Head-Mounted Displays as Opera Glasses: Using Mixed-Reality to Deliver an Equalitarian User Experience During Live Events

-
-
-
-

-
-
-
-

-
-
-
-

ABSTRACT

This paper explores the use of head-mounted displays (HMDs) as a way to deliver a front row experience to any audience member during a live event. To do so, it presents a two-part user study that compares participants reported sense of presence across three experimental conditions: front row, back row, and back row with HMD (displaying 360° video captured live from the front row). Data was collected using the Temple Presence Inventory (TPI), which measures presence across eight factors. The reported sense of presence in the HMD condition was significantly higher in five of these, including spatial presence, social presence, passive social presence, active social presence, and social richness. We argue that the non-significant differences found in the other three factors – engagement, social realism, and perceptual realism – are either artefacts of participants' personal taste for the song being performed, or the effects of using a mixed-reality approach. Finally, the paper presents a system description for low-latency, 360° video live streaming using off-the-shelf, affordable equipment and software.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI** • *Human-centered computing* → *Empirical studies in ubiquitous and mobile computing* • Human-centered computing → Displays and imagers

KEYWORDS

Mixed-reality, augmented-reality, virtual-reality, live event, embodied interaction, equality, presence, live performance

1 INTRODUCTION AND RELATED WORK

From the acoustics of ancient Greek and Roman theaters, to Opera glasses, we have always looked for ways to improve our experience of live events. The latter has been popularized in the 1850s, due to the poor lighting conditions of the theaters of the day, and the poor eyesight of some viewers. Nowadays, while still used to improve visual access to plays, Opera glasses are mostly provided as a prop to immerse viewers in the classical theatre experience [17]. They have been replaced in most venues by over-head or in-seat displays, which provide an up-close view of the event, regardless of where one sits in the arena.

What we propose in this paper is to explore use head-mounted displays (HMDs) to provide viewers with not only a closer look of the action, but to make them *feel* as if they are sitting in the front row. This sense of presence is a different proposition from Opera glasses and in-house displays, but it is not completely novel. 360° video and virtual-reality (VR) headsets have been used to immerse audiences in pre-recorded events [18], where they can look around as if they were there. What we suggest instead is the use of VR headsets to consume a *live* event, from within the arena where the event is taking place, but from the perspective of someone sitting in the front row. This is more akin to augmented-(AR) or mixed-reality, where unlike VR, users do not experience a fully virtual environment, but instead augment their experience of the real-world with digital content [9,14]. In the case of a live event, this real-world experience includes not only acoustic and vibrotactile feedback from both the performers and the audience, but also thermal and scent information. Furthermore, there is a collaborative dynamic between a live audience and a performer that a pre-recorded VR experience can hardly emulate [3].

This idea is grounded on a body of work that demonstrates how digital technology can be used to *trick* users' sense of embodiment and presence. Early examples used VR and mixed-reality [5,13] to replicate the "rubber hand illusion" [2], a seminal experiment in which users reported a sense of touch while watching a rubber hand being stroked. Other examples have shown that users can be tricked into believing they have longer arms [6], larger hands [10], or that they are interacting with objects with sizes, weights and materials different from the ones handed to them [12]. Furthermore, mixed-reality has been used to elicit out-of-body experiences [4], and facilitate telepresence and collaboration during work related tasks [1]. What we suggest instead is the use of mixed-reality to explore if users can be given a sense of sitting in the front row of a live show, while sitting far behind.

As such, the contributions of this paper are twofold. First, a two-part study that reports on users' sense of presence while watching a live performance from afar (with and without an HMD connected to a 360° camera positioned up front). And second, a system description that supports low-latency live streaming of 360° content with available, consumer grade equipment.

2 STUDY 1: PRESENCE IN OFFLINE EVENTS

The first study is an early exploration on using mixed-reality to provide a front row experience to all attendees of a live event. A study of this nature has many challenges, including the delivery of a consistent musical experience across participants and conditions, or the live-streaming of 360° video in real-time, with low latency and high video quality. As a first exploratory study, and for the sake of simplicity and quickness, these challenges were avoided by having participants interact with a pre-recorded video clip of a musical performance (recorded in the study venue).

