Grammar of VR Storytelling: Analysis of Perceptual Cues in VR Cinema

Jayesh S. Pillai
IDC School of Design
Indian Institute of Technology Bombay, India
jay@iitb.ac.in

Manvi Verma
School of Design
UPES Dehradun, India
manvi.verma1936@gmail.com

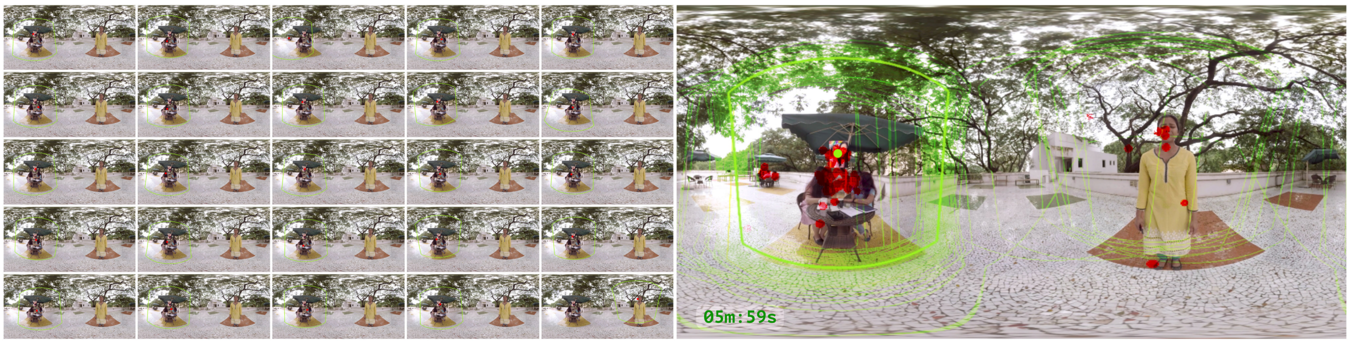

Figure 1: Grid showing the participants' experience (left) and the corresponding combined data for visual analysis (right)

ABSTRACT

With the advent of 360° film narratives, traditional tools and techniques used for storytelling are being reconsidered. VR cinema, as a narrative medium, provides users with the liberty to choose where to look and to change their point-of-view constantly. This freedom to frame the visual content themselves brings about challenges for the storytellers in carefully guiding the users so as to convey a narrative effectively. Thus researchers and filmmakers exploring VR cinema are evaluating new storytelling methods to create efficient user experiences. In this paper, we present, through empirical analysis, the significance of perceptual cues in VR cinema, and its impact on guiding the users' attention to different plot-points in the narrative. The study focuses on examining the experiential fidelity using *"Dragonfly"*, a 360° film created using the existing guidelines for VR cinema. We posit that the insights derived would help better understand the evolving grammar of VR storytelling. We also present a set of additional guidelines for effective planning of perceptual cues in VR cinema.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *User studies*; • **Applied computing → Media arts**.

KEYWORDS

360° Narrative, VR Cinema, Storytelling, Eye-Tracking, Presence, Technical Immersion

ACM Reference Format:
Jayesh S. Pillai and Manvi Verma. 2019. Grammar of VR Storytelling: Analysis of Perceptual Cues in VR Cinema. In *European Conference on Visual Media Production (CVMP '19), December 17–18, 2019, London, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3359998.3369402

1 INTRODUCTION

Virtual Reality (VR) is principally about understanding - whether it is an abstract concept, an entertaining story, or practising a skill [Jerald 2015]. VR researchers and content creators are continually experimenting with methods to leverage 'presence' - the experience of a mediated spatio-temporal reality in VR [Pillai et al. 2013; Sanchez-Vives and Slater 2005; Slater and Usoh 1993; Slater et al. 1994]. These explorations in VR include both interactive content and 360° videos [Brillhart 2016; Pillai et al. 2017]. Recent advancements in video recording and creating technologies have yielded more opportunities for creating 360° video content, otherwise known as omnidirectional videos, cinematic VR or VR cinema. However, the storytelling techniques for VR cinema have been often adapted from traditional frame-bound films and are still in the primitive stages of its evolution towards a well-defined grammar [Pillai et al. 2017]. Planning for VR cinema differs significantly from that of conventional cinema, and is emerging as a domain of interest for researchers and filmmakers [Xu et al. 2018]. There have been previous works that suggest guidelines and techniques for enhancing the VR cinema experience [Gödde et al. 2018; Pillai et al. 2017; Rothe et al. 2017; Vosmeer and Schouten 2017]. However, a detailed and systematic analysis of narratives created using these guidelines will

CVMP '19, December 17–18, 2019, London, United Kingdom

ACM ISBN 978-1-4503-7003-5/19/12…$15.00
https://doi.org/10.1145/3359998.3369402

further validate their effectiveness in VR storytelling [Gödde et al. 2018].

2 BACKGROUND

The factors that contribute to evoking the illusion of reality can be broadly classified into two - perceptual illusion and psychological Illusion [Pillai et al. 2013]. In VR, appropriate stimuli could be placed and presented in a manner that creates this illusion of reality [Jerald 2015]. VR storytellers usually place visual and audio cues in the 360° space to guide the user (experiencer) in a specific manner so as to communicate the narrative effectively. The degree to which the user's personal experience matches the intended experience of the VR storyteller is referred to as 'experiential fidelity' [Lindeman and Beckhaus 2009]. The main difference between a classical and 360° narrative is the 'frame'. In traditional cinema, spectators experience the story through a fixed frame that separates what is visible in the narrative and what is left to the viewers' imagination [Lescop 2017]. In 360° narratives, the experiencer is placed inside the environment of the story, and the frame continuously transforms with the head orientation of the viewer [Vosmeer and Schouten 2017]. Complete freedom lies with the experiencers to look around in a 360° scene and to choose their point-of-view [Ko et al. 2018]. This freedom poses a challenge when filmmakers intend to present a narrative that relies on a specific sequence of events to unfold the plot. Thus the experience of traditional cinema and VR cinema differs extensively [Xu et al. 2018]. Hence the adoption of conventional filmmaking techniques in VR cinema require reconsideration, and this new medium requires a grammar of its own.

