitions of performers to elicit the specific potentialities of XR media [30]. There is a risk that the narrative, story, or artistic vision becomes completely negated by engaging the audience through interactive mechanisms, which may destroy the art idea. For example, if the audience has too much agency, this could negatively impact the engagement, ultimately bringing about ennui.

Due to COVID 19, the number of participants was low, and the gender balance was heavily weighted [35]. Therefore, the data captured during this pilot study must be used carefully to inform the design of volumetric music video experiences. In this formative study, we have highlighted specific qualities that audiences seek during the consumption of such materials. Future work will directly compare traditional music videos to XR technology-mediated experiences. Moreover, we will endeavor to make VV content more affordable, accessible, and life-like to overcome the issues presented.

7 Conclusions

Overall, our participants expressed an awareness of the disruptive nature of volumetric music videos. Although the technology is still emerging and requires formal grammar [2, 32], much work remains to make VR more affordable and integrate VV techniques into current workflows. Our users struggled to suspend their awareness of real-world performance; therefore, the content was overtaken by expectations. Furthermore, the novelty of VV in VR made it hard to measure its effect, as audiences resolve to separate their awareness of physical performance. Addressing the issues raised by our participants will help content creators become comfortable using it and make audiences more enthusiastic about engaging with it.

Music can be described as an evolutionarily deep-rooted, abstract, real-time, complex, non-verbal, social activity [16]. Therefore, the original contribution of this research rests in the creation of user-focused data for future 6DoF volumetric VR music videos and identifies the potential impact of this type of musical experience on audiences. Future iterations of this novel experiment are expected to focus on differences between traditional media and new VR experiences and expose and build upon existing HCI studies that focus on music and technology in use, specifically those concerning how users experience music videos presented via XR technologies.

Acknowledgments

This publication has emanated from research conducted with the financial support of the Science Foundation Ireland (SFI) under Grant Number 15/RP/2776.

References

[1] A. Adams, P. Lunt, and P. Cairns. A qualititative approach to hci research. In P. Cairns and A. Cox, eds., *Research Methods for Human-Computer Interaction*, pp. 138 — 157. Cambridge University Press, Cambridge, UK, 2008.

[2] J. Bailenson. *Experience on demand: What virtual reality is, how it works, and what it can do*. WW Norton & Company, New York, 2018.

[3] R. Barrios. *A song in the dark: The birth of the musical film*. Oxford University Press on Demand, Oxford, 1995.

[4] A. Beacco, R. Oliva, C. Cabreira, J. Gallego, and M. Slater. Disturbance and plausibility in a virtual rock concert: A pilot study. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 538–545. IEEE, 2021.

[5] B. Bengler and N. Bryan-Kinns. Designing collaborative musical experiences for broad audiences. In *Proceedings of the 9th ACM Conference on Creativity & Cognition*, pp. 234–242, 2013.

[6] T. Blaine and S. Fels. Collaborative musical experiences for novices. *Journal of New Music Research*, 32(4):411–428, 2003.

[7] E. J. Boothby, M. S. Clark, and J. A. Bargh. Shared experiences are amplified. *Psychological science*, 25(12):2209–2216, 2014.

[8] J. Burgess and J. Green. *YouTube: Online video and participatory culture*. John Wiley & Sons, Hoboken, 2018.

[9] A. Collet, M. Chuang, P. Sweeney, D. Gillett, D. Evseev, D. Calabrese, H. Hoppe, A. Kirk, and S. Sullivan. High-quality streamable free-viewpoint video. *Transactions on Graphics*, 34(4):1–13, 2015.

[10] M. Edmond. Here we go again: Music videos after youtube. *Television & New Media*, 15(4):305–320, 2014.

[11] L. Evans. *The re-emergence of virtual reality*. Routledge, Abingdon-on-Thames, 2018.

[12] S. Frith, A. Goodwin, and L. Grossberg. *Sound and vision: The music video reader*. Routledge, Abingdon-on-Thames, 2005.

[13] K. Guo, P. Lincoln, P. Davidson, J. Busch, X. Yu, M. Whalen, G. Harvey, S. Orts-Escolano, R. Pandey, J. Dourgarian, et al. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Transactions on Graphics (TOG)*, 38(6):1–19, 2019.

[14] C. H. Hansen and R. D. Hansen. Music and music videos. *Media entertainment: The psychology of its appeal*, pp. 175–196, 2000.

[15] S. Holland, A. J. Bouwer, M. Dalgelish, and T. M. Hurtig. Feeling the beat where it counts: fostering multi-limb rhythm skills with the haptic drum kit. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, pp. 21–28, 2010.

[16] S. Holland, A. P. McPherson, W. E. Mackay, M. M. Wanderley, M. D. Gurevich, T. W. Mudd, S. O'Modhrain, K. L. Wilkie, J. W. Malloch, J. Garcia, et al. Music and hci. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 3339–3346, 2016.

[17] K. Höök, P. Sengers, and G. Andersson. Sense and sensibility: evaluation and interactive art. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 241–248, 2003.

