TeleFest: Augmented Virtual Teleportation for Live Concerts

Jacob Young
jacob.young@vuw.ac.nz
Victoria University of Wellington
New Zealand

Stephen Thompson
stephenthomp09@gmail.com
Computational Media Innovation Centre
Victoria University of Wellington
New Zealand

Holly Downer
holly.downer@vuw.ac.nz
Computational Media Innovation Centre
Victoria University of Wellington
New Zealand

Benjamin Allen
ben.allen@vuw.ac.nz
Victoria University of Wellington
New Zealand

Nadia Pantidi
nadia.pantidi@vuw.ac.nz
School of Design Innovation
Victoria University of Wellington
New Zealand

Lukas Stoecklein
lukas.stoecklein@gmail.com
Computational Media Innovation Centre
Victoria University of Wellington
New Zealand

Taehyun Rhee*
taehyun.rhee@ecs.vuw.ac.nz
Computational Media Innovation Centre
Victoria University of Wellington
New Zealand

Figure 1: TeleFest, our system for livestreaming events using tailored 360° content edited in real time by a producer. (Left): A real live performance streamed live to almost 2,000 people using TeleFest. Three 360° cameras were placed among the stage and crowd, and were livestreamed to YouTube with augmented 3D virtual content to enhance the remote viewing experience. (Right): The resulting livestream.

ABSTRACT

We present TeleFest, a novel system for live-streaming mixed reality 360° videos to online streaming services. TeleFest allows a producer to control multiple cameras in real time, providing viewers with different locations for experiencing the concert, and an intermediate software stack allows virtual content to be overlaid with coherent illumination that matches the real-world footage. TeleFest was evaluated by livestreaming a concert to almost 2,000 online viewers, allowing them to watch the performance from the crowd, the stage, or via a catered experience controlled by a producer in real time that included camera switching and augmented content. The results of an online survey completed by virtual and physical attendees of the festival are presented, showing positive feedback for our setup and suggesting that the addition of virtual and immersive content to live events could lead to a more enjoyable experience for viewers.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools**; **Mixed / augmented reality**; *Web-based interaction.*

*Corresponding Author

IMX '22, June 22–24, 2022, Aveiro, JB, Portugal

ACM ISBN 978-1-4503-9212-9/22/06…$15.00
https://doi.org/10.1145/3505284.3529968

KEYWORDS

extended reality, telepresence, live events, live visual effects

ACM Reference Format:
Jacob Young, Stephen Thompson, Holly Downer, Benjamin Allen, Nadia Pantidi, Lukas Stoecklein, and Taehyun Rhee. 2022. TeleFest: Augmented Virtual Teleportation for Live Concerts. In *ACM International Conference on Interactive Media Experiences (IMX '22), June 22–24, 2022, Aveiro, JB, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3505284.3529968

1 INTRODUCTION

Virtual and *Mixed Reality* (VR/MR) are known for their ability to make users feel as though they are present in another place. Whether it be for socialisation [1], recreation [25, 30], or collaboration [10, 12, 18, 22], VR/MR makes interacting with remote places more engaging by increasing the user's sense of *spatial presence* within them; that is, making them feel as though they really are there rather than viewing the space through some device [17, 19].

Much research has focused on how MR can be used to improve small, intimate experiences such as one-to-one communication. However, there is limited work exploring how it could be used to enhance the remote experience for larger events. With recent advancements in immersive interfaces and telepresence research, MR could allow people to attend such events no matter where they are and have a comparable experience to attending in person when they otherwise couldn't attend due to sickness, inability or unwillingness to travel, or simply due to limited spaces at the venue.

We present TeleFest, a system for allowing remote audiences to virtually attend real events through livestreamed mixed reality video. TeleFest allows 360° video from multiple locations to be streamed over the internet, allowing potentially thousands of users to engage with the content in a more immersive way than provided by conventional video. TeleFest also provides live audio and video (AV) controls to allow switching between multiple cameras placed around the venue, and an interface for augmenting virtual content onto the video in real time to augment and adapt to the live performance on the fly, creating an engaging experience like that shown in Figure 1.

The implementation details of our system are described in full, allowing replication of this experience at future events, as well as the lessons learned after its deployment at a real concert that was physically attended by over 20,000 people and watched virtually by almost 2,000 through the TeleFest system. We also report the results of a survey gathering feedback from the remote attendees of this concert, giving insight into how live capturing, editing, and streaming of MR content can be used in the future to provide an immersive online experience. Our contributions are thus as follows:

- We present the complete pipeline for TeleFest, a novel MR system that allows live concerts to be livestreamed in 360° to an arbitrary number of viewers using camera switching, synchronised audio, and seamless blending of 3D virtual augmentations for an immersive online MR experience
- We detail the successful deployment of our system to capture, edit, control, augment, and immersively livestream a real festival to almost two thousand people
- We discuss the findings of a user survey gathering feedback for TeleFest from virtual attendees of this festival, showing a generally positive response to the experience and providing valuable insight into how such mixed reality hybrid concerts can be improved in the future.

2 RELATED WORK

Concerts are about more than music. What's really important is the atmosphere [13]: the experience of being there, interacting with the music in a way not possible with audio or video alone. This experience can sometimes be the main motivator for attending events [21], with the actual entertainment as an afterthought. We must therefore consider how this experience can be replicated for those unable to attend such events in person.

2.1 Mixed Reality Telepresence

VR has long been researched as a tool for experiencing remote places in real time. A major focus of this research has been to create virtual yet realistic representations of real environments to allow remote users to visit and interact within them [20, 22, 27, 28]. A common focus in such work is to increase users' sense of spatial presence within the space, or their sense of "being there" within it [19], making them feel as though they really belong in that space rather than simply viewing it from outside.

