

Cinematic Virtual Reality: Evaluating the Effect of Display Type on the Viewing Experience for Panoramic Video

Andrew MacQuarrie, Anthony Steed*

Department of Computer Science, University College London, United Kingdom



Figure 1: A 360° video being watched on a head-mounted display (left), a TV (right), and our SurroundVideo+ system (centre)

ABSTRACT

The proliferation of head-mounted displays (HMD) in the market means that cinematic virtual reality (CVR) is an increasingly popular format. We explore several metrics that may indicate advantages and disadvantages of CVR compared to traditional viewing formats such as TV. We explored the consumption of panoramic videos in three different display systems: a HMD, a SurroundVideo+ (SV+), and a standard 16:9 TV. The SV+ display features a TV with projected peripheral content. A between-groups experiment of 63 participants was conducted, in which participants watched panoramic videos in one of these three display conditions. Aspects examined in the experiment were spatial awareness, narrative engagement, enjoyment, memory, fear, attention, and a viewer's concern about missing something. Our results indicated that the HMD offered a significant benefit in terms of enjoyment and spatial awareness, and our SV+ display offered a significant improvement in enjoyment over traditional TV. We were unable to confirm the work of a previous study that showed incidental memory may be lower in a HMD over a TV. Drawing attention and a viewer's concern about missing something were also not significantly different between display conditions. It is clear that passive media viewing consists of a complex interplay of factors, such as the media itself, the characteristics of the display, as well as human aspects including perception and attention. While passive media viewing presents many challenges for evaluation, identifying a number of broadly applicable metrics will aid our understanding of these experiences, and allow the creation of better, more engaging CVR content and displays.

Keywords: Cinematic virtual reality, panoramic video, user study.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Evaluation/methodology

*andrew.macquarrie.13@ucl.ac.uk, a.steed@ucl.ac.uk

1 INTRODUCTION

Cinematic virtual reality (CVR) is a broad term that could be considered to encompass a growing range of concepts, from passive 360° videos, to interactive narrative videos that allow the viewer to affect the story. Work in lightfield playback [25] and free-viewpoint video [8] means that soon viewers will be able to move around inside captured or pre-rendered media with six degrees of freedom. Real-time rendered, story-led experiences also straddle the boundary between film and virtual reality.

By far the majority of CVR experiences are currently passive 360° videos. Oculus recently stated that 1m people had used the Gear VR head-mounted display (HMD) during the month of April 2016, roughly 80% of whom had watched video on the device [47]. Facebook revealed that 1m hours of video was watched in the Gear VR in the four months after its launch, at which time 20k 360° videos were available on the Facebook platform with “hundreds more added daily” [12]. These videos span a broad spectrum of genres, from news and journalism to comedy and horror. In this paper we look specifically at monoscopic, passive, fixed-viewpoint 360° videos, as these are by far the most commonly available type of video for virtual reality.

There are several issues with HMD-based playback of 360° videos. Arguably the most detrimental problem is the lack of directorial control over what the viewer sees, as the director cannot dictate in which direction the viewer is looking at any given time. This presents issues for narrative understanding, and may lead to the viewer feeling they have missed important elements. In a recent user study, “audiences expressed FOMO (what the kids are calling ‘fear of missing out’)” [24]. Additional issues include the physical comfort of wearing the headsets, and the feelings of vulnerability and unsociability generated by being cut off from the physical world.

The SurroundVideo+ (SV+) display is designed to mitigate some of the issues associated with HMD playback, while attempting to retain some of the immersive characteristics. SV+ is an immersive display, featuring a TV as a focal point to provide a directed viewing experience, as well as projection-based peripheral content to provide immersive visuals. The peripheral projections in our SV+ are provided by a CAVE™-like display [10]. The use of projection

allows SV+ to provide a social viewing experience. The use of projection also means viewers do not need to wear any equipment, and that no setup is required following initial calibration. By using projection mapping and radiometric compensation techniques, SV+-like systems could be one model for the living room of the future. The use of a CAVE™-like display can be seen as an ideal version of SV+ where any furniture or fixings in the room are effectively imperceptible or irrelevant in the context of the video presentation.

Evaluation of passive CVR experiences presents additional issues. Task completion is almost always used in HCI evaluation, with metrics such as speed or accuracy forming the basis of most analyses. As CVR does not involve active tasks, these frameworks are not suitable. Evaluation techniques used in media psychology, however, do not specifically address the issues related to VR, such as immersion, presence, spatial awareness, physical comfort, etc. In this paper, we present literature relevant to the evaluation of such experiences, and explore a subset of identified measures through a user study.

The user study conducted was a between-groups experiment of 63 participants. The study was designed to evaluate the effect of display type on the viewing experience. In doing so, we were also able to critique each measure's ability to differentiate between display systems and to evaluate the CVR experience in general. While many aspects of the CVR experience are interesting, our study focused on spatial awareness, narrative engagement, enjoyment, memory, fear, attention, and a viewer's concern about missing elements of the story.

2 RELATED WORK

2.1 Immersive Displays

There have been many displays that envisage a living room of the future, and aim to extend the content beyond the TV to create a more immersive experience. The automatic extension of existing 16:9 content using vague peripheral shapes and colours was researched by MIT in their Infinity-by-nine system [23]. This concept was commercialised to some degree by Philips with their Ambilight TV [35].

The SV+ system, however, can be seen as an evolution of the Surround Video system designed by the BBC [22], and the Illumi-Room system designed by Microsoft [18]. Both displays extend content beyond the boundaries of the television screen, using projectors to augment the surrounding surfaces. Surround Video featured two video streams captured simultaneously using a standard lens and a 180° fisheye lens. The content captured using the standard lens was displayed on a large HD TV, while the content from the fisheye lens was projected around the TV, extending the content into the viewer's peripheral vision. As the output of a single projector was bounced off of a parabolic mirror to cover a large area of the room, the system suffered from issues related to brightness, colour and distortion. This was compounded by the fact that the system did not perform any kind of projection mapping or radiometric compensation.

The IllumiRoom system, by contrast, used projection mapping to negate or incorporate the room into the “peripheral illusions”, for example making the room look like a cartoon by saturating the colours and projecting black borders onto the surrounding objects [18]. While mostly designed for real-time rendered gameplay, the team also demonstrated the IllumiRoom's potential for CVR, by capturing two streams of video using a setup similar to that of the Surround Video. This produced better visual fidelity than the Surround Video system, due to the projection mapping techniques employed and the limited surface area covered by the single projector employed. Microsoft continued this concept in their RoomAlive system, blending multiple projectors together to allow an entire room to be spatially augmented [17]. The RoomAlive system,

however, did not feature a TV and therefore removed the focal point from the display.

Arguably, the most immersive type of display currently available is the HMD. These displays entirely occlude the real world. Head-tracking is used to allow the viewer to look around inside the space. For real-time rendered graphics, many of these devices allow movement with six degrees of freedom. As most CVR content is currently captured, monoscopic, panoramic video, only the three degrees of freedom associated with orientation are usually available.

