

A comparison of presence and emotion between immersive virtual reality and desktop displays for musical multimedia

Thery Al Alam[†]

Nicola Dibben

Department of Music, University of Sheffield, Sheffield, UK

[†] Corresponding author: thery.alam@live.com

Published 16 December 2021; <https://doi.org/10.18061/FDMC.2021.0017>

Author video presentation and/or other conference material: <https://doi.org/10.17605/OSF.IO/86K93>

Abstract

Visual and auditory immersion influence sense of presence, and in some cases emotional response. However, these have previously been tested separately and it is unknown whether and how immersion influences experiences combining both, as found in musical multimedia. A within-participants experiment compared participants' sense of presence (feeling of 'being there'), and induced and perceived emotions, when experiencing a classical concert performance presented over headphones on a desktop PC (2D video with headlocked sound reproduction) or in virtual reality (VR) (head-mounted display presenting 360° video and reactive spatial sound reproduction). Emotion was measured by self-report questionnaires based on a three-dimensional model (pleasantness, tense arousal and energetic arousal) and presence was measured using four items (Sense of physical space, Engagement, Ecological Validity and Negative Effects). Participants experienced a significant increase in pleasantness and all dimensions of presence in VR compared to the desktop condition. Ratings of felt arousal and perceived emotions did not significantly differ between conditions. The "engagement" component of presence was positively correlated with emotion felt and perceived, whereas "ecological validity" and "sense of physical space" were related to emotions experienced in the desktop condition only. The role of presence should therefore be considered when investigating emotional responses to musical multimedia.

KEYWORDS: *virtual reality, music, emotion, immersion, presence*

Introduction

'Immersive' technologies for three-dimensional sound and visual reproduction present new means to engage and emotionally affect audiences. Understanding the experience of immersive musical multimedia can extend existing theories of emotional responses to multimedia and potentially inform their design.

Immersive VR is a market-leading technology that enables a high sense of interaction with multimodal experiences. A VR environment often consists of a 3D audiovisual space with visual information presented in 360° and spatialised audio that responds to users'

movement. As we use the term, immersion describes objective properties of the technology (Slater & Wilbur, 1997). Virtual environments (VEs) mediated by VR technologies ('Immersive VR') may be considered 'high immersion' because they usually involve more than two spatially oriented sensorial modalities, a high-fidelity representation of sensorial information and the ability to modify the VE in response to body movement. This contrasts with the most commonly encountered technologies for audio-visual consumption, namely desktop and mobile video. Environments mediated by a desktop PC ('Desktop Video') may be considered 'low immersion' because they offer limited field of vision and sound reproduction and do not respond to body motion (Miller & Bugnariu, 2016).

Immersive VR increases users' sense of presence, or feeling of 'being there', in a VE (Slater & Wilbur, 1997). Presence is associated with allocation of attention to the VE and reduction in resources available for processing other stimuli (Kober & Neuper, 2012), although the definition of presence remains controversial (Coelho et al., 2006). In this paper, 'presence' describes the feeling evoked in a user of being physically located in a display environment mediated by technology (e.g. in a video, game or film), and comprises different experiential domains (Lessiter et al., 2001).

There is evidence that presence is heightened by increased immersion: in the visual domain, presence increases with headset relative to desktop display (Nichols et al., 2000), and with stereoscopic information, head tracking and field of view (Hendrix & Barfield, 1996); in the auditory domain, presence increases with spatialized sound relative to no sound or non-spatialized sound (Kobayashi et al., 2015). However, it is difficult to predict the sense of presence in a VE solely by manipulation of technological factors due to other variables, such as previous experience with VR (Coelho et al., 2006).

Whether emotion is influenced by immersion is less clear as is the relationship between presence and



emotions. Some authors suggest that presence and emotions are related only when the media content is emotionally arousing (Freeman et al., 2005) but further evidence is required.

This study investigated whether levels of visual and auditory immersion together influence emotional experiences of music. We compared induced and perceived emotions, sense of presence, and the relationship between presence and perceived and induced emotions with a Desktop Video (low immersion) and Immersive VR (high immersion) musical multimedia product.