2.1 Participants

20 participants were recruited (9F), aged between 20 and 55 years (*M* = 33.24, *SD* = 12.61). All participants were students or staff at a local institution, and except for four, had minimal experience with head-mounted displays, virtual-reality, or augmented-reality. Participants confirmed having normal hearing for their age, and either did not require vision aids, or used contact lenses.

2.2 Experimental Setup and Design

The experiment was conducted in an empty lecture hall, and followed a within-subjects repeated measures design with three conditions: front row (FR); back row (BR); and back row with head-mounted display (HMD). Participants in the FR and BR conditions watched the same pre-recorded video clip of a musician performing a short musical piece (approximately one minute long). This was recorded in 1080p at 24fps with a standard camera (50mm lens). To emulate a real performance, the clip was projected at floor level, with the approximate proportions of the musician (see Fig. 1). Participants in the HMD conditions watched the same video clip, recorded using a 360° camera in 1080p at 30fps (Ricoh Theta S); and displayed using the Samsung GearVR (Samsung S6). Both standard and 360° recordings happened simultaneously in the same lecture hall where the study took place, with the cameras positioned side-by-side from the second row, at the height of an average person sitting down. In all conditions, sound was played using the lecture hall's sound system. Participants in the FR condition were sited at ~350cm from the projection wall, while participants in the BR and HMD conditions were sited from ~950cm away. Finally, all three

Figure 1: View from the back row condition in Study 1.

conditions were counterbalanced.

2.3 Procedure and Metrics

Participants were given a brief explanation of the study and signed the appropriate consent form and demographics prior to entering the lecture hall where the study took place. At the start of each condition, participants were instructed on where to sit, and in the HMD condition, how to setup the HMD comfortably on their heads. Participants filled in the Temple Presence Inventory (TPI) [8] immediately after each condition. The TPI measures telepresence and is sensitive to media content, and unlike the more commonly used ITC-SOPI [7], also measures various aspects of social presence. At the end of the study, the researcher and participants informally discussed the study experience.

2.4 Results

The TPI measures presence across eight factors: *spatial presence*, *social presence (actor within medium)*, *passive social presence*, *active social presence*, *presence as engagement*, *presence as social richness*, *presence as social realism*, and *presence as perceptual realism*. Each factor was compared across conditions using the Friedman test, followed by Wilcoxon signed-rank tests with Bonferroni correction. Significant differences were found in all factors, due a reported higher sense of presence in the HMD condition when compared to both FR and BR conditions. The only exception are the results for passive social presence, where no differences were found between the HMD and FR conditions – see Table 1.

Table 1: Temporal Presence Inventory results for Study 1, including Friedman tests (all factors with two degrees of freedom). Mean results on a 7-point Likert scale (higher is better), with standard deviation in brackets. Postdoc tests with Wilcoxon signed-rank tests with a Bonferroni correction (significance level set at p < .017). Significant differences in bold.

Factor	First row	HMD	Back row	$\chi2$	p	FR-BR Z, p	FR-HMD Z, p	BR-HMD Z, p
Spatial Presence	2.16 (0.77)	4.77 (1.08)	2.29 (1.23)	27.80	**.000**	-0.18, .861	-3.92, **.000**	-3.66, **.000**
Social Presence (SP)	1.97 (0.80)	4.17 (1.28)	1.79 (0.88)	29.95	**.000**	-1.07, .285	-3.92, **.000**	-3.83, **.000**
Passive SP	4.17 (1.11)	5.15 (1.33)	3.73 (1.38)	6.19	**.045**	-2.53, **.011**	-1.29, .197	-2.65, **.008**
Active SP	1.60 (0.75)	2.65 (1.45)	1.53 (1.04)	11.66	**.003**	-1.66, .097	-2.73, **.006**	-2.91, **.004**
Engagement	2.90 (0.77)	5.13 (1.24)	2.67 (0.90)	27.82	**.000**	-0.99, .324	-3.92, **.000**	-3.88, **.000**
Social Richness	3.77 (0.86	5.44 (0.77)	3.22 (0.83)	23.57	**.000**	-1.93, .053	-3.79, **.000**	-3.89, **.000**
Social Realism	4.58 (1.49)	6.13 (1.00)	4.65 (1.15)	23.01	**.000**	-0.26, .796	-3.63, **.000**	-3.60, **.000**
Perceptual Realism	2.87 (1.21)	3.84 (1.17)	2.60 (0.77)	16.92	**.000**	-1.09, .276	-3.14, **.001**	-3.35, **.001**

3 STUDY 2: PRESENCE IN LIVE EVENTS

Motivated by these results, the second study tackles several of the logistical and technical challenges avoided earlier to examine a more representative experience in which a musician plays live for participants watching from afar, with and without an HMD.