2.1 VR Cinema and Grammar

In VR cinema, the relationship between the experiencer and the narrative can not be accurately anticipated. Each viewer will have a version of their own experience as it will never be exactly the same for each individual. A VR cinema experience is not entirely passive (lean-back, as in a traditional film) and nor is fully active (lean-forward, as in the case of video games or VR games) [Vosmeer and Schouten 2017]. It lies in between the two experiences, as shown in Figure 2.

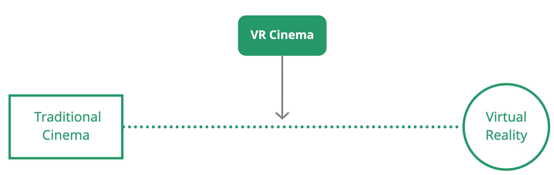

Figure 2: VR Cinema experience with respect to Traditional Cinema and VR

Traditional filmmaking techniques for planning and shooting a film need to be carefully interpreted for a 360° film [Fearghail et al. 2018; Nielsen et al. 2016; Pillai et al. 2017; Xu et al. 2018]. The new medium poses challenges for filmmakers, because in VR the person experiencing the content becomes the storyteller [Brillhart 2016; Gödde et al. 2018]. Creators can never be entirely sure where one is going to look, and a viewer won't always focus where a filmmaker expects one to. Due to the nature of the 360° spaces, some users also experience a fear of missing the plot-points [Rothe et al. 2017]. Likewise, if the pacing is wrong, viewers can either get bored or become completely overwhelmed by a scene [Gödde et al. 2018].

VR cinema has commonalities with immersive theatre, where the viewer feels immersed and co-present with the actors in the 360° space [Ko et al. 2018]. The position of the audience and the surrounding narrative world could be configured in multiple ways to evoke different experiences. Lescop [2017] suggests four such configurations - (1) where the user is placed inside a static 'virtual bubble', (2) where both the user and the virtual bubble move at the same time, (3) where the user moves, but the virtual bubble is fixed, and (4) where the user's position is fixed, but the bubble moves. For a VR cinema created using equirectangular images, the configurations (1) and (4) are befitting as they are essentially panoramic experiences where the user's position is fixed. The configurations (2) and (3) are nevertheless possible using emerging technologies like light field and real-time rendered 3D films.

Concerning the role of the viewer and the associated point-of-view (POV) in a VR film, there are two possibilities of experience. One is them being part of the scene, i.e. first-person perspective, and another as an observer, i.e. third-person perspective. The role of the viewer needs to be clearly defined as both experiences are quite distinct and have a significant impact on the experience [Gödde et al. 2018]. Viewers who are primarily used to traditional movies might passively experience VR cinema. Hence placing of action and story elements in the 360° space are critical. Framing in 360° is correlated to the head direction of the user. Thus the distance between the objects in the virtual environment and the user should also be considered as it might frighten the users or make them empathize more [Sheikh et al. 2016].

VR cinema being a new medium of expression, storytellers have a greater responsibility to leverage this new freedom (that allows the viewers to explore themselves), to intrinsically guide them to have a profound experience [Vosmeer and Schouten 2017]. Therefore a specific grammar for storytelling in VR is essential [Gödde et al. 2018; Ko et al. 2018; Pillai et al. 2017]. Although new innovative techniques and methods are already being experimented with, they are yet to be verified [Gödde et al. 2018].

2.2 Narrative and Perceptual Cues

Narrative and technical aspects of VR cinema support each other and allow users to be immersed in the 360° content [Elmezeny et al. 2018; Ryan 2015]. Narrative immersion focuses on the impact of the structure and content of the story on viewers. It is influenced by setting, place and time of the story, composition of the world, the structure of the plot, genre, and interplay of the story [Lescop 2017]. Technical immersion manifests through perceptual cues to direct and acknowledge viewers attention in the virtual environment [Zhu et al. 2018].

Cues have been used in traditional films, but in VR they are particularly useful in creating attention spots or reinforce the existing points of interest (POIs) and strengthen the kind of user experience filmmakers wish to create [Brillhart 2016]. Cues could be further classified as audio and visual cues. Visual cues include movement of objects or characters, alignment of objects, gaze and gestures of characters, lighting, extreme contrast, semantic opposites, text, graphics, special effects and camera movements. Audio cues include characters conversations, spatial audio, sound effects, and music - both diegetic and non-diegetic [Elmezeny et al. 2018; Gödde et al. 2018; Vosmeer and Schouten 2017].

The gazes of characters work as attentional cues, and if in a specific direction, are most likely to be followed, and guide viewers attention beyond their field-of-view (FOV) [Gödde et al. 2018]. More eye contact also has a satisfying experience for viewers [Ko et al. 2018]. In a 360° space, a viewer is more likely to be attracted by moving objects [Rothe et al. 2017; Xu et al. 2018]. Changing the depth-of-field or artificially creating a partial 'blur' has been a popular method in guiding viewers attention in visual narratives [Fearghail et al. 2018]. In VR, it is possible also to leverage on the stereoscopic disparity [Fearghail et al. 2018], which creates depth cues offering spatial immersion and in turn, stronger emotional immersion [Gödde et al. 2018].

Sound adds to the level of presence and periodical awareness of the surrounding [Bala et al. 2018]. Spatial audio can be an effective tool to guide the viewer to another area of the scene as are directional cues by the characters present within the environment [Dowling et al. 2018; Elmezeny et al. 2018; Fearghail et al. 2018]. Audio cues are also good for alerting the viewer during the narrative [Lescop 2017]. Combination of audio and visual cues is more powerful than visual cues alone, as audio cues are less dependent on the focus of attention at the time of the cue, and could be linked to elements outside the viewer's FOV [Sheikh et al. 2016].