A common method for achieving this sense of presence is by presenting the environment as 360° video [8, 22, 27]. This allows the viewer to freely look around, allowing them to obtain their own viewpoint of the space which is not possible using standard cameras [9]. Jo and Hwang [7] found that such view independence can significantly increase the sense of spatial presence induced in users remotely viewing a space, even with standard video images where no additional content can be seen. Young et al. [27] found that introducing 360° video further increases this sense of presence by allowing remote users to view environments completely independently of the capturing device. Tang et al. [22] further found that this view independence can provide remote users the ability to actively guide their own interactions with the captured area rather than passively view it, resulting in a more immersive and engaging experience.

Such 360° videos have been used in the past to allow for immersively viewing concerts or other events. Perhaps most famously, in 2017 the band Queen recorded a concert using 360° video with the intention that fans watch it through a VR device[1]. A 40-second clip of the concert freely available on YouTube has attracted over 200,000 views at the time of writing, showing an appetite (or at least a curiosity) for 360° concerts. However, such recordings fail to capture the experience of attending such events live [2], suggesting a need to support both immersive viewing and real-time capturing of the event.

2.2 Virtual Augmentations and Rendering

Several events have been broadcast live in 360° in the past: for instance, the opening of the Elbphilharmonie was streamed over YouTube in an attempt to replicate the experience of being at the opening in person [16]. However, live broadcasts can go beyond what is possible in in-person experiences through the augmentation of virtual content. This is commonly used in sports broadcasting

[1] https://www.imdb.com/title/tt11162500/

to show information such as player statistics and ball trajectories, giving viewers at home more information than would be available if they were at the stadium in person [30], or for enhancing storytelling at cultural festivals [25]. In the context of musical events, such effects could be used to add emphasis to certain parts of the music, provide pyrotechnics in a safer manner, or add weather effects such as snow to suit the mood of the music.

Recent research incorporated such augmented content into 360° video, however this must be done in a visually coherent way so as not to compromise the viewer's sense of presence within the presented environment [3]. Ideally, virtual content is rendered with environmental conditions such as shadows [17] or light rays [24] taken into consideration, making virtual content indistinguishable from the real world. This is done by capturing the real world as a 360° image and using this as the light source, giving believable results. Spherical harmonics can also be calculated from a 360° image to produce an efficient and cheap-to-calculate representation of ambient light [14].

One challenge with this is that 360° video is usually recorded at a low dynamic range, causing bright areas to get clipped and leading to darker, less accurate results. This lost information can be approximated using cheap scaling operations [6] or AI based approaches [29]. Although objects can be rendered using a single 360° image, the results will appear less accurate the further away the object is from the capturing 360° camera because reflections and localized lighting will no longer accurately match. One solution to this problem is to use multiple 360° images scattered around the scene [23] and interpolate between them based on the relative position of the virtual objects.

3 TELEFEST SYSTEM DESIGN AND IMPLEMENTATION

TeleFest allows real-world concerts to be viewed immersively through a mixed reality 360° livestream controlled in real time by a producer. The system is designed to take multiple 360° video streams from cameras placed around the stage and crowd and presents them to this producer, who can freely choose which stream is being showed to the remote audience. The producer can also control and place virtual 3D content around the captured space in real time, allowing them to add effects such as weather, pyrotechnics, creatures, and other augmentations to enhance and react to the mood of the performance. These augmentations are realistically lit based on real-world lighting conditions detected from the incoming video streams, seamlessly blending them into their surroundings and making them appear as a real part of the performance.

TeleFest was implemented using the Unity game engine. We built two main interfaces to handle the cameras and asset management, shown in Figure 2; both are broken into multiple windows so that we can take advantage of Unity's window tab system and rearrange the interface as desired. FFmpeg[2] is used for all video and audio related tasks, and is used to decode the incoming video streams from the 360° cameras. Real-world lighting conditions are then detected from this video and used to coherently illuminate any introduced virtual content using the MR360 Unity plugin [17]. The virtual content is then baked into the currently selected camera stream and rendered as an equirectangular texture, which is then re-encoded using FFMPEG and streamed to YouTube. We describe each step of this pipeline in greater detail below.

3.1 360° Video Streaming

We developed our own plugin to use FFmpeg from within Unity. Textures within Unity were able to directly interact with the encoder and decoder, reducing expensive copying between the CPU and GPU. Because we deal with 4K video streams, all video encoding and decoding is done on the GPU using Nvidia's NVENC and NVDEC encoding/decoding features to maintain real-time performance.

One challenge was ensuring all the 360° videos and the audio were correctly synced, given that 360° cameras usually incur several milliseconds to seconds worth of latency. This was solved by syncing all video streams with the audio input, which was given a constant delay to account for latency in the system. The timing on the video streams could then be altered to either drop or delay frames in order to appear better synced with the audio.

Streaming of media over the network is handled using Nginx[3] streaming with RTMP. The Nginx server is set up to provide URLs for pulling the video streams from the 360° cameras and for pushing video streams to YouTube. We also use another URL for forwarding the high quality audio to any other streaming computer. Multiple instances of FFmpeg are run concurrently to supply audio input for the audio URL and for providing the AV syncing and forwarding between the camera streams and YouTube streaming destinations. Video transcoding is avoided when possible to maximise performance. This setup allows us to distribute the workload of streaming and rendering multiple 4K video streams across several computers, and means there are fewer single points of failure in the case that technical issues arise during a live event.

3.2 Real-Time 360° Camera Switching

We built an interface (shown in Figure 2) that allows a producer to manage the camera setup in real time to adapt to the performance. Because only a few objects are rendered at any time, we are able to render two 4K equirectangular videos simultaneously, allowing for a smooth fade between them whenever the camera is changed. This prevents sudden cuts between cameras which could potentially cause a break in presence for users [11].

When switching between the 360° cameras, we ensure that the focus point of the user remains consistent even after switching between cameras. To help maintain this consistency, we added the ability to define a focal point along the x-axis of the equirectangular video. Switching between cameras will shift the equirectangular 360° video along the x-axis to match the focal point of the previously displayed 360° video.