2.2 Panoramic Video

By far the majority of CVR is currently monoscopic 360° videos. These videos are usually filmed using several cameras capturing overlapping views that are stitched together in software, allowing a full sphere to be captured which is referred to as the “viewing sphere”. Each view is warped into alignment, corrected for lens distortion and blended together using image processing techniques [42][43][33][6]. The video can then be watched back on a desktop or phone, with Facebook and YouTube both supporting 360° video. For immersive viewing experiences, HMDs can be used.

As discussed in the Introduction, 360° video suffers from some issues. The lack of directorial control is arguably the biggest barrier to adoption that panoramic video faces. While techniques to encourage users to look in a certain direction have been employed, such as the use of lighting [40] or binaural audio [44], there is no definitive way to be sure that a viewer will pick up on these cues. It has been noted that “VR is a major new medium on the scale of photography and film ... In the early days of film people had to learn a new visual grammar – close ups, establishing shots, etc. It will be the same for VR” [45]. This lack of visual conventions makes storytelling difficult, as directors rely on an agreed visual grammar to relay complex ideas.

Due to the current difficulties in storytelling and the expense of production, 360° content is generally of fairly short length, often five minutes or less.

2.3 Evaluation Techniques

Evaluation of CVR experiences presents several issues. Although immersive displays have been evaluated throughout the history of VR to identify benefits over traditional format displays (for example [28]), elements inherent in CVR make adapting existing techniques challenging. Evaluation of virtual environment (VE) displays generally focuses on task completion (for example [4]). This presents a particular problem for CVR evaluation, as these experiences are generally passive and therefore do not have tasks. Likewise, evaluation techniques common for standard format passive displays may lack the ability to assess the impact of highly immersive experiences. Although CVR experiences are considered “passive”, the viewer can look around and is still engaging in multiple cognitive activities. These activities may or may not be supported by the display configuration, such as the field of view (FOV), light levels, etc., and these are definitely of more general interest in VR.

2.3.1 Evaluation of Passive Experiences in Immersive Displays

Philips evaluated an early version of the Ambilight TV [35]. This repeated-measure experiment, in which participants rated criteria such as presence and naturalness on a five-point numeric scale, was informed by the International Telecommunication Union's BT500 methodology [15]. While the BT500 is a common tool for measuring subjective perception of image quality, it is not clear that it can be extended to immersive experiences and concepts such as presence. While significant results were found, the repeated-measure

design may be susceptible to demand characteristics, as participants can easily guess the hypotheses under test.

Further research by Philips for an Advanced Ambilight system included objective measures via physiological monitoring, such as heart rate, skin conductance and respiration [50]. Using physiological monitoring can be challenging, however, and significant results between conditions were not obtained for these measures. Heart rate and skin conductance responses were used successfully by Reeves et al. in their investigation of the effects of screen size on arousal and attention [30]. Their technique for attention requires content with frequent cuts in order to trigger orienting responses, however, and their technique for arousal is best suited for arousing content (e.g. videos containing violence and sex).

A review of literature relevant to immersive display evaluation was completed by Schnall et al. in their investigation of fulldomes, the immersive dome-based projection displays most known for their use in planetariums [34]. With a focus on the educational benefits of fulldome displays, their review included research on immersion, presence, attention, memory and social factors. They presented several suggestions that we have taken into consideration, including ensuring as much consistency as possible between display conditions to reduce the risk of confounding factors. They did not propose a framework for evaluating such displays, however, and did not conduct any experiments.

2.3.2 Presence

Immersive displays have often been measured by the sense of presence they create. In this paper, we follow the convention of using the term “immersion” to mean an inherent characteristic of the display system in use, while “presence” denotes the viewer’s sensation of “being there” [39]. There have been many techniques suggested for the measurement of presence. These include subjective self-assessment questionnaires such as that proposed by Witmer and Singer [51], physiological monitoring [21], and realistic physical responses to situations (“respond-as-if-real”) [37].

While often used, questionnaire-based methods of presence measurement have been criticised [32][36]. Physiological monitoring has also been criticised as a measure for presence, and has shown limited ability to produce significant results outside of stressful experiences [38]. Additionally, it is unclear if a respond-as-if-real measure could be used in a passive display context, as viewers are likely to be aware that any action taken would have no impact. Thus, while it might reasonably be expected that the different displays would elicit different levels of presence response, we decided not to pursue this through questionnaires. As the experience was not active or stressful, other measures would not be suitable. Thus we postpone consideration of this aspect of the displays to future work.

2.3.3 Spatial Awareness

While many measures previously used for assessing VEs cannot easily be adapted to CVR due to their focus on task completion, some measures that use a task that is performed following the experience are viable. One such technique is the measurement of spatial awareness (SA) using a map placement task. In this technique, following a VE experience, participants are asked to mark the locations of objects that were visible in the VE on a map of the environment. The SA metric can be taken as the summed Euclidean distance of these objects from a ground truth. This technique has been used to compare different non-immersive representations of 360° video [2]. A related concept, spatial orientation, was used by Bowman et al. in their comparison of HMD and CAVE™ displays [3]. They found that HMD users were more likely than CAVE™ users to favour natural turning over manual turning using a joystick. Participants were therefore more likely to maintain spatial orientation in an HMD than a CAVE™ display.

2.3.4 Memory

Another objective metric that can be measured after an experience is memory. This can be achieved by asking questions following the stimulus that require the viewer to remember elements of the video. There are several different kinds of memory. Memories can be “incidental” [13], i.e. memories made naturally during an experience, or intentional, where participants consciously try to retain memories during an experience. The popular model of memories proposed by Atkinson and Shiffrin has two memory stores: short-term memory, which lasts in the order of seconds, and long-term memory, which may be held indefinitely [1].

Measuring memory can be achieved in a number of ways, three of which are used in the Wechsler Memory Scale [49]. Free recall is a technique where participants are asked to remember elements without assistance, for example, “List as many character names as you can remember.” A second technique is cued recall, in which a related concept to the subject of the question is provided in order to aid memory retrieval. A third form, recognition, can be assessed using multiple choice questions. Testing memory immediately following an experience is an indicator of immediate recall, which relates to short and long-term memory. Evaluating retention of long-term memories requires a study that spans several days or weeks; participants must be tested after a break to measure the amount of content that has been retained.

The effect of immersion on memory is not well understood. Regan et al. showed that procedure memorisation improved with higher immersion, however their metric focused on physical tasks that relied on the improved spatial cues higher immersion provides [29]. In terms of CVR, while it might be reasonable to hypothesize that a more engaged viewer would remember more, and that a more immersive system would lead to higher engagement, this does not appear to be the case. A memory study in some ways similar to ours was conducted by Rizzo et al. in their investigation of the use of 360° video for memory assessment of persons with cognitive and functional impairments [31]. Their results indicated that participants showed poorer free recall and recognition when a 360° video was watched in a HMD over a standard 16:9 TV. Explanations for this effect focused on mental load, in that participants had needed to expend mental processing to handle the complex visuals, and therefore had less available for storing memories. To ensure fairness between conditions, Rizzo et al. took all memory questions from the audio track, so participants in all display conditions received the same information. Therefore an alternative explanation for the reduced memory performance may be the novelty effect of immersive displays, as participants were distracted by the visual “wow factor” and not attending to the audio track. A similar result was found by Mania et al. in their investigation on the impact of immersion on memory in virtual environments [20]. In their study, a 15-minute seminar was consumed in one of four conditions: audio-only, desktop, HMD, and the real world. They found that immersion level was not positively correlated with recall, however participants did remember more in the real-world condition.