We expected that: (H1): Participants' sense of presence will be higher in the high than the low immersion condition. (H2): Ratings of induced and perceived emotions will differ between the high and low immersion conditions. (H3): sense of presence will correlate with induced and perceived emotions.

Method

Participants

The twenty-eight participants (13 female; M age = 26.4 years, SD = 6.98 years, range 19–52 years) comprised students (82.1%) and staff (17.8%) at the University of Sheffield, of varied nationality, most of whom had prior experience with VR (82.2%). Participants had no history of seizures, blackouts or psychiatric problems nor implanted medical devices. Participants gave informed written consent in keeping with approval granted by the Ethics Committee of the University of Sheffield.

Stimuli

The experimental manipulation of immersion comprised a 'high immersion' head-mounted display (HMD) 360° spherical video (hereafter 'Immersive VR') and 'low immersion' desktop PC (2D) video (hereafter 'Desktop Video') of the same performance of Mozart's 'Agnus Dei', from Mozart's Coronation Mass (mass No.15 in C Major, K.317), scored for soloists, choir and orchestra, and lasting 4 minutes. This duration was sufficient to induce emotions (> 30s) (Garrido, 2014) and avoid adverse physiological reactions (< 20 mins) (Regan, 1995). This performance is in slow (54 bpm) three-four time, with a relatively simple texture and structure featuring solo soprano and orchestra in a cathedral without an audience.

The VE "Mozart 360: Agnus Dei" (Within Unlimited, 2016) places the viewer at a fixed position that follows the movement of a camera amongst the orchestra. The video is 360 degree spherical and the sound, described as '3D spatial audio' is reactive to head

rotation. Participants can interact with the VE via head rotation. The video includes neutrally valenced text overlays that directs the experiencer's attention to particular musicians and compositional processes. An additional VE of 'Google earth VR' (Google, 2016) was used as an introductory activity to reduce novelty effects (Hopp & Gangadharbatla, 2016), and was chosen because it was familiar to participants from Google Street View technology and was unrelated in content to the experimental stimuli.

The 2D music video was created by recording a playback of 'Mozart 360: Agnus Dei' (Within Unlimited, 2016) from the WITHIN mobile application on an iPhone 8 (Software version 12.2). The movement of the camera, determined by the authors, tracked the appearance of the text on the image. The resulting properties of the sound were 108kbps in mono transmitted on two channels with 44.100 kHz sample rate. The total bit rate for the image was 3206kbps at 59.84 frames/second. Unlike the VE version, the sound was headlocked.

Equipment

Stimuli were presented via i) desktop PC (including 17-inch monitor with a 2560 x 1440 pixels resolution, a 60Hz refresh rate and Intel® Xeon® CPU E5-2630 v3 @ 2.40GHz as processor) and, ii) VR HMD with two base stations tracking head movement (HTC vive GPU NVIDIA Quadro K5200, GPU driver 416.78, the custom resolution was 1524x1696 for each eye with 90Hz refresh rate). Music was presented over circumaural headphones (Sony MDR-V150).

Measures

The Goldsmiths Musical Sophistication Index (Gold-MSI) indicates the population sample is representative of the level of musical sophistication in the UK population (Mdn = 71.5, SD = 19.25, range = 40-105) (Müllensiefen et al., 2014). Familiarity with the piece and the performance, and attention to the music were each measured on 5-point rating scales (1 = not at all and 5 = very). Perceived and felt emotions in the music were measured on a 9-point bipolar scale for each emotion and emotion dimension (felt pleasantness, felt tense arousal, felt energetic arousal, perceived pleasantness, perceived tense arousal and perceived energetic arousal) (Schimmack & Grob, 2000). *The ITC-sense of presence inventory* was chosen due to its use in cross-media comparisons (Lessiter et al., 2001). Background information collected included participants' experience with audio media.