3.3 Virtual Augmentations in 360° Videos

Environment maps for reflections and spherical harmonics for ambient lighting are calculated each frame for each 360° camera to allow seamless blending of virtual objects into their real-world surroundings. Each object is assigned environment maps and ambient light calculated from the closest 360° camera. Previous work [17] can automatically detect light sources in order to provide high frequency

[2]https://ffmpeg.org/

[3]https://www.nginx.com/

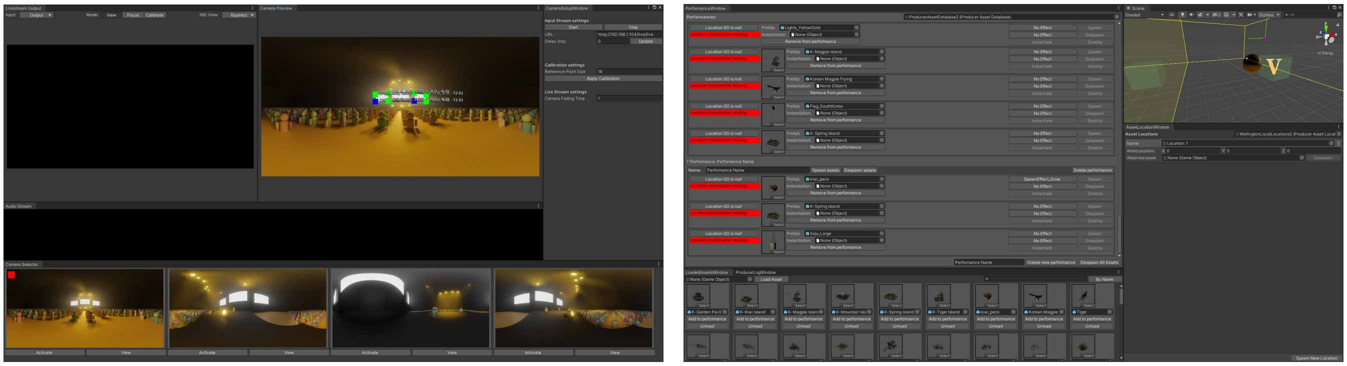


Figure 2: The interfaces used by the producer to control the livestream. (Left): The window for calibrating and switching between camera streams. The top frame shows the current camera stream in high detail, while thumbnails of all four streams are shown along the bottom edge of the display. (Right): The window for controlling augmented content. Individual assets can be added and moved around the scene in a preview window before being published to the livestream.

lighting with shadows; two limitations of this are that 1) lights are assumed to be directional, and 2) because the 360° video has low dynamic range and thus loses information, any over-exposed area can be assumed to be a light source, resulting in occasionally incorrect light detection and shadow casting.

Because the setup for TeleFest is assumed to be indoors with very localised lighting, virtual objects should be lit differently depending on which area of the room that object is located. We adapt the light detection of MR360 [17], opting for a more manual setup for light detection in order to ensure stable and accurate light detection. Areas where physical lights are located are manually annotated. Because we use multiple 360° cameras, we distributed our light detection across all cameras to cover a broader range of locations. Lights are set as spotlights directed at the 360° cameras and the positions are calculated using a combination of the position of the light annotation as defined by the user and a user defined distance away from the 360° camera. Lighting colour and intensity calculation is still calculated the same way as in MR360 [17].

3.4 Scene Calibration

All cameras are calibrated to find their real-world position relative to their surroundings so that virtual content appears anchored to the physical world when switching between 360° cameras. Each camera is calibrated independently so that if a camera needs to be replaced during an event it can be reset and re-calibrated without impacting the other live-streaming cameras.

Our calibration method relies on manual identification of 3D points in the real world. During testing of TeleFest at real events, we achieved good calibration results by using the floor plan for the venue as a guide to approximate the placement of 3D points (seen in Figure 3). For each camera, these points are then matched with their corresponding location in the 360° video.

Our process for matching between the 2D and 3D points is performed using raycasting and relies on several assumptions. First, given the correct 360° camera rotation, all rays should converge at the location of the 360° camera. Second, each ray should intersect their respective 3D real-world point. Because the position of the 360° camera is not known, we can instead use the 3D real-world points as origins for each respective ray. The direction for each ray in relation to all other rays should remain consistent because the rays are generated from an equirectangular image.

The calibration can then be broken down to two further steps: a coarse estimate followed by a refinement step. For the coarse estimation, the equirectangular camera image is "rotated" around all axes at fixed angles. We used a step size of 45° because it easily divides into 360°, was a small enough number to not overstep the solution, and was a large enough number to not result in redundant computation. For each step, the amount of error in the estimation is calculated as the sum of distances between the point closest to the rays and the rays themselves. If any of the rays are pointing towards the 360° camera then the rotation is rejected as this scenario can result in a false positive result. After testing all possible coarse rotations, the rotations with the lowest error is chosen as the correct approximate camera rotation.

The camera rotation then is further optimised through a final refinement step. This is done by rotating the camera around each axis by small steps, halving and inverting the direction of the step whenever the amount of error grows. The error is calculated in the same way as the rough estimation step. This refinement process is repeated until the estimation error is less than a predefined threshold or the number of iterations reach the allowed maximum number. Then, the rotation of the 360° camera is found, and the point closest to where the rays converge is assumed to be the 360° camera's location. A limitation of this algorithm is that it requires a good distribution of points in the 360° video. If the distribution is too small, the problem becomes poorly defined and can have multiple solutions.

3.5 Asset Management

To create the MR content, virtual assets need to be placed around the physical venue by the producer during the event to adapt to the live performance. TeleFest supports two types of assets - visual effects and 3D models. Visual effects are particle based visuals, which can include effects such as fireworks, snow, fire flies, and confetti. 3D models can include animated models, such as animals

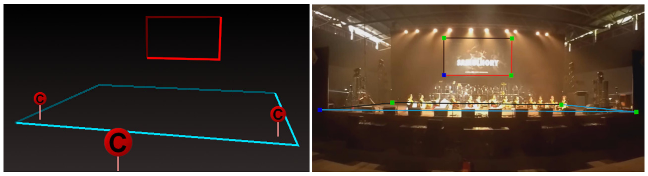


Figure 3: The manual calibration process for aligning the positions of virtual augmentations between multiple cameras. (Left): 3D points are found in relation to the camera's real-world positions; in this case we use the corners of the stage and the large display behind the performers. (Right): These points after being matched to the equirectangular image. This is done independently for each camera.