2.3.5 Narrative Engagement

In their Measuring Narrative Engagement Questionnaire (MNEQ), Busselle et al. proposed a set of 12 questions that measure four aspects of engagement: narrative understanding; attentional focus; emotional engagement; and narrative presence [7]. While this technique uses a questionnaire-based approach, physiological monitoring has been used to validate the self-reported narrative engagement scale [41].

These questions were distilled, through a series of experiments, from a much larger set of questions that covered an array of media engagement aspects such as empathy, narrative realism, and transportation. There is overlap between the MNEQ and questionnaires designed for interactive content, for example the Immersive Expe-

rience Questionnaire (IEQ) [16]. The MNEQ, however, is designed for passive experiences, so inappropriate aspects of the IEQ such as flow are not examined.

2.3.6 Simulator Sickness

A known issue of immersive displays is the risk of simulator sickness. While the exact mechanism is not fully understood, it is believed that a mismatch between optical flow detected by the eyes and balance as detected by the inner ear can cause nausea (for a review of the literature, see [14]). In certain cases, the effects of simulator sickness can be severe. There is evidence that peripheral vision plays a central role in detectingvection, and therefore a display with a very wide field-of-view may produce a stronger effect [5]. This is of particular concern for CAVE™-like environments, which can match the horizontal field-of-view of human vision. Therefore measuring simulator sickness is very important. The Simulator Sickness Questionnaire (SSQ) is the gold standard for measuring these effects, and is used extensively in the field [19].

2.4 CVR and the Horror Genre

It has been suggested that “VR horror experiences can be much more intense, isolating and terrifying than when played on a standard 2D display” [48]. While this statement was directed at VR gaming, CVR is likely to be similar. The complete occlusion of the real world in HMD experiences gives media producers the opportunity to control many elements of the experience, allowing them to create highly atmospheric content. Additionally, the isolation and physical vulnerability of viewers would likely increase any sense of fear. While this will likely produce more powerful horrors, it has been suggested that “jump scares” in VR may be too intense, prompting Oculus VR to “strongly discourage” content creators from using them [48].

3 METHOD

A between-groups experiment of 63 participants was conducted, in which participants watched 360° videos in one of three display conditions. The details of this study are described below. The study was approved by the UCL Research Ethics Committee.

3.1 Subjects

Participants were recruited through a university mailing list and two participant pool websites. A total of 65 participants took part in the study, but data from two was excluded due to procedural issues. Of the 63 remaining participants, 27 were male and 36 were female. Ages ranged from 19 to 76 years (mean: 27.78; standard deviation: 9.27). Participants were randomly assigned to a display condition until a condition reached 21 participants, at which point participants were randomly assigned between the remaining conditions. As a result, each display condition had 21 participants.

3.2 Experimental Conditions

Each participant watched all videos in one of three display conditions, which are described below.

3.2.1 Head-Mounted Display

The HMD used in this condition was the Oculus Rift CV1, driven by a Windows 8.1 desktop PC with an Intel i7-4790 CPU running at 3.6GHz with 8GB of RAM. The video card in use was a NVIDIA GeForce GTX 1080. The Oculus CV1 has a refresh rate of 90Hz. Whirligig was used as the video playback software, in which orientation tracking allowed viewers to look around the viewing sphere naturally. The videos played with no visible lag at their expected frame rate of 30fps. Attempts were made to keep as many aspects as possible consistent across all three display conditions. To that end, in all display conditions the viewer was seated in the CAVE™-like display even when the displays were turned off, as was the case in

the HMD condition. Audio across all three conditions was provided by the same stereo speakers, mounted above the corners of the walls of the CAVE™-like display. The built-in headphones of the Oculus Rift CV1 were removed, and the audio volume was the same across all conditions.

3.2.2 SurroundVideo+

The SV+ display is intended as a middle ground between the highly immersive visuals of a HMD – in that SV+ entirely fills the horizontal FOV of the viewer – and the directed experience of a TV. In Microsoft’s IllumiRoom display, peripherally projected content was designed to spatially augment a living room. In our SV+ display, we imagine a future scenario in which projection mapping can be used to visually negate the appearance of the surrounding room and replace it with the desired peripheral content.

The SV+ system used a three-walled CAVE™-like system to display the peripheral content, with a TV placed centrally to provide a focal point, as shown in the centre image in Figure 1. The TV was placed on a 70cm high table. Each wall of the CAVE™-like display was 3m long and 2.2m high, with each wall having a resolution of 1400x1050. While the CAVE™-like display in use also had a floor projector available, this was not utilised. The viewer sat on a fixed chair, positioned slightly back from the centre of the CAVE™-like display, meaning the peripheral projections covered the entire horizontal visual field of a viewer looking forward. No head tracking was used, as the videos do not support parallax, and head turning does not have an effect on the display.

The TV in use was a 60” LED HD TV made by LG, model number 60LA620V. The TV and projectors were driven by a Windows 7 desktop PC with an Intel i7-3930K CPU running at 3.2GHz with 32GB of RAM. The video card in use was a NVIDIA Quadro K5000. With a viewing distance of just over 2m, the TV subtended a solid angle of approximately 36° horizontally and 21° vertically for the head position of an average viewer. This meant the diagonal FOV of the TV was around 41°.

In order to allow 360° videos to be repurposed for use in the SV+ display, in a pre-processing step, tracking data was created to ensure the content deemed important at any given time in the video was displayed centrally on the TV. This data was produced by manually annotating the videos with keyframes. Each keyframe specified a “forward” direction that indicated what the authors felt was the most important content in the viewing sphere at that moment. An example of this would be that, in general, a character was considered to be the most important scene element when they were speaking. These keyframes were then linearly interpolated to produce a “forward” direction per frame. This meant that, when played, the viewing sphere rotated, tilted and cut around the viewer to maintain the important content on the TV as it moved through the scene. This tracking data produced video playback with a clear narrative, however the authors are non-expert directors so the visual experience may not be optimal.

Video playback in the SV+ system was achieved using an adapted version of the open source OmiPlayer 360° video player written by Omar Mohamed Ali [26]. While the projectors and TV had high refresh rates of 96Hz and 200Hz respectively, the videos were only available with a frame rate of 30fps. The system was capable of displaying the videos at this frame rate with no visible lag.

3.2.3 Television

The TV display condition was identical to the SV+ display condition, except that the peripheral projections were disabled. For ease of swapping conditions, the projectors were not switched off but instead they projected solid black. For this reason, the walls of the CAVE™-like display during the TV condition were at the projectors’ black level.