The common practice in film editing defines a continuity domain along the dimensions of space, time and action, which are classified into three different categories. First, edits that are discontinuous in space, time and action (action discontinuities); second, edits that are discontinuous in space or time, but continuous in action (spatial/temporal discontinuities); and third, edits that are continuous in space, time and edits (continuity edits) [Magliano and Zacks 2011; Serrano et al. 2017]. In VR, careful editing for continuity is essential, as frequent scene transitions often cause nausea [Ko et al. 2018]. Hence the transitions have to be natural and psychologically comfortable. By identifying patterns of the shift in viewing behaviour, the beginning and ending points of their experience can be estimated. These beginning and ending points can be intuitively connected for smooth transitions from one scene to another [Brillhart 2016; Pillai et al. 2017]. Alignment of Region of Interests (ROIs) in two consecutive scenes is recommended for fast-paced 360° content, to avoid misalignment during explorations in VR cinema [Serrano et al. 2017]. Maintaining common visual elements across two shots, while they change, would also be a useful technique for effective transitions [Pillai et al. 2017]. Viewers require some time to adapt to the visual content before their gaze fixates on ROIs [Serrano et al. 2017]. This orientation time in a VR setting should also be considered during transitions, which can vary depending on the users and their level of experience with VR [Gödde et al. 2018; Rothe et al. 2017; Vosmeer and Schouten 2017].

A 360° film *"Dragonfly"* was created in our research lab using the existing guidelines for VR cinema which became the foundation of our research experiments. The results were both quantitatively and qualitatively analyzed. In this research, the primary focus is on examining the experiential fidelity in terms of technical immersion, using *Dragonfly*. The objective was to study the impact of perceptual cues - visual, audio and depth cues on viewers experience and their effectiveness in guiding them to follow different plot-points in the narrative. A comprehensive analysis, particularly concerning the users' narrative immersion in *Dragonfly* has been presented in [Pillai and Verma 2019].

2.3 Analysis Methods

Gaze prediction in 360° space is based on temporal and spatial saliency as well as history gaze path [Xu et al. 2018]. Attentional Synchrony is a phenomenon in which gaze behaviours of viewers predominantly cluster around predictable areas. In VR cinema, although viewers are allowed to choose their POV in the 360° space, eye-tracking will enable us to visualize where attentional synchrony (clustering) occurs. Visualizing this clustering can aid identify content or features helpful in guiding the users [Bender 2018]. In a VR setting, viewing behaviour can be studied using patterns of eye gaze and head orientation of the experiencers [Löwe et al. 2015; Xu et al. 2018]. In VR cinema, this data can be then be compared with the director's experience to observe the closeness of the user's experience to that of the intended one. Different visualization methods for the eye gaze and head orientation patterns for VR cinema have been proposed by researchers [Blascheck et al. 2017; Fearghail et al. 2018]. However, the representation of sound in 360° videos is still an open task [Xu et al. 2018]. In our empirical study, eye gaze data and head orientation of the experiencers were recorded to study their viewing behaviour pattern.

2.4 Research Questions

The purpose of this study is to examine and understand technical immersion through perceptual cues and the experiential fidelity associated with the plot-points in VR cinema. The following questions guided the overall research:

(1) How do the perceptual cues guide the users in following different plot-points in the narrative?
(2) How can one predict the users' entry and exit POVs in their experience of VR cinema?
(3) How do users consume the content in 360° space?
(4) How these questions may lead to guidelines for designing effective narrative structure for VR cinema?

Figure 3: Stereo view of a frame in *Dragonfly* (left) with respect to the top-bottom equirectangular frames (right)

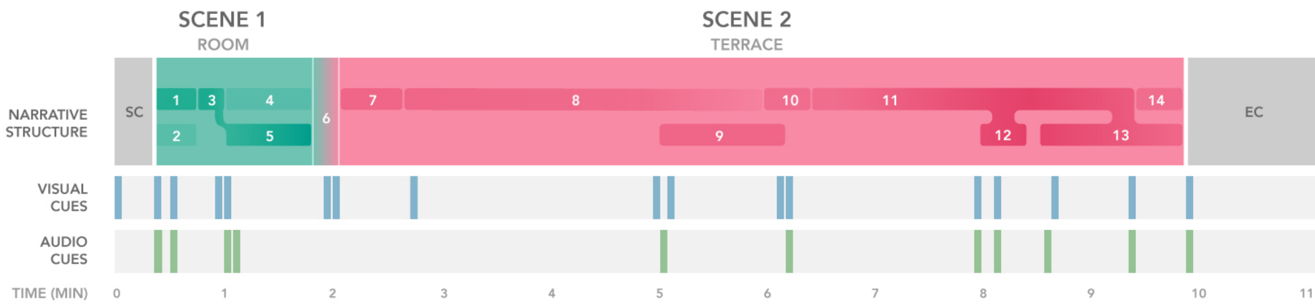

Figure 4: Temporal narrative structure of *Dragonfly*, depicting the primary and secondary plot-points along with the audio and visual cues

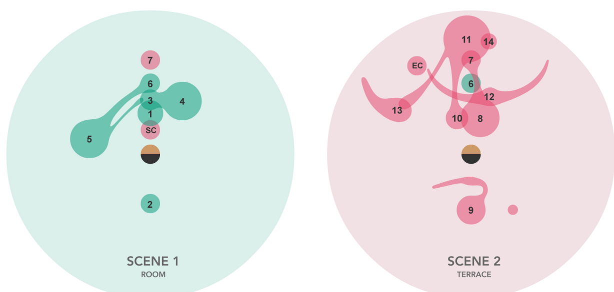

Figure 5: Different plot-points in *Dragonfly* for Scene 1 and Scene 2. The numbers represent the intended POVs and its size represents its duration (a detailed version derived from [Pillai and Verma 2019])

Plot Points	Description
SC	Starting Credits
POV 1	Aisha watching TV
POV 2	A talk-show on TV
POV 3	Aisha making video call
POV 4	Priya on call, driving the car
POV 5	Aisha on call
POV 6	Accident, transition to Dragonfly Title

Plot Points	Description
POV 7	Aisha standing on the terrace
POV 8	Aisha comes closer for the video call
POV 9	Memory, showing Aisha and Priya
POV 10	Aisha walks towards the edge of terrace
POV 11	Aisha on the parapet of the terrace
POV 12	Dragonfly flutters
POV 13	Priya comes on a wheelchair, Aisha meets her
POV 14	Mobile rings
EC	End Credits

Figure 6: POV description corresponding to different plot-points in scene 1 and 2 of *Dragonfly*

To unravel the potential of VR storytelling and to explore these research questions, the film *Dragonfly* [Pillai 2018] was planned and designed, which formed the basis of our study.