Figure 4: Examples of virtual augmentations used during our live performance test. (Left): Virtual fireworks are let off to add emphasis to certain parts of the music. (Right): Virtual wildlife is introduced to add a serene feeling to the music. Their red hue illustrates how the virtual lighting is based on real-world conditions, in this case a heavy use of red stage lights.

and flags, or static objects such as buildings or trees. Examples of both can be seen in Figure 4.

One challenge with placing assets in real-world scenes is handling depth occlusion between virtual and physical objects. Without live depth information in the 360° video, assets need to be placed carefully to avoid incorrect occlusion with physical objects. We thus carefully designed our assets to avoid critical occlusion issues. For example, we designed floating islands that assets could be placed on: this way, the assets can appear to be floating over any potential sources of occlusion (eg. the crowd) in a believable way while avoiding potential breaks in presence.

All assets are controlled by the producer through a custom-made interface within the Unity editor. During the concert, assets can be placed into the scene, and the producer has the ability to adjust and preview their placement before they are published to the livestream.

TeleFest currently supports two ways to spawn assets into the livestream depending on the type of assets. Visual effects can be spawned and despawned gracefully by gradually fading in and out respectively. 3D assets likewise grow into or shrink out of vision. Alternatively, assets can instantly appear or disappear if required. These assets are locked to unique, predefined locations in the scene, but those locations can also be moved if required. When choosing which assets to place in the scene, the producer can list these by performance prior to the start of the event.

4 LIVE DEPLOYMENT OF TELEFEST

TeleFest was used to livestream the K Festival[4], a cultural festival physically attended by thousands of people. The livestream lasted approximately three hours, consisted of several live musical performances of varying genres, and was watched live by 1,908 people via the livestream.

4.1 Technical Setup

Four camera streams were available to watch via YouTube or a dedicated website[5] (see Figure 5): one on each corner of the stage, one in the crowd 10m from the stage, and one that utilised our proposed system to switch between these cameras and overlay 3D visual effects, as shown in Figure 6. These effects and the camera switching were controlled by a producer in real time. We used three Insta360 Pro 2 cameras to capture the 360° video; these cameras are also capable of capturing audio, but the sound quality proved inadequate for a musical performance so for the best audio quality we manually synced the audio from the sound crew with our 360° videos. This unfortunately meant spatial audio was not possible as the sound desk used a stereo mix.

Ideally our 4k video would be streamed at 40Mbps or more to ensure high visual quality, but electrical interference with the ethernet cables used during the concert resulted in a high loss of quality and an unstable connection, so we had to run the four video streams at only 15Mbps.

[4]https://www.kfestival.co.nz/
[5]https://xr-hotspot.live/

5 EVALUATION

To gather feedback on our experimental setup we offered an online survey that physical and virtual attendees of the festival could complete. Those that did were put in the draw to win a $100, $50, or $25 supermarket voucher. Questions included general demographics and questions evaluating how respondents enjoyed various aspects of the festival, with separate questions presented to the participant based on whether they watched the concert in person, virtually through their browser, or virtually in VR. Virtual attendees were asked about their impression of the livestreaming setup, and were administered a slightly altered *iGroup Presence Questionnaire* (IPQ) [19] using a 7-point Likert scale to determine the degree to which they felt spatially present at the festival.

We had two hypotheses:

- (H1): There would be no significant difference in enjoyment between physical and virtual attendees of the festival
- (H2): Virtual attendees would feel spatially present at the festival

Due to the nature of the live event we were unable to have a control group for the experiment without unfairly limiting the experience for some attendees. As we had no point of comparison for the spatial presence scores, we consider it sufficient for H2 to be proven if the IPQ scores average above the midpoint (> 4).

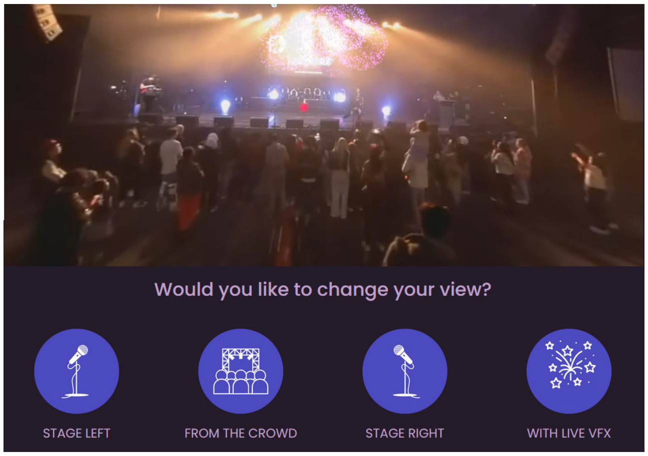


Figure 5: The web interface for watching and switching between the four 360° live streams.

5.1 Results

1,908 people watched the livestream in total. The most popular stream was the crowd camera which was viewed by 1,019 people (53%). This was followed by the live VFX stream with 448 viewers, stage right with 238 viewers, and stage left with 203 viewers.

47 people completed the survey: 34 of these attended physically and 13 virtually. 33 of these respondents were female, 14 were male, and none were gender diverse. Of the 13 virtual attendees only 4 watched the livestream through a VR headset. Due to this low engagement with the survey all results presented here are unfortunately only indicative.

As shown in Figure 7, nine of the thirteen virtual attendees watched the festival alone, while only four of the 34 physical attendees did the same. Families were more popular than friend groups, with 41% and 38% of physical attendees attending with family or friends respectively, and with 8% and 15% of virtual attendees watching with family or friends respectively.