3.3 Stimuli

Videos that matched the genre requirements were selected based on several factors. The videos needed to be suitably engaging, and high in audio and video quality. To ensure a contiguous image between the projected content and the TV, the FOV of the content displayed on the TV was fixed at the angle subtended by the physical TV within the CAVE™-like display. As discussed in section 3.2.2, the diagonal FOV of the content displayed on the TV was 41°. This meant the videos needed to match certain character placement requirements, and videos in which characters were too close to the camera had to be discarded.

Due to the small solid angle subtended by the TV, videos with as high a resolution as possible were required. Source videos were only available with a maximum resolution of 4K. Even with a 4096x2048 pixel equirectangular video, the effective resolution of the content displayed on the TV was only 409x250 pixels. It is fair to say that this is a noticeably low resolution for a TV capable of displaying up to 1080p. As the participant was seated just over 2m from the display, this did not seriously distract them.

Videos were also selected based on their appropriateness for the designed measures, which are discussed in full in section 3.5. For example, the spatial awareness task required participants to mark the locations of objects that had been seen in the video on a floor plan of the scene. This required videos that were staged largely in a single space, and had objects that could reasonably be remembered. The MNEQ required narrative content, which proved particularly difficult to source. This was due to the limited duration of content, as well as the fact that difficulties in storytelling mean many 360° videos are experiential in nature rather than narrative based.

It was decided that participants would sit in a fixed chair across all conditions. During an informal pilot study, participants in the HMD condition failed to observe any action happening behind them, including a dramatic fight sequence. As a result, content was reselected in which most action happened in the forward direction, and all critical action happened within ±100° from “forward”. This may have had unintended consequences for some of our measures, as will be discussed later in section 4.

Four videos were selected. The first, a 2m19s music video, was used to reduce the novelty effect of the display. The DOCUMENTARY stimulus was a 5m53s documentary piece about a beekeeper. The HORROR stimulus, which was of the slasher sub-genre, was 2m47s long. The final video, the NARRATIVE stimulus, told the story of a murder and lasted 4m43s.

3.4 Hypotheses

The following hypotheses were investigated:

H1 There will be a difference in spatial awareness between conditions

H2 There will be a difference in incidental memory between conditions

H3 There will be a difference in narrative engagement between conditions

H4 There will not be a difference in video enjoyment between conditions

H5 There will be a difference in display enjoyment between conditions

H6 Attention can be guided in TV/SV+ conditions

H7 There will be a difference in participant’s concern about missing something between conditions

H8 Participants will be more afraid during a horror video in the HMD condition

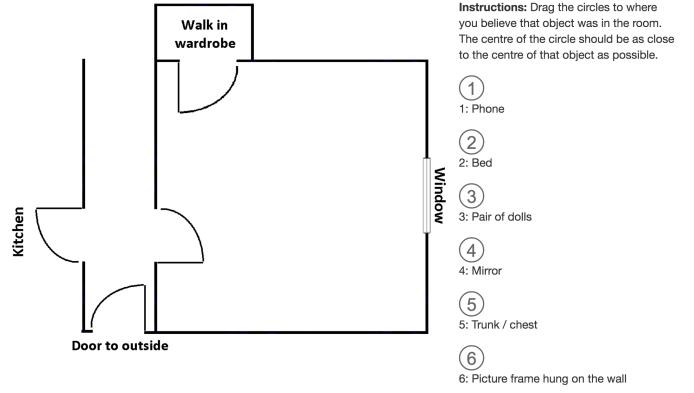


Figure 2: Object placement task for HORROR stimulus. Instructions read: “Drag the circles to where you believe that object was in the room. The centre of the circle should be as close to the centre of that object as possible.”

3.5 Measures

There is a vast range of questions that can be asked about immersive media. It was essential that we selected a subset of these measures so as not to overwhelm participants, and to keep each trial to under an hour to prevent fatigue. As CVR is a new field of study, our focus for the selection of metrics has been based on conversations with the CVR community – such as concerns that have been raised about the format by producers – as well as areas we believe are promising and applicable based on previous VR research. While there is little academic basis so far, open-ended, qualitative studies are emerging that indicate that these concerns are not just being felt by content producers, but are indeed critical aspects of the end-user experience [27]. Part of the selection process was determining what effective measures can be used, and what measures might be generally useful. To aid the design of future studies, the complete list of questions and tasks are presented in the supplementary materials.

H1: Spatial awareness To measure spatial awareness, participants were asked to mark the locations of objects seen in the video on a floor plan of the room featured in the video. As two of the stimuli were set mostly within a single location per video (HORROR and NARRATIVE), the spatial awareness test was administered for these videos. The map placement task for the HORROR stimulus can be seen in Figure 2. Objects were represented as circles to ensure participants did not focus on orientation, and were labelled to ensure each placed object had a known corresponding object in the ground truth.

Objects were chosen at varying levels of difficulty, from items that characters interacted with that formed an element of the plot, to more difficult items that would be easy to miss. For reasons of fairness, all objects were visible in all display conditions. At least one participant in each condition correctly placed all objects.

The final measure was taken as the summed Euclidean distance of each object from a “ground truth” object placement. If a participant failed to place an object, the object’s original unplaced position was used, thus incurring a consistent penalty between participants for each unplaced object. As the videos in question were not filmed by us, the “ground truth” and floor plans were to some extent approximations based on close inspection of the videos.

H2: Incidental memory To measure incidental memory, participants were asked ten questions about content from the DOCUMENTARY stimulus. All questions could be answered from the audio track, and no additional information could be gained from the visuals. While the video features a visible narrator, his face is

masked by a beekeeper's veil and therefore no additional information is gained by looking at him. The audio was delivered through the same speakers and at the same volume across all three display conditions.

The questions varied in difficulty. The easiest was a fact that was said twice in different wordings, "On this frame there's a couple of microphones. I like to record stereo in the hive". Memory of this statement was checked with the question, "How many microphones were on the honeycomb?". The hardest questions related to difficult-to-remember concepts that were mentioned in passing, for example the statement "I'm not allergic to stings, which one in every thousand people are" for the question "What ratio of people are allergic to bee stings?".

H3: Narrative engagement Narrative engagement was measured using the Measuring Narrative Engagement Questionnaire (MNEQ) following the two narrative stimuli (HORROR, NARRATIVE). The MNEQ was not applied following the DOCUMENTARY stimulus, as this piece is non-narrative and many of the MNEQ questions would not make sense in this context.

H4, H5: Enjoyment Enjoyment was measured using five-point Likert scale indications of agreement with two statements. We wished to measure enjoyment of the display and enjoyment of the video separately. In an attempt to tease apart these enjoyment levels, participants were asked to indicate their level of agreement with two statements: "Considering the display and the video separately, I enjoyed watching this video" and "Considering the display and the video separately, I enjoyed using this display". These two questions were placed side by side on the questionnaire to ensure participants answered them in tandem.