3 DESIGNING THE NARRATIVE

The process of planning and creating the VR film *Dragonfly* was carried out entirely in our research lab. Insights from previous studies laid the foundation of our experiment in terms of the narrative structure and arrangement of the plot-points in the film. Previous guidelines, including the ones proposed in [Gödde et al. 2018; Pillai et al. 2017; Rothe et al. 2017; Vosmeer and Schouten 2017], were considered while designing potential perceptual and narrative cues for an intrinsic guided experience.

3.1 Planning the Film *"Dragonfly"*

The film is about Aisha's (protagonist) waiting and yearning for the return of her beloved (Priya) who has undergone a painful incident. *Dragonfly* immerses the viewer in emotions of Aisha - her reality, memories, and hopes [Pillai 2018]. The 360° film experience was designed with stereoscopic 3D and spatial audio to provide the user with an additional sense of depth. Figure 3 shows a stereo view of one of the frames in *Dragonfly* along with the corresponding top-bottom equirectangular images. The primary intention of the narrative was to communicate the emotions of the protagonist effectively. The narrative also traversed the possibilities of different plot-points in 360° space for such a genre exploring emotion and drama.

3.2 Narrative Structure and Plot-Points

The narrative of *Dragonfly* was intended to be experienced as an observer (from a third-person perspective) in a sitting position. Figure 4 shows the narrative structure of *Dragonfly*, with the temporal placement of different plot-points. The figure also presents all the primary visual and audio cues planned at various points during the experience. The narrative is temporally divided into starting-credits (SC), Scene1, Scene2 and end-credits (EC). Different positions for the plot-points were experimented to utilize the 360° space. Figure 5 depicts the top view of different plot-points planned within the two scenes (in a detailed form, derived from [Pillai and Verma 2019]). Scene1 is of approximately 2 minutes duration and is an indoor shot. In this scene, primary plot-points are 1, 3, 5 (dark shade), and secondary plot-points are 2, 4. Plot-point 1 and 2 are 180 degrees spatially apart and occur concurrently in the narrative, and thus, appropriate audiovisual cues were planned to guide the experiencers. Plot-points 3 and 5 are spatially 90 degrees apart. Along with visual cues, spatial audio prompts the experiencer to shift their POV from 3 to 5. Plot-point 6 is the transition from Scene1 to 2. Scene2 is of approximately 8 minutes duration and takes place on a rooftop terrace. Primary plot-points in Scene 2 are 7, 8, 9, 11, 12 and 13. Plot point 8 becomes secondary when story action begins at plot-point 9, and both 8 and 9 are 180 degrees apart. Here as

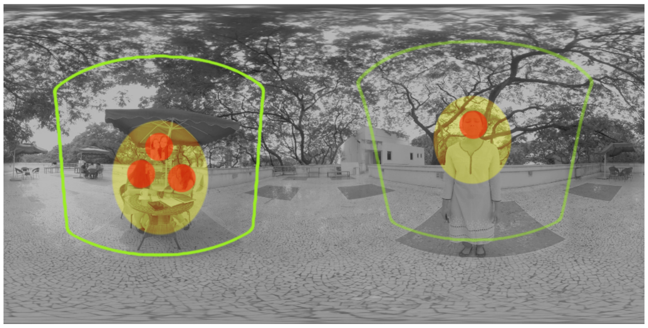

Figure 7: Primary(left) and secondary(right) Intended POVs (green), ROIs (yellow) and POIs (red)

well, spatial audio and gaze of the protagonist (visual cue) were planned to guide the experiencer towards plot-point 9. Plot-point 10 results in 11 and moves from secondary to primary plot-point as the narrative progresses. Story action from plot-point 11 and 13 meet at a common point as shown in figures, and continues till the scene ends. The end credits were spatially placed in between plot-points 13 and 14 (Figure 5) for a seamless transition into the credit sequence. At the time of the transition from Scene1 to Scene2, the direction of the plot-points were aligned in the same direction. In *Dragonfly* we experimented with techniques such as fading of one scene to another, plot-points fading-in and out within a scene, merging of simultaneous plot-points that were initially spatially apart, depth-of-field related blurs to bring attention to the storyline, and using common visual elements to orient experiencers in the intended direction.

3.3 Perceptual Cues

The narrative has a total of sixteen plot-points (including starting and ending credits) spread over two scenes. In order to ensure experience fidelity, and that the user experiences the narrative as close as possible to the director-intended experience, perceptual cues were carefully planned within the narrative. The temporal narrative structure (Figure 4) depicts the exact positions in time where these cues were included to guide the experiencer to focus on primary plot-points and areas of interest.

3.4 POVs and POIs, ROIs

Each plot-point in the narrative connects to a particular POV, towards which the user requires to orient. Description of the POVs corresponding to different plot-points in the narrative is shown in Figure 6. Specific story elements within a POV form potential points-of-interest (POIs) for the user. A collection such POIs in close proximity form a region of interest (ROI) otherwise called an area of interest (AOI). Intended POVs, ROIs and POIs, are based on the director's intended viewing experience. An example of such intended POVs, ROIs and POIs for a specific frame is shown in Figure 7.

4 EXPERIMENT AND EVALUATION

Initially, a pilot study was conducted with five participants, recording only the POV data (and no eye-tracking data) to understand the

Figure 8: Participant experiencing the film *Dragonfly* during the experiment (top). Equirectangular snapshot of the frame with the participant's POV and eye-gaze point (bottom) [Pillai and Verma 2019]

process and to design the main experiment. It helped us in planning the background recording of POV and eye-tracking data as well as framing the right questions for the qualitative interview. In the main experiment, *Dragonfly* was experienced by the participants using a head-mounted display fitted with an eye-tracker, in a closed room on a swivel chair.