Participants were asked their overall enjoyment of the festival on a scale from 1 to 7. A Wilcoxon Rank Sum Test ($\alpha = 0.05$) found no significant difference in enjoyment between physical and virtual attendees ($p = 0.84$), with physical attendees giving an average score of 5.58 ($\sigma = 1.16$) and virtual ones giving an average score of 5.64 ($\sigma = 1.22$).

Virtual attendees overall tended to feel spatially present at the festival, as shown in Figure 8, with IPQ scores averaging 4.27 ($\sigma = 0.23$) which a Wilcoxon Signed Rank Test showed to be significantly above the midpoint ($p = 0.01$). This was due to high scores in the general ($\mu = 5$, $\sigma = 1.26$) and spatial presence ($\mu = 4.52$, $\sigma = 0.39$) subscales, which also both scored above the midpoint ($p = 0.04$, $p = 0.01$, respectively), and were enough to counteract low scores in the involvement ($\mu = 3.79$, $\sigma = 0.95$) and realness ($\mu = 4.33$, $\sigma = 1.04$) subscales, neither of which were significantly different from the midpoint ($p = 0.54$, $p = 0.45$, respectively).

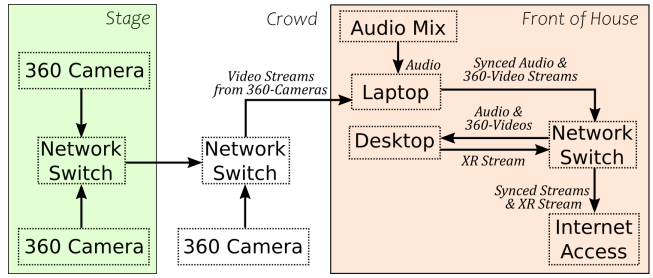


Figure 6: The setup for the cameras, audio, PCs and internet, as used during the live event. Audio was imported into the laptop by using an interface to convert XLR to USB. The other connections shown here all represent Ethernet cables.

6 DISCUSSION

Our system and the subsequent livestream were overall very well received. Remote attendees stated how they enjoyed being able to watch the event live even though they couldn't be there in person and would have loved for the whole day to have been live-streamed:

> *"I wasn't able to attend in person so being able to watch it on my computer was great! I do wish I could've seen other performances that weren't a part of the virtual festival"*

6.1 Enjoyment

Our first hypothesis was that there would be no significant difference in enjoyment between physical and remote attendees, or in other words, that attending the concert virtually would be just as enjoyable as the real thing. Our survey results showed no significant difference between the two attendance methods, though due to the low number of respondents our quantitative data on the matter is unfortunately inconclusive.

Qualitative feedback, however, was overall very positive. Specific positive comments were received with regards to the quality of the sound, the 360° views, the VFX visualisations, and their synchronisation with the performances:

> *"I like the sound is clear, 360 degree view."*
> *"I liked the XR visualizations, and the 360 degree cameras."*
> *"I liked the visualizations which were in time with the performances."*

The main complaint with the livestream was the low video quality, particularly when zooming in; this is despite the video being native 4K throughout the streaming pipeline. We found that YouTube's compression algorithm was particularly harsh here, as suggested by a visual comparison with pre-compressed video, as it removed many small details and in particular the performers' faces. This could be due to the low lighting conditions, possible bandwidth issues from the venue's network connection, or a lack of optimisation on YouTube's part for 360° content.

Viewers appreciated the addition of virtual content to the 360° video stream. Many cited this as their favourite part of the concert, and only one complained that the XR lighting sometimes failed to match the real world:

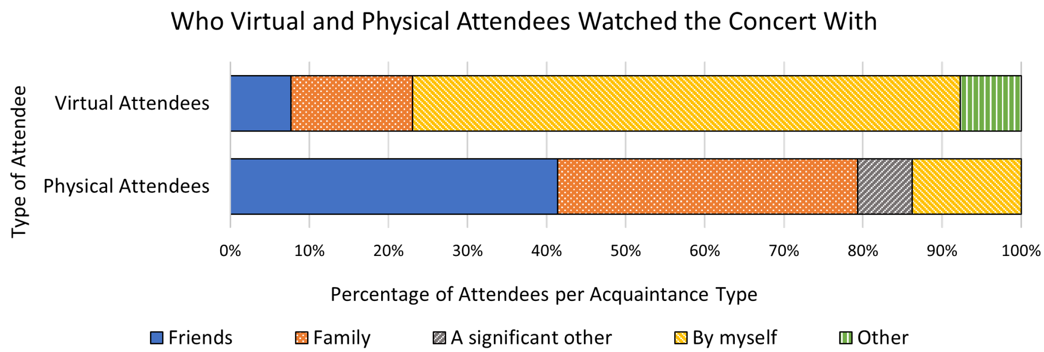


Figure 7: Who survey respondents attended the festival with, separated by physical and virtual attendees. Virtual attendees overwhelmingly tended to watch the concert alone, while physical attendees tended to go with family or friends.

> *"The VFX were nice, but at some times, because the background video was dark, the blending of the 3d objects was not as good as it would be in outdoor lighting."*

This was an issue with the low lighting in the venue sometimes making the virtual content difficult to see. In future we may test the idea of adding ambient light to virtual content whenever insufficient light is detected in the 360° video which would solve this at the cost of visual consistency.

There were surprisingly no complaints about the lack of socialisation options provided by the live stream, despite this being a major focus of festival attendance [13, 21]. This may be due to the virtual attendees' overwhelming tendency to watch the festival alone; it could be that the virtual livestream attracted those that may not wish to attend a festival because of this socialisation aspect, and who may have appreciated the opportunity to watch the concert without a crowd. It could also be that they already weren't expecting a social situation when they began watching due to this isolation; unfortunately none of the survey respondents provided any insight here.