H6: Attention guided To test how successfully attention could be directed in the TV/SV+ conditions over the theoretically undirected HMD condition, elements of the video that could be examined via questionnaire were highlighted in the TV/SV+ conditions by centring them on the TV. For example, this included a lingering shot of the murder weapon prior to the murder in the NARRATIVE stimulus, and cuts to a telephone call being initiated by a victim in the HORROR stimulus. These were examined using the questions "What was the murder weapon? Describe it as specifically as you can (colour, shape, material)", and "Who initiated the phone call?", respectively. In total four such questions were asked. Three questions had pass/fail answers that contributed zero or one to the total attention score, and one was marked out of three depending on the level of detail provided of the murder weapon. This gives a total possible score range of zero to six. To ensure fairness, the answers were marked blind, i.e., the display condition of the participant was not known when their attention score was tallied.

H7: Concern about missing something Concern about missing something was measured by five-point Likert scale responses of agreement to the statements, "At times, I was worried I was missing something", and, "My concern about missing something impacted my enjoyment of the video". These questions were not placed side by side on the questionnaire. The responses to these two questions were summed together to give an indication of a participant's general concern about missing something.

H8: Fear during horror To improve the validity of questionnaire responses, participant's fear during the HORROR stimulus was measured using two questions. The question, "I felt afraid while watching this video", was designed to measure fear directly. The question, "I felt nervous while watching this video", was designed to measure anxiety, a state associated with fear during horror media [46]. Participants responded to both questions on a five-point Likert scale of agreement. These questions were not placed side by side on the questionnaire. These answers were summed to give a total score.

3.6 Procedure

Due to the learning effects that would exist in the metrics for memory and spatial awareness, a between-groups design was required. Each participant, therefore, watched all videos in one of the three display conditions. As many elements as possible outside of the display device were held constant, including the video order, audio and chair used.

Before arrival, participants were assigned to a display condition using a random number generator. Upon arrival, participants were given an information sheet to read. Important aspects of the information sheet were reinforced verbally, such as the procedure and the risk of simulator sickness. Participants then signed a consent form, and were introduced to the experiment environment.

The experiment consisted of four videos. The first – a 2m19s music video – was designed to reduce the novelty effect of the display and was not followed by any questions. Immediately following each of the other three videos, participants were given a laptop to complete the appropriate questions and tasks for that stimulus. Before beginning the experiment, all participants were given identical advice about the nature of the questions, specifically that they would relate "to the content and their experience".

As the DOCUMENTARY stimulus was the video used to measure incidental memory, it was viewed second. This was to reduce the likelihood that participants' viewing of the video would be influenced by the questions for other metrics. For example, by knowing that questions regarding spatial awareness would be asked, it may have caused a participant to focus on remembering objects in the video, rather than viewing the video naturally. At the end of the DOCUMENTARY stimulus, participants answered questions to test incidental memory, as well as generic questions about the experience such as their concern about missing something and enjoyment.

The HORROR stimulus was watched third. Following this video, participants answered questions related to their feelings of fear during the video, and questions that tested their memory of visual aspects that the TV and SV+ conditions deliberately drew attention to. Participants also completed the MNEQ. Generic questions such as enjoyment and concern about missing something were also answered. Finally, participants were asked to place objects from the video on a floor plan of the room to test spatial awareness.

The forth and final video was the NARRATIVE stimulus. As with the HORROR stimulus, participants answered questions on attention, narrative engagement, enjoyment and concern about missing something. Participants also completed a spatial awareness task. Participants then completed a SSQ. Participants then took part in a short, unstructured interview to gain insights into their reasoning and opinions.

After the experiment, participants were debriefed about the purpose of the study and given the opportunity to use the HMD if they has not had the chance during the experiment. Finally, participants were given £10 for taking part.

4 RESULTS AND DISCUSSION

Due to the ordinal nature of the Likert, attention and memory data – and outliers in the Euclidean error distance data – analysis was conducted using the Kruskal-Wallis H Test for all hypotheses. Scores were not similar for all groups, as assessed by visual inspection of boxplots of the data (for boxplots, see supplementary material). The results of the Kruskal-Wallis H Tests can be seen in Table 1. As eight hypotheses were being tested, Bonferroni correction was applied where the statistical significance required was $p < .00625$.

H1: Spatial awareness There was a statistically significant difference between display conditions in our measure of spatial awareness, as shown in Table 1. Pairwise comparisons were performed using Dunn's procedure with a Bonferroni correction for

Table 1: Results of Kruskal-Wallis H tests across all hypotheses.

Dependant variable	Stimulus	Mean rank			<i>df</i>	χ^2	Asymp. Sig.	η_H^2 [‡]
		HMD	SV+	TV				
Spatial awareness [↓]	Ensemble	19.90	36.19	39.90	2	14.146	.001*†	.202
	HORROR	23.29	35.05	37.67	2	7.334	.026	.089
	NARRATIVE	21.81	35.29	38.90	2	10.145	.006	.136
Incidental memory	DOCUMENTARY	31.33	30.02	34.64	2	0.723	.697	-0.021
Narrative engagement	Ensemble	35.17	32.50	28.33	2	1.485	0.476	-0.009
	HORROR	34.93	31.62	29.45	2	.953	.621	-0.017
	NARRATIVE	35.74	34.12	26.14	2	3.306	.191	.022
Enjoyed video	Ensemble	37.69	29.36	28.95	2	3.152	0.207	.019
	DOCUMENTARY	37.10	26.10	32.81	2	4.556	.102	.043
	HORROR	32.81	35.50	27.69	2	2.363	.307	.006
	NARRATIVE	34.64	33.38	27.98	2	1.960	.375	-0.001
Enjoyed display	Ensemble	42.83	32.00	21.17	2	15.196	0.001*†	.220
	DOCUMENTARY	40.67	31.81	23.52	2	10.383	.006	.140
	HORROR	40.69	35.21	20.10	2	16.378	.0003	.240
	NARRATIVE	41.38	30.29	24.33	2	10.584	.005	.143
Attention	Ensemble	31.62	28.38	36.00	2	1.969	.374	-0.001
	HORROR	32.00	32.00	32.00	2	.000	1.000	-0.033
	NARRATIVE	31.31	27.74	36.95	2	2.884	.236	.015
Concern about missing something [↓]	Ensemble	27.10	37.00	31.90	2	3.103	.212	.018
	DOCUMENTARY	27.64	38.21	30.14	2	3.942	.139	.032
	HORROR	32.40	33.02	30.57	2	.210	.900	-0.030
	NARRATIVE	25.55	35.48	34.98	2	4.025	.134	.034
Felt afraid	HORROR	28.14	31.55	36.31	2	2.155	.341	.003

^{*}p < .05[†]p < .00625 (Bonferroni corrected for 8 hypotheses)[↓]A smaller value indicates a better result[‡]Effect size calculated from Kruskal-Wallis H value [9]

multiple comparisons [11]. Adjusted p-values are presented. Values are mean ranks unless otherwise stated. This post hoc analysis revealed statistically significant differences in ensemble Euclidean error scores between the HMD (19.90) and SV+ (36.19) ($p = .012$), and HMD and TV (39.90) ($p = .001$) display conditions, but not between the TV and SV+ display conditions ($p = 1.0$).