3.1 Participants

10 participants were recruited (3F), aged between 21 and 28 years (M = 24.9, SD = 2.18). All participants were students or staff at a local institution, and except for one, had minimal experience with head-mounted displays, virtual-reality, or augmented-reality. As before, participants confirmed having normal hearing for their age, and either did not require vision aids, or used contact lenses. None of the participants took part in the first study.

3.2 Experimental Setup and Design

The second study was conducted in an empty lecture hall, and followed a within-subjects repeated measures design with two conditions: back row (BR); and back row with head-mounted display (HMD). Participants took part in pairs, watching the live performance side-by-side from the back row (one participant per condition) – see Fig. 2. At the end of the performance participants swapped experimental conditions, and the musician replayed the song. This setup enabled a more consistent experience across participants (as the song was played 10 times, instead of 20); reduced fatigue effects on the performer, and provided a more realistic social experience, as one is not usually alone during live shows. Due to availability issues, a second musician was recruited half-way through the study, performing a different song of the same length (approximately four minutes long).

Participants in the HMD condition used an HTC Vive headset to experience the live performance. This displayed a real-time video feed from a 360° camera (Ricoh Theta S) positioned in the front row of the lecture hall, ~250cm from the performer, and at the height of an average person sitting down. The camera was wired to a computer located behind the participants using a standard 5m USB cable connected to a 10m powered USB extension cable that minimized latency. The video feed was processed using bespoke software programmed in Max [19], which was responsible for not only stitching the 360° video (using the Ricoh's Theta UVC Blender driver), but also displaying it correctly in the HTC Vive headset (using the *worldmaking* Max patch [16]). The headset was wired to the computer, and visuals were rendered at ~90fps. Finally, the musicians' instruments were connected to the lecture hall's speakers to provide real-time audio, independent of experimental condition.

3.3 Procedure and Metrics

Each pair of participants were introduced to the musician and given a brief explanation of the study prior to starting to its start. This explanation included how the HTC Vive works, and how to set it up comfortably. After listening to a live song for the first time, participants were asked to complete the Temple Presence Inventory (TPI) individually. Afterwards, the HTC Vive

Figure 2: A pair of participants taking part in Study 2. Each participant is experiencing a different experimental condition. The participant in the HMD condition is watching a live feed from a 360° camera positioned in the front row.

Table 2: Temporal Presence Inventory results for Study 2, with Wilcoxon signed-rank tests. Mean results on a 7-point Likert scale (higher is better), with standard deviation in brackets. Significant differences in bold.

Factor	HMD	Back row	Z, p
Spatial Presence	5.33 (1.31)	3.46 (1.37)	-2.29, **.022**
Social Presence (SP)	5.04 (1.17)	3.60 (1.64)	-1.99, **.047**
Passive SP	5.83 (1.07)	4.08 (1.65)	-2.26, **.024**
Active SP	4.47 (1.44)	3.03 (1.34)	-2.04, **.042**
Engagement	6.03 (1.20)	4.45 (1.41)	-1.84, .066
Social Richness	5.87 (1.18)	3.91 (1.58)	-2.24, **.025**
Social Realism	6.23 (1.01)	6.30 (1.55)	-0.25, .799
Perceptual Realism	5.48 (0.77)	5.32 (1.38)	-0.24, .813

exchanged hands under the supervision of the researcher, and the musician played the same song again. Finally, participants completed one final TPI, and had an informal group discussion with the researcher about their experience.

3.4 Results

Each TPI factor was compared across experimental conditions using the Wilcoxon signed-rank test. Participants reported a higher sense of presence in the HMD condition for five factors: spatial presence, social presence (including passive and active), and social richness. No differences were found for engagement, social realism, or perceptual realism – see Table 2.