6.2 Induced Presence

Our second hypothesis was that presenting the concert live in 360° would induce a sense of spatial presence within viewers. This was partially confirmed by our data, which showed a mean IPQ score significantly above the midpoint, however the low number of survey respondents means that this data is unfortunately only indicative.

When participants were asked "what, if anything, made them feel as if there were there", the 360° camera viewing accounted for the majority of the responses, with specific mentions around the ability of being able to look around and see other people:

> *"The 360 cameras were great to be able to look around me." "I like the 360 degree viewing (...) Seeing all the people there." "The ability to see 360"*

This aligns with prior research that suggests the ability to obtain independent viewpoints is a large contributor to the sense of spatial presence within remote environments [7, 27].

Involvement was the lowest-scoring of the IPQ subscales, which in this case is probably not too surprising. While virtual attendees could choose to view the concert from three separate viewpoints, the interface to do so was slightly unwieldy due to our reliance on YouTube for the networking infrastructure, and the locations were fixed and discrete. Respondents requested the ability to freely move around the concert without being restricted to these predefined locations, noting that the current experience *"limited [their] experience"*, and another felt that *"being stuck to the three cameras"* prevented them from feeling present at the venue. Despite this, viewers still appreciated the ability to change their viewpoint, stating that they *"loved being able to watch live and move the cameras and pick and choose from which angle I watched"*.

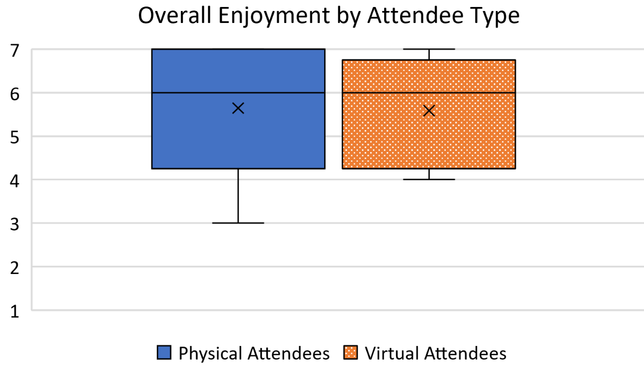
This low score for involvement may have also been exacerbated by the low incidence of VR use, resulting in low scores for the questions "Did you still pay attention to your real environment?" and "Were you fully captivated by the festival?". Half of the respondents who watched the livestream in their browser stated they would be interested in using a VR headset, provided they had one, to watch the event. Several reasons were offered for this, such as they felt it would make *"for a more immersive experience"* or that they just wanted to try it on as they have never had that opportunity. However, when asked whether they consider VR a promising technology for hosting events such as music festivals, some respondents noted that they still see value in other non-VR ways of watching live events such as Facebook live or traditional broadcasts.

6.3 Lessons Learned

This test provided us with valuable insight into how future events could be streamed using an immersive XR platform. Here we outline the lessons we learned throughout this process in the hopes that they assist future researchers and event organisers.

One of our main issues was the incompatibilities experienced between the conventional stage setup and ours. An example of this was that the 360° camera placed at stage left consistently had its view blocked by a cameraman with a conventional camera; this rendered that camera's view almost useless, and consequently it had the lowest viewer count of the four streams.

The setup required for the conventional concert also limited how the 360° cameras could be placed in other ways. The ideal camera placement would have been at the front-centre of the stage,



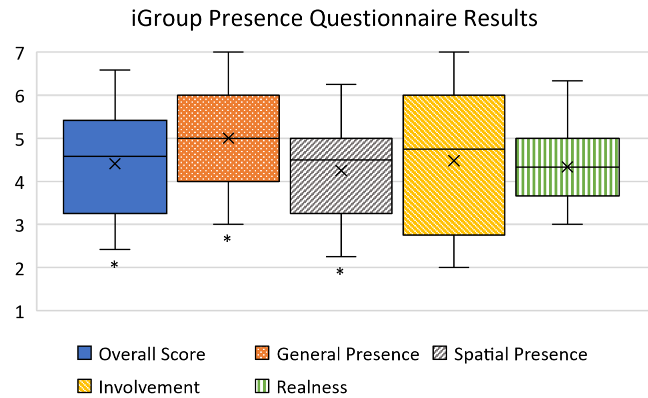


Figure 8: Quantitative results from our survey of concert attendees. (Left) Overall enjoyment of virtual and physical attendees. (Right): The results of the iGroup Presence Questionnaire separated by subscale. * = significantly above the midpoint, indicating generally high scores for that category.

ensuring that all performers could be seen from a natural angle. However, this would have blocked the views of both the performers and the physical crowd. We attempted to get around this by placing the camera on a 2.5m-high platform, however venue restrictions meant that this had to be placed at least ten metres from the stage which made making out the performers difficult once the video had been compressed by YouTube. This was the reason for us placing a camera in each corner of the stage, however these also proved too far from the performers to make out small details such as facial expressions during the performance.

We also experienced issues with the inconsistent lighting within the venue. Despite having six discrete lenses, the Insta360 Pro2 is only capable of adjusting to a single exposure value. As the stage was always brightly lit, and the crowd dimly lit in comparison, only one could be in focus at a time: increasing the exposure to make the crowd visible washed out the performers, and reducing the exposure to make the performer visible made the crowd too dark to see. This was exacerbated in performances with rapidly flashing lights, which often meant constant adjustments to the exposure were needed. Since the virtual content was lit according to the lighting conditions detected in the 360° video, this also meant that the 3D assets were often too dark to see as well.

The final issue experienced with the venue setup was the lack of audio captured from the crowd. As our audio was taken from the sound desk, it only included audio from the on-stage microphones. Several of the performances involved crowd participation, however due to this the crowd's responses to the performer weren't included in the livestream which led to several reported breaks in presence from virtual attendees. At future events we will consider the addition of sound from the 360° camera placed in the crowd to avoid this.