These results indicate that the HMD produced better spatial awareness than SV+ and TV displays. SV+ does not offer a statistically significant improvement to spatial awareness over the TV. This is somewhat unexpected, as the SV+ provides a horizontal FOV beyond that of human vision, while the TV offers only a 36° horizontal FOV. We propose two possible explanations for this. The first is that the key to producing good spatial awareness is exploration of the space, and the SV+ discourages exploration by providing a focal point. An alternative explanation is that rotation of the virtual space played a role. The HMD display was the only condition in which the virtual world is fixed with regards to the participant. In the SV+ and TV conditions, the world rotates and tilts to ensure the important content remains centred on the TV. It is possible that this rotation disoriented viewers, resulting in poorer spatial awareness. Further investigation is required to clarify these results.

H2: Incidental memory No significant difference in incidental memory was found between conditions, as shown in Table 1. We were unable to recreate the results of [31], which indicated that incidental memory may be lower in a HMD over a tracked TV experience. Memory is highly variable between individuals, however, and therefore more than 63 participants may be required to produce a statistically significant result, if such an effect exists.

H3: Narrative engagement While the order of mean rank scores for narrative engagement increased from HMD, to SV+ and

then TV, the results were not statistically significant, as shown in Table 1. Additionally, none of the four sub-scales of the MNEQ showed a statistically significant difference (see supplementary material for sub-scale results). While this may indicate that narrative engagement is not strongly affected by display condition, it may also be a result of the short-form content that was used. While the short films used in the experiment were generally well received by participants, narrative engagement as measured by the MNEQ may require a more substantial and engrossing plot than was offered in these clips. For example, questions such as, “During the program, when a main character succeeded, I felt happy, and when they suffered in some way, I felt sad”, may require a stronger connection with the characters than was achieved, as well as a more substantial narrative arc.

H4, H5: Enjoyment A statistically significant difference between display conditions was present for display enjoyment, but not for video enjoyment, as shown in Table 1. This indicates that participants were able to separate these concepts.

Pairwise comparisons of display enjoyment results were performed using Dunn’s procedure with a Bonferroni correction for multiple comparisons [11]. Adjusted p-values are presented. Values are mean ranks unless otherwise stated. This post hoc analysis revealed statistically significant differences in ensemble display enjoyment scores between the HMD (42.83) and TV (21.167) ($p = .000$), but not between HMD and SV+ (32.00) ($p = .154$) display conditions, or between SV+ and TV display conditions ($p = .154$).

While the ensemble pairwise comparisons did not produce a statistically significant difference between the TV and SV+ conditions, pairwise comparisons do produce a statistically significant difference between TV and SV+ for the HORROR stimulus ($p = .012$). While this must be considered a post hoc analysis, the differ-

ence in the level of display enjoyment between stimuli may indicate that certain types of content are more appropriate for a given display. Further investigation would be required to determine what may have made the HORROR stimulus particularly well suited for the SV+ display. Possible aspects that may have contributed include the genre, character placement, the rapidity of cuts required, and the characteristics of the captured environment.

The resolution of the TV may have had an impact on this metric. During the unstructured interview at the end of the experiment, several participants mentioned that the low effective resolution of the TV had impacted their enjoyment of the display. While the central content had an identical resolution in the TV and SV+ conditions, it is possible that the highly immersive visuals of the CAVE™-like display partially compensated for this issue in the SV+ condition, resulting in a stronger enjoyment result.

H6: Attention guided No significant difference in attention was found between conditions, as shown in Table 1. This was an unexpected result, as it seems almost certain that a TV can draw attention to a specific aspect of the scene more effectively than a HMD. We propose two possible explanations for this. The first is that the chosen metrics for attention were not suitable, in which case a repeat of this study using different videos and attention measures may produce a statistically significant result. Another explanation is that the fixed chair used across all conditions limited the field of regard (FOR) of the HMD such that the risk of a viewer missing some aspect was substantially reduced.

As mentioned earlier, during a pilot study a video featuring a fight scene was replaced as important material was missed in the HMD condition. This decision was made to encourage fairness between the conditions. In the removed video, a character starts in front of the viewer, at roughly 0° from “forward”. This character then walks through the scene and becomes engaged in a fight directly behind the viewer, at roughly 180° from “forward”. No viewers were willing to crane in their seat to follow this character, despite his prominence, and became confused by the audio of a fight they could not locate in the scene. This may indicate that the FOR of a HMD when watched in the fixed chair is more limited than the available FOV. This may have interesting practical implications for current 360° content creation, as home viewers will most likely be seated on a fixed chair or couch.

H7: Concern about missing something In none of the videos did the HMD receive the highest mean rank score for concern about missing something, although the difference between display conditions was not significant as shown in Table 1. This result was unexpected, as there is a prevailing belief that HMD experiences often leave viewers feeling they have missed something. As discussed in the Attention results section, this may be because of the partially restricted FOR caused by the non-swivel chair, as well as the intentional choice of videos in which all action happens within ±100° of forward. Due to this soft limit on the FOR, the HMD condition might also be considered partially guided.

The SV+ display received the highest mean rank score across all stimuli for this measure. This result is also unexpected, as the SV+ was intended as a partially guided experience. This may have been due to the feeling that viewers should be looking at the TV, but were aware that additional content was available in the periphery that they were unable to digest. These findings may also be a result of viewers’ subtle awareness that the content was designed for a 360° viewing experience, and that only a portion of the FOV was available in the TV and SV+ conditions. An alternative explanation is that rotations in the SV+ and TV conditions contributed to the feeling of missing something, for example if some previously available content rotated out of view.

H8: Fear during horror Displays with higher immersion produced lower levels of fear during the horror video, although not sig-

nificantly as shown in Table 1. These unexpected scores may have been a result of the deliberate decision – due to concern about upsetting participants – to use a horror video that was extremely mild. One participant reported finding the video “lame”, while none answered “Strongly agree” to the question “During this video, I felt afraid”. The type of horror – the slasher sub-genre – perhaps does not make best use of the characteristics of the HMD, such as being isolated, feeling physically vulnerable, and being entirely surrounded by the content. More psychological horrors, or horrors that rely on jump scares, might prove more effective in producing stronger differences between the display conditions. Worthy of note is that a participant in the HMD condition – who appeared visibly shaken by the video – reported in the post-experiment interview that the audio that had played a key role in eliciting their fear.

Simulator sickness A Kruskal-Wallis H test was conducted to determine if there were differences in SSQ scores between the three display conditions: HMD ($n = 21$), SV+ ($n = 21$), and TV ($n = 21$). Values are mean ranks unless otherwise stated. Distributions of SSQ scores were not similar for all groups, as assessed by visual inspection of a boxplot (see supplementary material for boxplot). SSQ scores increased from TV (28.14), to HMD (31.50), to SV+ (36.36) display conditions, but the differences were not statistically significant, $\chi^2(2) = 2.192$, $p = .334$.

5 LIMITATIONS

Some issues were encountered when repurposing 360° videos that were designed for HMD viewing for display in the TV and SV+ conditions. Finding content that worked in all three display conditions, and also met our genre and quality requirements, meant some compromises were required. This meant that some videos were not ideal for viewing in the SV+ condition. For example, the NARRATIVE stimulus featured conversations between two characters who were not standing near each other. To keep the speaking character on the TV required several cuts in quick succession. This was not a problem in the TV condition, as this is a standard visual technique used in TV. Some participants in the SV+ condition – who could see the other character in their peripheral vision – found this cutting irritating, prompting one to comment the display seemed to “flit about”. If the content had been filmed specifically for the purpose of this test or these displays, care could have been taken in character placement to ensure both characters could be framed centrally on the TV during conversations.