[18] R. Howells and J. Negreiros. *Visual culture*. John Wiley & Sons, Hoboken, 2019.

[19] Z. Huang, T. Li, W. Chen, Y. Zhao, J. Xing, C. LeGendre, L. Luo, C. Ma, and H. Li. Deep volumetric video from very sparse multi-view performance capture. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 336–354, 2018.

[20] J. Jerald. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool, San Rafael, 2015.

[21] M. B. Korsgaard. *Music video after MTV: Audiovisual studies, new media, and popular music*. Routledge, Abingdon-on-Thames, 2017.

[22] J. Lanier. *Dawn of the new everything: A journey through virtual reality*. Random House, New York City, 2017.

[23] B. Laugwitz, T. Held, and M. Schrepp. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*, pp. 63–76. Springer, 2008.

[24] L. Liikkanen, C. Amos, S. J. Cunningham, J. S. Downie, and D. McDonald. Music interaction research in hci: let's get the band back together. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pp. 1119–1122. ACM, 2012.

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

[26] J. Nielsen. *Usability engineering*. Morgan Kaufmann, Burlington, 1994.

[27] L. S. Nowell, J. M. Norris, D. E. White, and N. J. Moules. Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods*, 16(1):1609406917733847, 2017.

[28] NPR. The fall of TRL and the rise of internet video. `https://www.npr.org/templates/story/story.php?storyId=96869060&t=1591716501367m`, Nov 2008 [Online].

[29] N. O'Dwyer and N. Johnson. Exploring volumetric video and narrative through samuel beckett's play. *International Journal of Performance Arts and Digital Media*, 15(1):53–69, 2019.

[30] N. O'Dwyer. *Digital Scenography: 30 Years of Experimentation and Innovation in Performance and Interactive Media*. Bloomsbury Publishing, London, 2021.

[31] N. O'Dwyer, N. Johnson, E. Bates, R. Pagés, J. Ondřej, K. Amplianitis, D. Monaghan, and A. Smolić. Virtual play in free-viewpoint video: Reinterpreting samuel beckett for virtual reality. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 262–267. IEEE, 2017.

[32] N. O'Dwyer, G. W. Young, N. Johnson, E. Zerman, and A. Smolic. Mixed reality and volumetric video in cultural heritage: Expert opinions on augmented and virtual reality. In *Culture and Computing: Part of the 22nd HCI International Conference (HCII 2020)*, pp. 195–214. Springer, 2020.

[33] N. O'dwyer, E. Zerman, G. W. Young, A. Smolic, S. Dunne, and H. Shenton. Volumetric video in augmented reality applications for museological narratives: A user study for the long room in the library of trinity college dublin. *Journal on Computing and Cultural Heritage (JOCCH)*, 14(2):1–20, 2021.

[34] R. Pagés, K. Amplianitis, D. Monaghan, J. Ondřej, and A. Smolić. Affordable content creation for free-viewpoint video and vr/ar applications. *Journal of Visual Communication and Image Representation*, 53:192–201, 2018.

[35] T. C. Peck, L. E. Sockol, and S. M. Hancock. Mind the gap: The underrepresentation of female participants and authors in virtual reality research. *IEEE transactions on visualization and computer graphics*, 26(5):1945–1954, 2020.

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

[37] PWC. Power shifts: Altering the dynamics of the e&m industry. `https://www.pwc.com/gx/en/industries/tmt/media/outlook/segment-findings.html`, 2021 [Online].

[38] Statista Research Department. Virtual reality (VR) - statistics facts. `https://www.statista.com/topics/2532/virtual-reality-vr/`, 2020 [Online].

[39] R. Tannenbaum and C. Marks. *I want my MTV: The uncensored story of the music video revolution*. Penguin, London, 2011.

[40] V-SENSE. XR play trilogy. `https://v-sense.scss.tcd.ie/research/mr-play-trilogy/`, 2019 [Online].

[41] L. M. Wilcox, R. S. Allison, S. Elfassy, and C. Grelik. Personal space in virtual reality. *ACM Transactions on Applied Perception (TAP)*, 3(4):412–428, 2006.

[42] G. W. Young and D. Murphy. HCI models for digital musical instruments: Methodologies for rigorous testing of digital musical instruments. In *Computer Music Multidisciplinary Research*, pp. 534 – 544, 2015.

[43] G. W. Young, D. Murphy, and J. Weeter. A qualitative analysis of haptic feedback in music focused exercises. In *Proceedings of the international conference on new interfaces for musical expression (NIME)*, pp. 204 – 209, 2017.

[44] E. Zell, C. Aliaga, A. Jarabo, K. Zibrek, D. Gutierrez, R. McDonnell, and M. Botsch. 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, 2015.

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

[46] K. Zibrek, S. Martin, and R. McDonnell. Is photorealism important for perception of expressive virtual humans in virtual reality? *ACM Transactions on Applied Perception (TAP)*, 16(3):1–19, 2019.