6.4 Future Work

While the survey gathered valuable feedback on the system as a whole, we plan on investigating particular elements of it in future to see how much each individual component contributed to participants' overall enjoyment. A study with conditions that study 360° video, virtual augmentations, and camera switching in isolation could be extremely valuable in determining which features contribute to providing an engaging and entertaining experience for viewers.

It may also be worth considering some of the feedback provided from respondents who attended the event in person. A large number of respondents stated enjoying the food, the merchandise, and especially the 'freebies' in addition to the musical performances:

> *"We were so happy to received lots of freebies."*
> *"I liked the variety of food. The various business stalls were good I especially enjoyed the ones where you could buy merch."*

Having giveaways, being able to buy merchandise, or meeting new people are often pivotal parts of the experience that can turn a good event into a great one [21]. Consequently, it is worth considering how such 'extras' can be translated or replicated for remote attendees. In this respect, including some digital 'freebie' or digital merchandise that can be unique to remote attendees can serve the purpose of a souvenir and be of sentimental value to attendees, but we could also go a step further and represent an ownership status that can be of reputational or transactional value similar to non-fungible token (NFT) art certificates.

Similarly, it may be worth looking into integrating multi-sensory aspects into the live experience to further give the impression of a festival atmosphere [5, 15]. Being able to replicate and deliver to the remote attendees the smells and even perhaps the room temperature of an event could provide a sensory authenticity that has been found to make people's experiences more memorable [26].

Finally, our reliance on YouTube for streaming meant that there were few opportunities for the virtual attendees to interact with the festival or with each other. This was a conscious tradeoff as using YouTube greatly increased the festival's reach by utilising a widely available and widely known platform that most people are already comfortable in using, but also meant we were limited to the functionality it already provides. In future we may consider developing our own streaming platform, which would allow for a greater degree of autonomy within the MR environment and enabled shared interpersonal experiences such as was provided by Hamilton et al. [4], hopefully convincing virtual attendees to watch the performance with others. This could also allow the system to be extended beyond 360° videos and reintroduce the third dimension, either through non-stationary cameras or synthesised 3D environments, and allow attendees to freely move around the venue.

7 CONCLUSION

We presented TeleFest, a solution for live-streaming events in an immersive way. Viewers can switch between multiple 360° camera streams to select their view of the venue, and intermediate software allows virtual content to be seamlessly augmented onto the live footage to enhance the performance.

TeleFest was tested by livestreaming a real cultural festival to 1,908 people over three hours. Although this test was successful with largely positive feedback from viewers, there were many areas that could be improved in the deployment of our streaming application. Incompatibility with traditional stage setups tended to be a consistent problem, resulting in awkward viewing angles, obstructed cameras, and sub-optimal lighting for the virtual content.

Despite this, viewer feedback was overall positive, with many enjoying the ability to freely choose their view of the performance and enjoying the addition of virtual content. Virtual attendance was subsequently just as enjoyable as attending in person, and remote attendees indicated that they felt physically present at the event despite a lack of true control, interaction, or socialisation with the physical attendees. Virtual attendance tended to be a solitary experience, in stark contrast to the family- and friend-focused physical experience, though viewers seemed to have a good time nonetheless. We hope that the lessons we learned through the development and deployment of this system at a live concert prove useful for future research, and pave the way for similar immersive events in the future.

ACKNOWLEDGMENTS

This project was supported by Victoria University of Wellington, K-Festival, and the Entrepreneurial University Programme by the Tertiary Education Commission in New Zealand. We appreciate K-Festival organisers, staff, volunteers, stage crews, performers, audiences (both in-person and online), and CMIC staff and students who supported this project. Special thanks to Gio Jin, Siyun Thompson, Lohit Petikam, Andrew Chalmers, and Chris Dean.

REFERENCES

[1] Andrew Au and Jie Liang. 2012. Ztitch: A Mobile Phone Application for Immersive Panorama Creation, Navigation, and Social Sharing. In *2012 IEEE 14th International Workshop on Multimedia Signal Processing, MMSP 2012 - Proceedings*. IEEE, Banff, AB, Canada, 13–18.

[2] Steven Caldwell Brown and Don Knox. 2017. Why Go to Pop Concerts? The Motivations Behind Live Music Attendance. *Musicae Scientiae* 21, 3 (2017), 233–249.

[3] Jonny Collins, Holger Regenbrecht, and Tobias Langlotz. 2017. Visual Coherence in Mixed Reality: A Systematic Enquiry. *Presence* 26, 1 (2017), 16–41.

[4] William A Hamilton, John Tang, Gina Venolia, Kori Inkpen, Jakob Zillner, and Derek Huang. 2016. Rivulet: Exploring participation in live events through multi-stream experiences. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, New York, NY, USA, 31–42.

[5] Daniel Harley, Alexander Verni, Mackenzie Willis, Ashley Ng, Lucas Bozzo, and Ali Mazalek. 2018. Sensory VR: Smelling, Touching, and Eating Virtual Reality. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, New York, NY, USA, 386–397.

[6] Thomas Iorns and Taehyun Rhee. 2015. Real-time image based lighting for 360-degree panoramic video. In *Pacific-Rim Symposium on Image and Video Technology*. Springer, Cham, Auckland, New Zealand, 139–151.

[7] Hyungeun Jo and Sungjae Hwang. 2013. Chili: Viewpoint Control and On-Video Drawing for Mobile Video Calls. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*. ACM Press, New York, New York, USA, 1425. http://dl.acm.org/citation.cfm?id=2468356.2468610

[8] Shunichi Kasahara and Jun Rekimoto. 2015. JackIn head: Immersive visual telepresence system with omnidirectional wearable camera for remote collaboration. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST* 13-15-November-2015 (2015), 217–225.

[9] Gun A. Lee, Theophilus Teo, Seungwon Kim, and Mark Billinghurst. 2019. A User Study on MR Remote Collaboration Using Live 360 Video. In *Proceedings of the 2018 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2018*. IEEE, Munich, Germany, 153–164.