Our measure of spatial awareness may have been susceptible to a confounding factor, in that rotation of the virtual world with respect to the viewer was only present in two of the three display conditions (SV+ and TV). To clarify that the significant benefit in spatial awareness offered by the HMD was not caused by this rotation, an experiment should be conducted that investigates the relationship between exploration of the space and spatial awareness, with rotation controlled as an independent variable.

Rotations of the scene were generally well received. Some participants in the SV+ condition, however, did comment that the rotations had felt somewhat intense. This is to be expected, as the CAVE™-like display entirely filled the horizontal FOV of participants. No participants reported feeling unwell, and there was no significant difference in SSQ scores between display conditions. Our videos were all of short duration, however, so the risk of cumulative effects caused by the increased FOV for longer experiences has not been addressed.

During the post-experiment interview, a number of participants also commented on the low resolution in the TV condition. This is also to be expected, as the effective resolution of the TV was only 409x250 due to the solid angle subtended by the TV inside the 4k viewing sphere. Interestingly, participants in the SV+ condition tended not to comment on the resolution of the TV, despite it being identical to the TV condition. This may indicate that the immersive

visuals provided by the CAVE™-like display compensated for the poor effective resolution of the TV to some degree. This low resolution was again caused by the decision to repurpose 360° video designed for HMD experiences. If content was to be shot specifically for the SV+ display, it is likely a similar approach to the original Surround Video system would be used: capturing a high-resolution inset using a dedicated camera, while using separate cameras to capture the lower resolution peripheral content. This again emphasises that immersive video may not be a single type of experience: producers need to be aware of the type of display viewers will use. There may not be a single form of immersive video and guidelines for producers will need to recognise this.

Several participants commented on the audio mix in the DOCUMENTARY stimulus. Background noise and music present in the audio track may have hindered a participant's ability to hear the dialogue. This may have impacted their ability to remember the content, and therefore affected the memory metric. This was not a confounding factor, however, as all participants received identical audio across conditions.

Other improvements to the testing method include using longer format content when measuring narrative engagement, using a more appropriate horror stimulus to test fear, and the use of specifically filmed content with resolution and character placements suitable for all display conditions.

6 CONCLUSION

In this paper, several metrics for measuring passive CVR experiences were discussed. These metrics related to areas in which CVR was likely to present an improvement over traditional film and TV experiences, as well as areas of concern for CVR that may impact its adoption. While there are many factors that are important to CVR experiences, such as display ergonomics, sociability, etc., it was important to limit the scope of our study to prevent participant fatigue. After consideration of the applicability and appropriateness of the metrics under review, the aspects investigated measured spatial awareness, narrative engagement, enjoyment, memory, fear, guiding attention, and the feeling of missing something.

A between-groups experiment was conducted with 63 participants. Three display conditions were investigated, including the radically different immersive displays of the HMD and the SV+, and a non-immersive TV condition. Our results indicated the HMD offered a significant benefit in terms of spatial awareness over both the TV and SV+ conditions. While it was expected that the TV would perform poorly on the spatial awareness metric, the SV+ was expected to perform better due to the large FOV provided. While this may have been caused by a lack of exploration encourage by the focal point of the TV, more experimentation is required to rule out the possibility of confounding factors. Significant improvements for enjoyment were present in the HMD over the TV and SV+, as well as the SV+ over the TV for the HORROR stimulus.

We were unable to confirm the work of a previous study that showed incidental memory may be lower in a HMD over a TV. We did not find a significant difference in narrative engagement between conditions, however this may be a result of the short-form stimuli that were available. The lack of a significant difference between displays in our measurement of fear during the HORROR stimulus may have been caused by the choice of an inappropriate video. The stimulus was not scary, and did not make use of the characteristics of the immersive displays likely to increase fear.

Drawing attention and a viewer's concern about missing something were also not significantly different between display conditions. These are unexpected results, as there is a commonly held belief that the undirected experience of the HMD causes viewers to miss visual events, as well as to experience a feeling of missing something. Our results may indicate that using a fixed chair in a HMD experience places a soft limit on the field of regard. This re-

sult would have important consequences for video production, as at-home viewers will likely be seated on a couch. Further experimentation is required to confirm this result, and to identify the FOR available for use in videos designed for consumption in the home.

It is clear that passive media viewing consists of a complex interplay of factors that present many challenges for evaluation. As CVR is becoming a common use for virtual reality hardware, it is essential for the virtual reality community to investigate this emerging field. Overall, one important steer for future research is the difficulty of finding content for each display. The affordances of the displays are actually quite different, and thus content developed for one may not work on another. We have highlighted that comparing SV+ to HMD is difficult. Even evaluating different types of HMD content may be challenging, because of assumptions made on whether the participant can turn freely or not. In future work, we hope to explore further this complex area and cover a wider range of scenarios, such as videos with camera motion and social viewing of videos.

ACKNOWLEDGEMENTS

This work was supported in part by grant EP/G037159/1 for University College London's Virtual Environments, Imaging & Visualisation Doctoral Training Centre (<http://engdveiv.ucl.ac.uk/>) from the UK Engineering and Physical Sciences Research Council (EPSRC) and the BBC. The authors would like to thank BBC R&D for their support and allowing us to use their content.

REFERENCES

- [1] R. C. Atkinson and R. M. Shiffrin. Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, 2:89–195, 1968.
- [2] W. Boonsuk, S. Gilbert, and J. Kelly. The impact of three interfaces for 360-degree video on spatial cognition. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2579–2588. ACM, 2012.
- [3] D. A. Bowman, A. Datey, Y. S. Ryu, U. Farooq, and O. Vasnaik. Empirical comparison of human behavior and performance with different display devices for virtual environments. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 46, pages 2134–2138. SAGE Publications, 2002.
- [4] D. A. Bowman, J. L. Gabbard, and D. Hix. A survey of usability evaluation in virtual environments: classification and comparison of methods. *Presence*, 11(4):404–424, 2002.
- [5] T. Brandt, J. Dichgans, and E. Koenig. Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16(5):476–491, 1973.
- [6] M. Brown and D. G. Lowe. Automatic panoramic image stitching using invariant features. *International journal of computer vision*, 74(1):59–73, 2007.
- [7] R. Busselle and H. Bilandzic. Measuring narrative engagement. *Media Psychology*, 12(4):321–347, 2009.
- [8] J. Carranza, C. Theobalt, M. A. Magnor, and H.-P. Seidel. Free-viewpoint video of human actors. In *ACM transactions on graphics (TOG)*, volume 22, pages 569–577. ACM, 2003.
- [9] B. H. Cohen. *Explaining psychological statistics*. John Wiley & Sons, 2008.
- [10] C. Cruz-Neira, D. J. Sandin, T. A. DeFanti, R. V. Kenyon, and J. C. Hart. The cave: audio visual experience automatic virtual environment. *Communications of the ACM*, 35(6):64–73, 1992.
- [11] O. J. Dunn. Multiple comparisons using rank sums. *Technometrics*, 6(3):241–252, 1964.
- [12] Facebook. New steps toward the future of virtual reality. <http://newsroom.fb.com/news/2016/02/new-steps-toward-the-future-of-virtual-reality/> [Online; accessed 26-August-2016].
- [13] E. L. Glisky. Incidental memory. In *Encyclopedia of Clinical Neuropsychology*, pages 1303–1304. Springer, 2011.
- [14] K. S. Hale, K. M. Stanney, B. Keshavarz, H. Hecht, and B. D. Lawson. Visually induced motion sickness: causes, characteristics, and