Procedure

Participants were tested individually. They were screened for general well-being, completed the background information questionnaire and Gold-MSI. Differences between emotions induced and perceived in music were explained. Participants then spent 3 to 5 minutes in ‘Google Earth VR’ VE presented through a HMD. They walked in a limited room scale and explored the environment using controllers and head motion. Next, they were instructed to focus on the music in the two conditions, to isolate the implications for presentation of the musical performance. Participants experienced the Desktop Video while seated at a desktop screen. Then completed questionnaires on familiarity, attention, emotions and presence. In the Immersive VR condition, participants stood, and explored the VE (presented through a HMD) via head rotation. Afterwards they completed the same questionnaires as for the Desktop Video condition. The order of presentation of the conditions was reversed for half the participants. The same headphones were used in both conditions and the sound level was kept constant. The experiment duration was 45 to 60 minutes.

Results

Group means were calculated for presence’s subscales, emotion felt and perceived, familiarity, and levels of attention. Participants’ familiarity with the composition ($M = 1.96, SD = 1.23$) and the performance ($M = 1.6, SD = 1.07$) were low. Levels of attention to the music were very similar in the Desktop Video ($M = 3.43, SD = 0.83$) and Immersive VR ($M = 3.43, SD = 1.1$) conditions. There were no significant carry over effects. These variables were excluded from further analysis.

Scores were compared to check for differences in the experience of the Desktop Video and Immersive VR versions. Variables were normally distributed except for felt tense arousal, spatial presence (i.e. sense of physical space) and negative effects in the Desktop Video condition. A paired t-test was run to compare the ratings of perceived pleasantness, perceived tense and energetic arousal, felt pleasantness and energetic arousal, ecological validity and engagement in the two conditions (Table1). Wilcoxon signed-rank test compared the ratings of felt tense arousal, spatial presence and negative effects in the two conditions (Table 2).

Ratings were significantly higher in Immersive VR for felt pleasantness $t(27) = 2.09, p = .047, d = .45$, and all measures of presence: engagement $t(27) = 8.15, p$

$<.001, d = 2.01$, spatial presence ($z = -4.47, p < .001, r = -.844$), ecological validity $t(27) = 6.08, p < .001, d = 1.17$, and negative effects ($z = -3.04, p < .001, r = -.57$) (calculated as one-tailed comparisons due to the directional hypothesis). However, there was no significant difference in the ratings of perceived pleasantness $t(27) = 1.78, p = .086, d = .29$, perceived energetic arousal $t(27) = 1.87, p = .073, d = .37$, perceived tense arousal $t(27) = .72, p = .477, d = .15$, felt energetic arousal $t(27) = 1.31, p = .202, d = .35$ nor felt tense arousal ($z = -1.19, p = .234, r = -.22$) between the two conditions.

Table 1: Means and standard deviations of Engagement, Ecological validity, Pleasantness, Energetic arousal and Tense arousal.

Variables	Desktop		VR	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Engagement	2.79	0.70	4.02	0.53
Ecological validity	2.95	0.85	3.82	0.64
Perceived pleasantness	6.36	1.45	6.75	1.24
Perceived energetic arousal	4.39	1.23	4.96	1.82
Perceived tense arousal	4.21	2.04	3.89	2.36
Felt pleasantness	6.54	1.57	7.21	1.42
Felt energetic arousal	4.29	1.44	4.82	1.63

Table 2: Medians and interquartile ranges of Spatial presence, Negative effects, Tense arousal.

Variables	Desktop		VR	
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>
Spatial presence	1.74	1.08	3.44	1.04
Negative effects	1.57	0.83	2.17	1.12
Felt tense arousal	3	2	3	3

Spearman’s correlation test revealed significant associations between factors of presence and induced and perceived emotions in both conditions (Table 3 and 4).

Table 3: Spearman’s correlations between the sense of presence and felt and perceived emotions in the Desktop Video condition.