[10] Stephan Lukosch, Ronald Poelman, Oytun Akman, and Pieter Jonker. 2012. A Novel Gesture-based Interface for Crime Scene Investigation in Mediated Reality. In *CSCW 2012: Exploring collaboration in challenges environments : From the car to the factory and beyond*. ACM, Seattle, Washington, USA, 3–6. http://resolver.tudelft.nl/uuid:b9a2b407-a286-41d9-9781-956cd96b3c4e

[11] Liang Men, Nick Bryan-Kinns, Amelia Shivani Hassard, and Zixiang Ma. 2017. The Impact of Transitions on User Experience in Virtual Reality. In *2017 IEEE virtual reality (VR)*. IEEE, Los Angeles, CA, USA, 285–286.

[12] Thammathip Piumsomboon, Gun A. Lee, Andrew Irlitti, Barrett Ens, Bruce H. Thomas, and Mark Billinghurst. 2019. On the Shoulder of the Giant: A Multi-Scale Mixed Reality Collaboration with 360 Video Sharing and Tangible Interaction. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM, New York, NY, USA, 1–17.

[13] Richard Prentice and Vivien Andersen. 2003. Festival as creative destination. *Annals of Tourism Research* 30, 1 (2003), 7–30. https://doi.org/10.1016/S0160-7383(02)00034-8

[14] Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. ACM, New York, NY, USA, 497–500.

[15] Nimesha Ranasinghe, Pravar Jain, Shienny Karwita, David Tolley, and Ellen Yi-Luen Do. 2017. Ambiotherm: Enhancing Sense of Presence in Virtual Reality by Simulating Real-World Environmental Conditions. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1731–1742.

[16] Simon Rein. 2017. The opening concert of the Elbphilharmonie Hamburg live and in 360°. https://www.blog.google/around-the-globe/google-europe/opening-concert-elbphilharmonie-hamburg-live-and-360/. Accessed: 2021-11-02.

[17] Taehyun Rhee, Lohit Petikam, Ben Allen, and Andrew Chalmers. 2017. MR360: Mixed Reality Rendering for 360° Panoramic Videos. *IEEE Transactions on Visualization and Computer Graphics* 23, 4 (4 2017), 1379–1388. https://doi.org/10.1109/TVCG.2017.2657178

[18] Taehyun Rhee, Stephen Thompson, Daniel Medeiros, Rafael Dos Anjos, and Andrew Chalmers. 2020. Augmented virtual teleportation for high-fidelity telecollaboration. *IEEE transactions on visualization and computer graphics* 26, 5 (2020), 1923–1933.

[19] Thomas Schubert, Frank Friedmann, and Holger Regenbrecht. 2001. The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments* 10, 3 (2001), 266–281. http://www.mitpressjournals.org/doi/10.1162/105474601300343603

[20] Rajinder S. Sodhi, Brett R. Jones, David Forsyth, Brian P. Bailey, and Guiliano Maciocci. 2013. BeThere: 3D mobile collaboration with spatial input. In *Conference on Human Factors in Computing Systems - Proceedings*. ACM, New York, NY, USA, 179–188. http://dl.acm.org/citation.cfm?doid=2470654.2470679

[21] Alina Striner, Sarah Halpin, Thomas Röggla, and Pablo Cesar. 2021. Towards Immersive and Social Audience Experience in Remote VR Opera. In *IMX 2021 - Proceedings of the 2021 ACM International Conference on Interactive Media Experiences*. ACM, New York, NY, USA, 311–318.

[22] Anthony Tang, Omid Fakourfar, Carman Neustaedter, and Scott Bateman. 2017. Collaboration in 360° Videochat: Challenges and Opportunities. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. ACM, Edinburgh, United Kingdom, 1327–1339. https://doi.org/10.1145/3064663.3064707

[23] Joanna Tarko, James Tompkin, and Christian Richardt. 2019. Omnimr: Omnidirectional mixed reality with spatially-varying environment reflections from moving 360 video cameras. In *2019 IEEE Conference on Virtual Reality and 3D*

User Interfaces (VR). IEEE, Osaka, Japan, 1177–1178.

[24] Stephen Thompson, Andrew Chalmers, and Taehyun Rhee. 2019. Real-time mixed reality rendering for underwater 360 videos. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Beijing, China, 74–82.

[25] Alexander Wiebel, Sascha Keuchel, Magnus Eckenfels, Jan Drengner, and Werner König. 2020. Challenges of and Recommendations for Combining 6-DOF Spatial VR-Interaction with Spherical Videos. https://www.techrxiv.org/articles/preprint/Challenges_of_and_Recommendations_for_Combining_6-DOF_Spatial_VR-Interaction_with_Spherical_Videos/11525586/1

[26] Johan Willander and Maria Larsson. 2006. Smell Your Way Back to Childhood: Autobiographical Odor Memory. *Psychonomic bulletin & review* 13, 2 (2006), 240–244.

[27] Jacob Young, Tobias Langlotz, Matthew Cook, Steven Mills, and Holger Regenbrecht. 2019. Immersive Telepresence and Remote Collaboration using Mobile and Wearable Devices. *IEEE Transactions on Visualization and Computer Graphics* 25, 5 (2019), 1908—-1918.

[28] Jacob Young, Tobias Langlotz, Steven Mills, and Holger Regenbrecht. 2020. Mobileportation: Nomadic Telepresence for Mobile Devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–16. https://doi.org/10.1145/3397331

[29] Jinsong Zhang and Jean-François Lalonde. 2017. Learning high dynamic range from outdoor panoramas. In *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, Venice, Italy, 4519–4528.

[30] Stefanie Zollmann, Tobias Langlotz, Moritz Loos, Wei Hong Lo, and Lewis Baker. 2019. ArSpectator: Exploring Augmented Reality for Sport Events. In *SIGGRAPH Asia 2019 Technical Briefs, SA 2019*. ACM, New York, NY, USA, 75–78.