4.1 Subjects

Participants, who were primarily from the creative field, were invited for the film experience. A total of 105 users participated in the experiment within the age range of 18 - 61 years. All had a basic understanding of VR, of which about half (56.1%) had previously

Figure 9: Participants watching the film *Dragonfly*, a sitting experience on a rotating chair

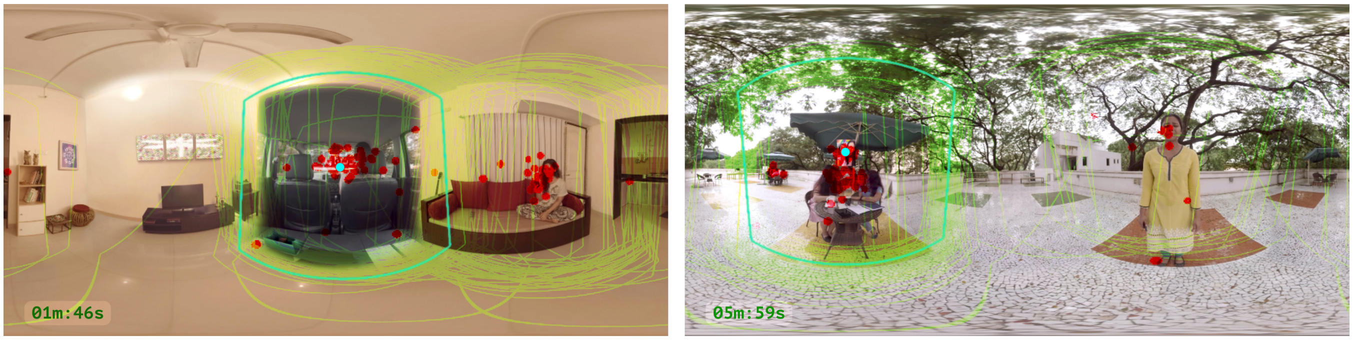

Figure 10: Combined POV and POI data of all participants, along with the Director's POI and POV (left - Scene1 and right - Scene2). The red spots and the green outlines are participant's POIs and POVs, respectively. The cyan spot and the cyan outline is the director's POI and POV respectively.

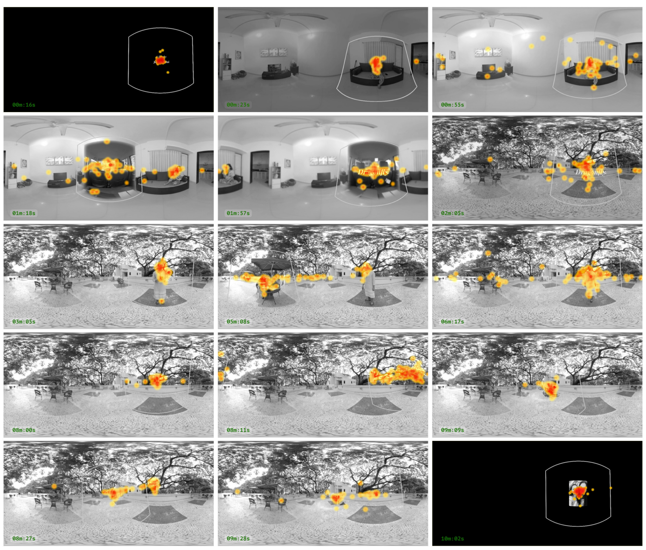

Figure 11: Heatmap for average viewing at specific plot-points

experienced VR. The participants were informed that it was a sitting experience, and they were free to rotate the chair to orient themselves during the experience.

4.2 Apparatus

The entire experience was using an HTC Vive VR headset fitted with aGlass eye-tracker that recorded the eye-gaze data of the participants. Help videos were played using a 360° video player (Vive Cinema), while *Dragonfly* was played using the Unity game engine. During the experiment, using Unity, the gaze positions (POIs) and head orientations (POVs) of the participants were recorded for every second of the film and saved as snapshots of equirectangular images. Figure 8 shows a participant experiencing the film *Dragonfly* during the experiment and the corresponding equirectangular snapshot of the frame having the participant's POV (green rectangle) and eye-gaze point (red dot) data.

4.3 Methodology

An hour session was allotted for each participant. The complete procedure was briefed before starting the experiment. After taking the written consent, participants were seated on a rotating chair (Figure 9). Controllers were not provided to the participants. Before showing *Dragonfly*, participants were shown two short VR films (one live-action and one animated film) to familiarise them with 360° film watching. *Replaced* [2:32 min, live-action film] was shown primarily to encourage the participants to utilise chair rotation and actively watch the film. The trailer of *Invasion* [4:05 min, 3D animation film] was shown to get them to experience stereoscopic 3D in VR cinema. After the help videos, the headset was calibrated for using the eye-tracker. This was followed by the experience of the main VR film *Dragonfly* [11:11 min]. Once the participants completed the film experience, they were interviewed regarding their experience and their understanding of the story. They then filled the post-experiment feedback form. The whole session was video-recorded as well. Director's experience was separately recorded in order to compare the participants' experience to that of the intended one.

5 RESULTS AND ANALYSIS

Qualitative and quantitative data of 99 participants out of 105 was considered, as the eye-tracking data for 6 participants were not properly recorded. Multiple visualizations and statistical methods were used to analyze the cues and understand the users' experience pattern, which are elaborated in the following subsections.

5.1 Visual Analysis - From Combined Experience Data

Visual Analysis gave insights on average viewing pattern of all the participants. In this analysis, eye-tracking data and POVs of all the participants were merged for each frame for every second of the film and compared with the director intended experience. Figure 10 shows two such frames from each scene. Additionally, heatmaps were created to study the pattern of average viewing experience (Figure 11). This helped in visually identifying primary and secondary areas of interests as observed from the average experience of all the participants.