	Spatial presence	Engagement	Ecological validity	Negative effects
Perceived pleasantness	$r = +.45$ ($p = .016$)	$r = +.63$ ($p < .001$)	$r = +.55$ ($p = .002$)	$r = -.16$ ($p = .425$)
Perceived energetic arousal	$r = +.05$ ($p = .786$)	$r = +.21$ ($p = .273$)	$r = +.28$ ($p = .145$)	$r = -.28$ ($p = .149$)
Perceived tense arousal	$r = -.19$ ($p = .323$)	$r = -.13$ ($p = .494$)	$r = -.11$ ($p = .581$)	$r = +.01$ ($p = .945$)
Felt pleasantness	$r = +.39$ ($p = .04$)	$r = +.66$ ($p < .001$)	$r = +.60$ ($p < .001$)	$r = -.38$ ($p = .047$)
Felt energetic arousal	$r = +.24$ ($p = .222$)	$r = +.38$ ($p = .047$)	$r = +.34$ ($p = .077$)	$r = -.28$ ($p = .145$)
Felt tense arousal	$r = -.19$ ($p = .34$)	$r = -.34$ ($p = .075$)	$r = -.27$ ($p = .162$)	$r = +.31$ ($p = .111$)

Table 4: Spearman’s correlations between the sense of presence and felt and perceived emotions in the Immersive VR condition.

	Spatial presence	Engagement	Ecological validity	Negative effects
Perceived pleasantness	$r = +.21$ ($p = .275$)	$r = +.49$ ($p = .008$)	$r = +.24$ ($p = .229$)	$r = -.28$ ($p = .150$)
Perceived energetic arousal	$r = +.12$ ($p = .555$)	$r = +.16$ ($p = .410$)	$r = +.09$ ($p = .637$)	$r = +.05$ ($p = .792$)
Perceived tense arousal	$r = -.07$ ($p = .740$)	$r = -.05$ ($p = .809$)	$r = -.16$ ($p = .424$)	$r = +.39$ ($p = .042$)
Felt pleasantness	$r = +.33$ ($p = .087$)	$r = +.76$ ($p < .001$)	$r = +.41$ ($p = .030$)	$r = -.44$ ($p = .021$)
Felt energetic arousal	$r = 0$ ($p = .999$)	$r = +.22$ ($p = .267$)	$r = +.02$ ($p = .932$)	$r = -.2$ ($p = .307$)
Felt tense arousal	$r = -.16$ ($p = .406$)	$r = -.26$ ($p = .177$)	$r = -.16$ ($p = .407$)	$r = +.48$ ($p = .011$)

Discussion

The study investigated whether people’s experience of musical multimedia differs with audiovisual formats and their associated levels of immersion. It revealed a significantly higher sense of presence, specifically of engagement and ecological validity, in an Immersive VR musical experience as opposed to the Desktop Video version, suggesting that presence increases with the level of immersion afforded by the technology (H1). This is the first time to our knowledge that this relationship has been empirically confirmed with commercially available musical multimedia.

Level of immersion had no significant effect on the perception nor induction of emotion, with the exception of increased pleasantness reported with the Immersive VR condition (H2). The lack of influence of immersion on other aspects of emotional experience may be due to the low-arousal character of the stimulus (Freeman et al., 2005); it scored a maximum mean rating of 3.89 and 4.96 on a 9-point scale respectively for perceived tense and energetic arousal.

The results confirmed associations between presence, and emotion felt and perceived (H3) for both levels of immersion in this low arousal performance, despite previous suggestions that presence and emotions are only related in the case of arousing stimuli (Freeman et al., 2005). The ‘engagement’ factor of presence was positively correlated with felt and perceived pleasantness in both conditions, indicating its importance regardless of the immersive properties of the technology. ‘Ecological validity’ and ‘physical presence’ seem to play a more important role in the Desktop Video condition: both factors correlated positively with perceived and felt pleasantness, whereas in Immersive VR, ‘ecological validity’ was positively associated with felt pleasantness only. It could be that in media experiences that lack ecological validity and do not elicit strong feelings of presence, participants’ affective experience is related to the resolution of the presentation and how much interaction and control is offered, since these variables influence presence (Lessiter et al., 2001). Finally, the relationship between ‘negative effects’ and emotion seems to be more comprehensive in Immersive VR: as negative effects of Immersive VR increased, so too did perceived and felt tense arousal, and felt pleasantness decreased, whereas negative effects of the Desktop Video condition were associated with a reduction in felt pleasantness only. Further empirical studies are needed. First, both high- and low-arousal stimuli should be included to determine

whether immersion influences other dimensions of emotional experience. Second, features which coincide in the multimedia products (i.e. visual immersion (2D versus 360° video), auditory immersion (mono versus 3D spatialised sound), interactivity (none versus head rotation) and body position (sitting versus standing)) should be investigated together and separately to understand how they interact to influence presence and emotions.