4 DISCUSSION

The results from Study 1 are encouraging, but not surprising. Participants reported a higher sense of presence in the HMD condition for most factors of the TPI. This happened when comparing data from the HMD condition to both the front and back row conditions. Looking at the study setup it is easy to understand why: while participants in the HMD condition watched a 360° video of a musician playing right in front of them, participants in the front and back row conditions watched a

standard video of the same musician displayed on a flat surface. Furthermore, the FR and BR conditions have more in common with the in-house displays available in my many live shows. As such, while not surprising, we argue these results demonstrate convincingly that HMDs and 360° video are better suited to elicit a sense of presence than the equivalent system currently available (flat displays). It is also worth noting that the only non-significant difference between the HMD and front row conditions was in the reported *passive social presence*. In this section of the TPI, participants were asked about observed social cues such as the musician's facial expressions or body language. Naturally, participants in the front row condition had similar responses to those in the HMD condition, and reported a significantly higher sense of passive social presence than those in the back row condition, where these cues are harder to observe.

The results from the Study 2 are more diverse. Participants in the HMD condition reported a higher sense of presence in five of the eight TPI factors. This included *spatial presence* (e.g., how much did it seem as if you could reach out and touch the person you saw/heard?); *social presence* (e.g., to what extent did you feel you could interact with the person you saw/heard?); *passive social presence*; *active social presence* (e.g., how often did you smile in response to someone you saw/heard?); and social richness (e.g., rank the experience from remote to immediate). On the other hand, participants reported very similar results across conditions for *engagement* (e.g., how relaxing or exciting was the experience?); *social realism* (e.g., the events I saw/heard could occur in the real world); and *perceptual realism* (e.g., how much did the people in the environment you saw/heard sound like they would if you had experienced them directly?). The seemingly negative results can be understood by how engagement seems to be directly related to participants' enjoyment of the song; and how the social and perceptual realism questions are not aimed at mixed-reality experiences, where most of what is experienced is real. In sum, these are very positive results that validate HMDs as a viable solution for providing audience members with the sense that they are sitting in the best seats in the house during a live performance.

5   LIMITATIONS

One of the biggest challenges with the study presented in this paper was finding a feasible solution to deliver live, wireless, high-quality 360° video to a VR headset with neglectable delay. Commercial solutions were explored (e.g., YouTube), but these introduced a delay that could reach 20s on an above-average internet connection. Other solutions describe the use of advanced hardware to achieve 360° live streaming with as little as 70ms delay (e.g., three video capture cards with a Genlock synchronization mechanism) [11]. Such hardware was not available to us at the time of the project. As such, we opted for a commercial 360° camera (<$350), a VR-ready desktop computer, and bespoke software that stitched and streamed the video data to the VR headset. This solution had two compromises: it required a wired connection to minimise lag; and because the highest video resolution supported was 1080p (1K), the video quality was

subpar (as evidenced by most 360° videos available on YouTube). Nonetheless, the system presented is a valid contribution to other researchers looking for an available and affordable solution to explore similar scenarios: despite the poor video quality, participants still reported a higher sense of presence in the HMD condition. An additional downside of our solution is that it does not support a multi-user experience, where multiple VR headsets could tap into a single 360° camera in the front row.

We expect these limitations to be mitigated in the near-future. Several 4K and 8K 360° cameras are expected to hit the market in 2017 (~$3500); and as of May 2017, NVIDIA has released their VRWorks 360° video SDK [15], which is aimed at facilitating the stitching and streaming of high-quality 360° video.

6   CONCLUSIONS AND FUTURE WORK

This paper introduced two studies that position HMDs as a viable mixed-reality solution for an age-old problem: the delivery of an equalitarian user experience during live events, where everyone has access to the best seats in the house. Our studies concluded that participants consistently reported a higher sense of presence when watching a live show through an HMD displaying live 360° video, compared to watching the same performance from afar. We envision HMDs being used in the same way as Opera glasses (or 3D glasses for that matter): the theatre makes these available prior to a show for a small fee, and users return them afterwards. Furthermore, future studies could explore the use of multiple 360° cameras to deliver different perspectives in real-time (e.g., the singer's or maestro's perspective). Such studies would have to explore different interaction techniques to facilitate the selection of different 360° cameras, and the effects of switching between these in the users' overall experience.