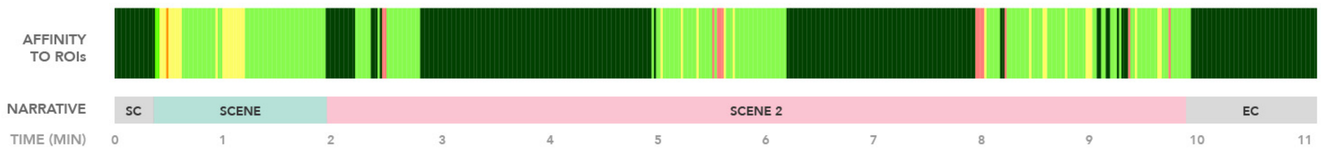

Figure 12: Experiential fidelity with respect to the Intended POVs and ROIs in *Dragonfly* [Pillai and Verma 2019]

5.1.1 Experiential fidelity with Intended POV and ROI. Figure 12 shows the temporal spectrum of experiential fidelity - the affinity (closeness) of the average observed ROIs of viewers with the intended ROIs. Dark green areas represent the times when all the viewers were perfectly oriented towards the intended ROI. Light green represents the times when the majority of the viewers (above 50%) were oriented towards the intended ROIs, while yellow represents low affinity (below 50%). Peach colour represents the times when the observed ROI was completely misoriented with respect to the intended one [Pillai and Verma 2019].

5.2 Statistical analysis - From Individual Experience Data

For in-depth analysis and to quantify previous results, individual viewing experience and how they responded to different cues during the film were visually analysed. For this, a 10x10 grid was created using the screen captures of all the 99 participants and the director's intended experience as the 100th frame. Figure 13 shows this grid for a specific frame, with the POV and POI of all the viewers (95th frame highlighted), and the intended POV and POI (100th frame highlighted). Visual comparisons were made for frames where any visual cues, audio cues or their combinations were present or a shift in POV was expected from viewers to follow the narrative.

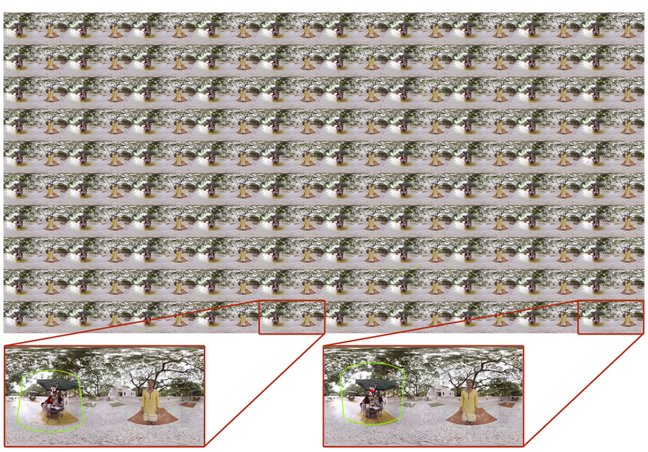

Figure 13: A grid created for visual analysis, showing the POIs and POVs of all 99 participants for a particular frame of *Dragonfly* (an example frame highlighted), with the corresponding intended POI and POV in the last frame (100th frame, highlighted) [Pillai and Verma 2019]

Equirectangular snapshots of individual participants were also studied to understand and quantify the results in terms of the effectiveness of all the cues associated with POV transitions. Individual video recordings of each participant were also analyzed to observe the viewing behaviour and to find correlations to their recorded data. Figure 14 represents the response of the participants during POV transitions (bottom graph), the corresponding reasons that triggered these transitions, and the audiovisual cues that were presented to them (top graph). The reasons observed in the study for a participant's POV shift from one plot-point to another, with respect to the cues planned, are represented using a colour code in the figure. These could be broadly classified into five categories - (1) exclusively visual cues (blue), (2) exclusively audio cues (green), (3) combination of audio and visual cues (Turquoise), (4) exploratory (Pink), and (5) same plot-point (Light pink). The percentage of participants who missed a particular plot-point is shown as light grey

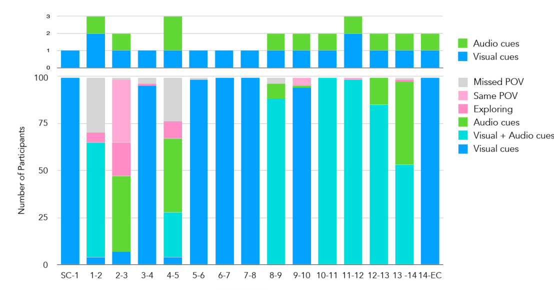

Figure 14: Graph representing the response of participants during POV transitions (bottom) for each cue category (top)

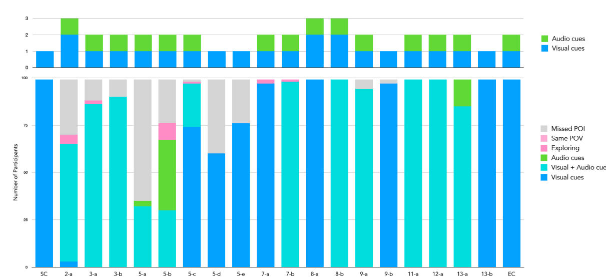

Figure 15: Graph showing the response of participants' POI transitions (bottom) for each cue category (top)

in the graph. The top graph shows the number of audio and visual cues available during each POV transition. A similar observation in terms of the participants' response to a particular POI within a POV is shown in the Figure 15 (bottom graph) and the corresponding audiovisual cues (top graph) that were planned to guide them.

Figure 16 shows the plot-point changes (POV transitions) and the duration of primary story elements within a plot-point (POIs) of the participants, represented on a timeline with respect to the narrative structure of *Dragonfly*. It gives a visual summary of the experiential fidelity in terms of POIs, their duration and the perceptual cues available at a particular instant in the storyline. In order to find the answers to our research questions, inferences were drawn from visual and statistical data analysis, as explained in the section below.

5.3 Observations and Insights

5.3.1 *Observations from Visual Data Analysis.*

(1) Exploratory behaviour was observed at the beginning of both the scenes and was comparatively more common in Scene 1 (in Figure 12 only light green and yellow shades in scene 1 with no dark green areas). This could be due to the presence of simultaneous plot-points placed spatially apart, users' impulse to get familiar with the environment, and also the presence of spatial audio.
(2) Maximum number of participants followed the primary plot-points compared to the secondary ones. The primary plot-points had either the protagonist, an action or some movement, and thus perceptual cues were successful at most of the points in guiding the experiencers.
(3) Users who extensively watch traditional films were observed to have less exploration of the VR environment [Pillai and Verma 2019].
(4) The protagonist's gaze directions guided the participants in facing towards the intended plot-points.
(5) The absence of action prompted participants to either look for a new story action or return to the previously established POV.
(6) In Figure 12 the peach colour representing the areas of complete misorientation is extremely less and comparatively for a shorter duration. The yellow and green areas observed immediately after the peach ones show that perceptual cues helped the experiencers in aligning back towards the intended experience.