Conclusion

This study provides empirical confirmation that Immersive VR can afford an enhanced sense of ‘being present’ at virtual music performances relative to Desktop Video and suggests that presence is related to users’ emotional experience of musical multimedia. Finally, the study suggests that presence should be considered in theories of musical emotion applied to multimedia contexts.

References

- Coelho, C., Tichon, J. G., Hine, T. J., Wallis, G. M., & Riva, G. (2006). Media presence and inner presence: the sense of presence in virtual reality technologies. *From communication to presence: Cognition, emotions and culture towards the ultimate communicative experience* (pp. 25-45). IOS Press, Amsterdam.
- Freeman, J., Lessiter, J., Pugh, K., & Keogh, E. (2005). When presence and emotion are related, and when they are not. *Proceedings of the 8th annual international workshop on presence (PRESENCE 2005)* (pp. 213-219). International Society for Presence Research.
- Garrido, S. (2014). A systematic review of the studies measuring mood and emotion in response to music. *Psychomusicology: Music, Mind, and Brain*, 24(4), 316. <https://doi.org/10.1037/pmu0000072>
- Google (2016). Google Earth VR [Computer software] Retrieved from : https://store.steampowered.com/app/348250/Google_Earth_VR/
- Hendrix, C., & Barfield, W. (1996). Presence within virtual environments as a function of visual display parameters. *Presence: Teleoperators & Virtual Environments*, 5(3), 274-289. <https://doi.org/10.1162/pres.1996.5.3.274>
- Hopp, T., & Gangadharbatla, H. (2016). Novelty effects in augmented reality advertising environments: The influence of exposure time and self-efficacy. *Journal of Current Issues & Research in Advertising*, 37(2), 113-130. <https://doi.org/10.1080/10641734.2016.1171179>
- Kobayashi, M., Ueno, K., & Ise, S. (2015). The effects of spatialized sounds on the sense of presence in auditory virtual environments: a psychological and physiological study. *Presence: Teleoperators and Virtual Environments*, 24(2), 163-174. https://doi.org/10.1162/PRES_a_00226
- Kober, S. E., & Neuper, C. (2012). Using auditory event-related EEG potentials to assess presence in virtual reality. *International Journal of Human-Computer Studies*, 70(9), 577-587. <https://doi.org/10.1016/j.ijhcs.2012.03.004>
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments*, 10(3), 282-297. <https://doi.org/10.1162/105474601300343612>
- Miller, H. L., & Bugnariu, N. L. (2016). Level of immersion in virtual environments impacts the ability to assess and teach social skills in autism spectrum disorder. *Cyberpsychology, Behavior, and Social Networking*, 19(4), 246-256. <https://doi.org/10.1089/cyber.2014.0682>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: an index for assessing musical sophistication in the general population. *PloS one*, 9(2), e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Nichols, S., Haldane, C., & Wilson, J. R. (2000). Measurement of presence and its consequences in virtual environments. *International Journal of Human-Computer Studies*, 52(3), 471-491. <https://doi.org/10.1006/ijhc.1999.0343>
- Regan, C. (1995). An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality*, 1(1), 17-31. <https://doi.org/10.1007/BF02009710>
- Schimmack, U., & Grob, A. (2000). Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14(4), 325-345. [https://doi.org/10.1002/1099-0984\(200007/08\)14:4<325::AID-PER380>3.0.CO;2-I](https://doi.org/10.1002/1099-0984(200007/08)14:4<325::AID-PER380>3.0.CO;2-I)
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, 6(6), 603-616. <https://doi.org/10.1162/pres.1997.6.6.603>
- Within Unlimited, Inc. (2016). Mozart 360: Agnus Dei. WITHIN [Computer software]. Retrieved from : <https://store.steampowered.com/app/458890/WITHIN>