7   ACKNOWLEDGEMENTS

REFERENCES

[1] M. Billinghurst, I. Poupyrev, H. Kato, and R. May. 2000. Mixing realities in Shared Space: an augmented reality interface for collaborative computing. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No.00TH8532)*, 1641–1644 vol.3. https://doi.org/10.1109/ICME.2000.871085

[2] Matthew Botvinick and Jonathan Cohen. 1998. Rubber hands "feel" touch that eyes see. *Nature* 391, 6669: 756–756. https://doi.org/10.1038/35784

[3] Michael Dowdy. 2007. Live Hip Hop, Collective Agency, and "Acting in Concert." *Popular Music and Society* 30, 1: 75–91. https://doi.org/10.1080/03007760500503459

[4] H. Henrik Ehrsson. 2007. The Experimental Induction of Out-of-Body Experiences. *Science* 317, 5841: 1048–1048. https://doi.org/10.1126/science.1142175

[5] Wijnand A IJsselsteijn, Yvonne A. W de Kort, and Antal Haans. 2006. Is This My Hand I See Before Me? The Rubber Hand Illusion in Reality, Virtual Reality, and Mixed Reality. *Presence: Teleoperators and Virtual Environments* 15, 4: 455–464. https://doi.org/10.1162/pres.15.4.455

[6] Konstantina Kilteni, Jean-Marie Normand, Maria V. Sanchez-Vives, and Mel Slater. 2012. Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion. *PLOS ONE* 7, 7: e40867. https://doi.org/10.1371/journal.pone.0040867

[7] Jane Lessiter, Jonathan Freeman, Edmund Keogh, and Jules Davidoff. 2001. A Cross-Media Presence Questionnaire: The ITC-Sense of Presence Inventory.

Presence: Teleoperators and Virtual Environments 10, 3: 282–297. https://doi.org/10.1162/105474601300343612

[8] Matthew Lombard, Theresa B. Ditton, and Lisa Weinstein. 2009. Measuring Presence: The Temple Presence Inventory. In *Proceedings of the 12th Annual International Workshop on Presence*, 1–15.

[9] Paul Milgram and Fumio Kishino. 1994. A Taxonomy of Mixed Reality Visual Displays (Special Issue on Networked Reality). *IEICE transactions on information and systems* 77, 12: 1321–1329.

[10] Francesco Pavani and Massimiliano Zampini. 2007. The Role of Hand Size in the Fake-Hand Illusion Paradigm. *Perception* 36, 10: 1547–1554. https://doi.org/10.1068/p5853

[11] Rodrigo M. A. Silva, Bruno Feijó, Pablo B. Gomes, Thiago Frensh, and Daniel Monteiro. 2016. Real Time 360° Video Stitching and Streaming. In *ACM SIGGRAPH 2016 Posters* (SIGGRAPH '16), 70:1–70:2. https://doi.org/10.1145/2945078.2945148

[12] Adalberto L. Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional Reality: Using the Physical Environment to Design Virtual Reality Experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3307–3316. https://doi.org/10.1145/2702123.2702389

[13] Filip Škola and Fotis Liarokapis. 2016. Examining the effect of body ownership in immersive virtual and augmented reality environments. *The Visual Computer* 32, 6–8: 761–770. https://doi.org/10.1007/s00371-016-1246-8

[14] Ina Wagner, Wolfgang Broll, Giulio Jacucci, Kari Kuutii, Rod McCall, Ann Morrison, Dieter Schmalstieg, and Jean-Jacques Terrin. 2009. On the Role of Presence in Mixed Reality. *Presence: Teleoperators and Virtual Environments* 18, 4: 249–276. https://doi.org/10.1162/pres.18.4.249

[15] 2016. NVIDIA VRWorks™. *NVIDIA Developer*. Retrieved May 21, 2017 from https://developer.nvidia.com/vrworks

[16] 2017. *Max_Worldmaking_Package: A package for Max/MSP/Jitter to support computational worldmaking*. Computational Worldmaking Lab. Retrieved May 15, 2017 from https://github.com/worldmaking/Max_Worldmaking_Package

[17] Theatregoers. *London Opera Glass Company*. Retrieved May 20, 2017 from http://operaglasses.co.uk/theatregoers/

[18] Virtual Reality Studio Session. Retrieved May 20, 2017 from https://www.edef.co.uk/event/154/

[19] Cycling '74. Retrieved May 15, 2017 from https://cycling74.com/