5.3.2 *Observations from Statistical data analysis.*

(1) Explorations were more common in Scene1 as compared to Scene2. In Scene 1, 5% of the participants shifted from POV 1 to 2, and 18% from POV 2 to 3 due to exploration. This could be due to the temporally close arrangement of plot-points and POIs in Scene1, and the location of Scene2 being a familiar space for many of the participants.
(2) Audio cues, whether exclusively present or in combination with the visual ones, have been effective to guide viewers.
(3) In the POV 8-9 transition, a combination of visual cue (gaze of the protagonist) and the audio cue was very effective. 88% of the participants followed the audio-visual cue.
(4) In POV 3-4 shift, the movement of the main character was a stronger cue over other action elements. 95% of the participant's head orientation followed the direction in which the protagonist shifted.
(5) POV 4-5 is a 90° shift in terms of placement where the spatial sound was observed to be an effective cue. 63% of the participants responded to the spatial sound.
(6) Shifts POV 2-3 and POV 10-11, fading of a plot-point, or in the absence of action, participants returned to the previously established plot-points.
(7) In POV 1-3 shift, 33% of participants chose to maintain the previous POV where the main character was in the plot-point.
(8) Exclusive visual cues were effective when two POVs were in close proximity to the field-of-view of the experiencer. For instance, in POV 3-4 transition, which was in close proximity, 100% of participants responded to the visual cues.
(9) POV 8-9 is a 180° shift, where spatial audio and gaze proved to be effective cues. 88% of participants responded to the audio-visual cues.
(10) During POV transitions, POIs within the field-of-view were given preference over the elements outside the field-of-view.
(11) The main character as POI was always given a higher priority. For instance, for POI 13-a, 85% of the participants followed the gaze of the protagonist.
(12) In four transitions (SC-POV1, POV5-6, POV6-7, POV14-EC) the arrangement of consecutive plot-points, one at the end of the POV and the other at the start of the next POV, were aligned in the same direction, and there was 100% experiential fidelity.

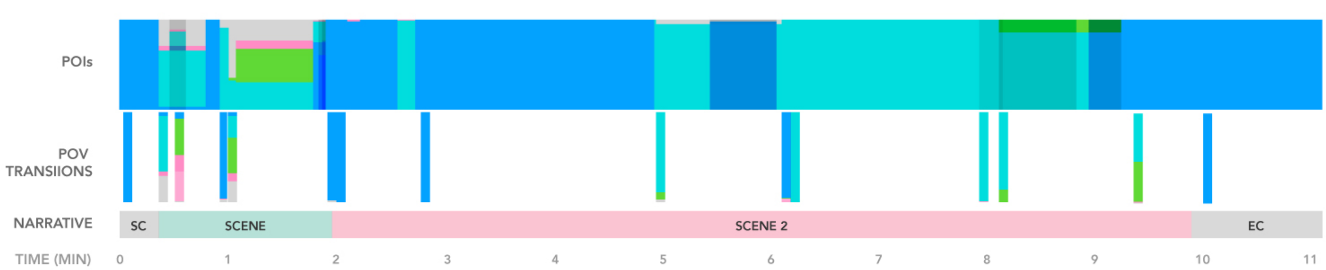

Figure 16: POIs and POV transitions of participants with respect to the narrative structure

(13) The affinity of the average viewing to the intended experience (experiential fidelity in terms of ROIs) was observed to be more in the Scene2.

5.4 Guidelines

Based on the above inferences, the following guidelines are proposed for VR Storytellers, which would be useful in planning the narratives efficiently.

(1) Narrative planning should consider sufficient orientation time for the viewers to get familiar with the 360° space (depending on the narrative features and the level of detail within the scene).
(2) Exclusive visual cues could be planned when two POVs are in close proximity.
(3) The gaze of the main character could be used to guide viewers attention to a particular POV or POI.
(4) Whenever there is a planned shift in POV beyond a viewer's FOV or not in close proximity, spatial audio would be most effective in guiding their attention.
(5) A combination of visual and audio cues is stronger in guiding the user's attention compared to a single visual cue since audio cues are independent of the user's POV and spatial audio cues can originate even from outside the POV.
(6) Having common visual elements across two consecutive shots, during a transition, proved to be extremely effective.
(7) Experimenting with the technique of using a camera shake in a scene which was coherent with narrative and gradually rotating the camera as well proved effective in guiding the experiencers. These techniques are generally not recommended in VR filmmaking guidelines, but when used appropriately, can lead to intended experiences.
(8) Presence of at least one plot-point (either primary or secondary) in the direction of the initial POV, is helpful for users with less exploration in the VR space to still be able to follow the narrative. In such cases, audio plays an important role in guiding them to follow the outline of the story.
(9) Having plot-points of consecutive scenes aligned in the same direction, was effective in achieving intended orientation of the experiencers during scene transitions.

6 DISCUSSION

The gaze of the main character proved to be an important cue in guiding attention towards different POIs and POV transitions [Xu et al. 2018]. Visual cues, before a significant plot-point, proved effective to direct audience attention towards intended POV [Sheikh et al. 2016]. One of the participants could hear only from her right ear. Visual cues were pivotal for her to follow the narrative.

Participants were able to follow the intended viewing experience more in the second scene, as in Scene1 they took some time to get used to the mode of watching a VR film, which is different from the passive laid back traditional film watching.

When the distance between the character and camera was very close, few participants felt uncomfortable and even tried to move backwards [Sheikh et al. 2016]. Participants even had the urge to stop the action of the main character at a pivotal climactic plot-point in the movie, and some were eager to examine the space around the character. Depth due to stereoscopic disparity contributed to emotional immersion as well, which made the experiencers feel more empathetic [Fearghail et al. 2018].