- countermeasures. In *Handbook of Virtual Environments: Design, Implementation, and Applications, Second Edition*, pages 647–698. CRC Press, 2014.
- [15] ITU-R. *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union, 2003.
 - [16] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, and A. Walton. Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661, 2008.
 - [17] B. Jones, R. Hodges, M. Murdock, R. Mehra, H. Benko, A. Wilson, E. Ofek, B. MacIntyre, N. Raghuvanshi, and L. Shapira. Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 637–644. ACM, 2014.
 - [18] B. R. Jones, H. Benko, E. Ofek, and A. D. Wilson. Illumiroom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 869–878. ACM, 2013.
 - [19] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
 - [20] K. Mania and A. Chalmers. The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior*, 4(2):247–264, 2001.
 - [21] M. Meehan, B. Insko, M. Whittton, and F. P. Brooks Jr. Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics (TOG)*, 21(3):645–652, 2002.
 - [22] P. Mills, A. Sheikh, G. Thomas, and P. Debenham. Surround video. In *Proceedings of the 2011 Networked and Electronic Media Summit*, pages 55–63, September 2011.
 - [23] MIT Object Based Media Group. Infinity-by-nine. <http://obm.media.mit.edu/> [Online; accessed 22-August-2016].
 - [24] K. Newton and K. Soukup. The storyteller’s guide to the virtual reality audience. <https://medium.com/stanford-d-school/the-storytellers-guide-to-the-virtual-reality-audience-19e92da57497> [Online; accessed 14-September-2016].
 - [25] Nozon. What is presenz? <http://www.nozon.com/presenz> [Online; accessed 26-August-2016].
 - [26] Omigamedev. Omiplayer: media player for oculus rift. <https://bitbucket.org/omigamedev/omiplayer> [Online; accessed 3-September-2016].
 - [27] P. J. Passmore, M. Glancy, A. Philpot, A. Roscoe, A. Wood, and B. Fields. Effects of Viewing Condition on User Experience of Panoramic Video. In *ICAT-EGVE 2016 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments*. The Eurographics Association, 2016.
 - [28] R. Pausch, D. Proffitt, and G. Williams. Quantifying immersion in virtual reality. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 13–18. ACM Press/Addison-Wesley Publishing Co., 1997.
 - [29] E. D. Ragan, A. Sowndararajan, R. Kopper, and D. A. Bowman. The effects of higher levels of immersion on procedure memorization performance and implications for educational virtual environments. *Presence: Teleoperators and Virtual Environments*, 19(6):527–543, 2010.
 - [30] B. Reeves, A. Lang, E. Y. Kim, and D. Tatar. The effects of screen size and message content on attention and arousal. *Media Psychology*, 1(1):49–67, 1999.
 - [31] A. Rizzo, L. Pryor, R. Matheis, M. Schultheis, K. Ghahremani, and A. Sey. Memory assessment using graphics-based and panoramic video virtual environments. In *Proc. 5th Intl Conf. Disability, Virtual Reality & Assoc. Tech*, 2004.
 - [32] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience*, 6(4):332–339, 2005.
 - [33] H. S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(3):235–243, 1999.
 - [34] S. Schnall, C. Hedge, and R. Weaver. The immersive virtual environment of the digital fulldome: Considerations of relevant psychological processes. *International Journal of Human-Computer Studies*, 70(8):561–575, 2012.
 - [35] P. Seuntiens, I. Vogels, and A. Van Keersop. Visual experience of 3D-TV with pixelated ambilight. *Proceedings of PRESENCE*, 2007, 2007.
 - [36] M. Slater. How colorful was your day? why questionnaires cannot assess presence in virtual environments. *Presence*, 13(4):484–493, 2004.
 - [37] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
 - [38] M. Slater, A. Brogni, and A. Steed. Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th Annual International Workshop on Presence*, volume 157. Citeseer, 2003.
 - [39] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997.
 - [40] SpeedVR. Lighting for virtual reality filmmaking. <http://speedvr.co.uk/360-vr-filmmaking/lighting-for-virtual-reality-filmmaking/> [Online; accessed 14-September-2016].
 - [41] F. Sukalla, H. Bilandzic, P. D. Bolls, and R. W. Busselle. Embodiment of narrative engagement. *Journal of Media Psychology*, 2015.
 - [42] R. Szeliski and S. B. Kang. Direct methods for visual scene reconstruction. In *Representation of Visual Scenes, 1995.(In Conjunction with ICCV'95), Proceedings IEEE Workshop on*, pages 26–33. IEEE, 1995.
 - [43] R. Szeliski and H.-Y. Shum. Creating full view panoramic image mosaics and environment maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 251–258. ACM Press/Addison-Wesley Publishing Co., 1997.
 - [44] The Verge. Surrounded by sound: how 3D audio hacks your brain. <http://www.theverge.com/2015/12/8021733/3d-audio-3d-binaural-immersive-vr-sound-times-square-new-york> [Online; accessed 14-September-2016].
 - [45] The Verge. This company is using virtual reality to make 360-degree movies. <http://www.theverge.com/2015/1/5/7490557/arkamys-360-degrees-audio-video-virtual-reality> [Online; accessed 14-September-2016].
 - [46] V. Vachiratamporn, R. Legaspi, K. Moriyama, and M. Numao. Towards the design of affective survival horror games: An investigation on player affect. In *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, pages 576–581. IEEE, 2013.
 - [47] J. Vanian. More than 1 million people used the samsung gear vr in april. <http://fortune.com/2016/05/11/oculus-samsung-gear-1-million-users/> [Online; accessed 26-August-2016].
 - [48] VRFocus. Oculus ‘strongly discouraging’ devs from creating jump scares in VR horror. <http://www.vrfocus.com/2016/01/oculus-strongly-discouraging-devs-from-creating-jump-scares-in-vr-horror/> [Online; accessed 14-September-2016].
 - [49] D. Wechsler et al. *Wechsler memory scale (WMS-III)*. Psychological Corporation San Antonio, TX, 1997.
 - [50] A. Weffers-Albu, S. de Waele, W. Hoogenstraaten, and C. Kwisthout. Immersive TV viewing with advanced Ambilight. In *Consumer Electronics (ICCE), 2011 IEEE International Conference on*, pages 753–754. IEEE, 2011.
 - [51] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and virtual environments*, 7(3):225–240, 1998.