Spatial Audio played an exceptionally important role in guiding the viewers' attention efficiently as intended by the storyteller [Gödde et al. 2018; Pillai et al. 2017; Rothe et al. 2017]. However, in one POV transition, a spatial sound cue intended for 90° POV shift, was mistaken by few participants as a traditional phone call which they heard on their left and hence, restricted themselves from turning to the intended POV.

Few participants, in the beginning, were watching the movie from the first perspective, assuming themselves as an object or a character in the scene. Storyline reshapes with cognitive and emotional inputs [Lescop 2017]. Certain visual cues, like fading of the scene, ending of an action which guided the users or explorations observed at a certain timeline of the movie were mainly because of narrative immersion. This is further discussed in another research article [Pillai and Verma 2019].

The entire experience of watching *Dragonfly* requires a rotating chair, which might not be feasible in all environments. Practical aspects of watching a VR film need to be considered especially at home or in social viewing. Similar to the grammar of VR storytelling, the etiquettes of watching a 360° film without breaking one's immersion is evolving as well.

7 CONCLUSION

In this research, the primary focus was to examine the experiential fidelity in VR Cinema, principally with respect to the perceptual cues planned. We used the 360° film *"Dragonfly"* that was created specifically for research purposes, using the existing guidelines for VR cinema. The objective was to study the impact of perceptual cues on viewers experience and their effectiveness in guiding them to follow different plot-points in the narrative. The experiment was planned to obtain answers to our research questions on the grammar of VR storytelling. The study primarily looked into qualitative analysis with information gathered from eye-tracking, screenshots and session recordings. Results from the experiment verified and supported previously proposed guidelines [Gödde et al. 2018; Pillai et al. 2017; Rothe et al. 2017; Vosmeer and Schouten 2017]. From the observations and analysis, inferences were derived, which led to the proposition of further guidelines for VR storytellers. We believe these guidelines would be helpful for VR filmmakers in planning 360° narrative experiences efficiently.

REFERENCES

Paulo Bala, Raul Masu, Valentina Nisi, and Nuno Nunes. 2018. Cue Control: Interactive Sound Spatialization for 360° Videos. In *International Conference on Interactive Digital Storytelling*. Springer, 333–337.
Stuart Bender. 2018. Headset attentional synchrony: tracking the gaze of viewers watching narrative virtual reality. *Media Practice and Education* (2018), 1–20.
Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2017. Visualization of eye tracking data: A taxonomy and survey. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 260–284.
Jessica Brillhart. 2016. The Language of VR: Concepts and Ideas. (2016). Retrieved April 7, 2019 from https://medium.com/the-language-of-vr.
Declan Dowling, Colm O Fearghail, Aljosa Smolic, and Sebastian Knorr. 2018. Faoladh: A Case Study in Cinematic VR Storytelling and Production. In *International Conference on Interactive Digital Storytelling*. Springer, 359–362.
Ahmed Elmezeny, Nina Edenhofer, and Jeffrey Wimmer. 2018. Immersive storytelling in 360-degree videos: an analysis of interplay between narrative and technical immersion. *Journal For Virtual Worlds Research* 11, 1 (2018).

Colm O Fearghail, Cagri Ozcinar, Sebastian Knorr, and Aljosa Smolic. 2018. Director's cut-analysis of aspects of interactive storytelling for vr films. In *International Conference on Interactive Digital Storytelling*. Springer, 308–322.

Michael Gödde, Frank Gabler, Dirk Siegmund, and Andreas Braun. 2018. Cinematic Narration in VR - Rethinking Film Conventions for 360 Degrees. In *International Conference on Virtual, Augmented and Mixed Reality*. Springer, 184–201.

Jason Jerald. 2015. *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

Dong-uk Ko, Hokyoung Ryu, and Jieun Kim. 2018. Making New Narrative Structures with Actor's Eye-Contact in Cinematic Virtual Reality (CVR). In *International Conference on Interactive Digital Storytelling*. Springer, 343–347.

Laurent Lescop. 2017. Narrative grammar in 360. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*. IEEE, 254–257.

Robert W Lindeman and Steffi Beckhaus. 2009. Crafting memorable VR experiences using experiential fidelity. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. ACM, 187–190.

Thomas Löwe, Michael Stengel, Emmy-Charlotte Förster, Steve Grogorick, and Marcus Magnor. 2015. Visualization and analysis of head movement and gaze data for immersive video in head-mounted displays. In *Proceedings of the Workshop on Eye Tracking and Visualization (ETVIS)*, Vol. 1. Citeseer.

Joseph P Magliano and Jeffrey M Zacks. 2011. The impact of continuity editing in narrative film on event segmentation. *Cognitive Science* 35, 8 (2011), 1489–1517.

Lasse T Nielsen, Matias B Møller, Sune D Hartmeyer, Troels Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ACM, 229–232.

Jayesh S Pillai. 2018. *Dragonfly, 360° VR Film.* Retrieved December 7, 2018 from https://dragonflymovie.in/.

Jayesh S Pillai, Azif Ismail, and Herold P Charles. 2017. Grammar of VR Storytelling: Visual Cues. In *Proceedings of the Virtual Reality International Conference-Laval Virtual 2017*. ACM, 7.

Jayesh S Pillai, Colin Schmidt, and Simon Richir. 2013. Achieving Presence through Evoked Reality. *Frontiers in psychology* 4 (2013), 86.

Jayesh S Pillai and Manvi Verma. 2019. Grammar of VR Storytelling: Narrative Immersion and Experiential Fidelity in VR Cinema [under review]. In *17th ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI) 2019, Brisbane, Australia*. ACM.

Sylvia Rothe, Heinrich Hussmann, and Mathias Allary. 2017. Diegetic cues for guiding the viewer in cinematic virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 54.

Marie-Laure Ryan. 2015. *Narrative as virtual reality 2: Revisiting immersion and interactivity in literature and electronic media*. Vol. 2. JHU Press.

Maria V Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience* 6, 4 (2005), 332.

Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. 2017. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 47.

Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. 2016. Directing attention in 360-degree video. (2016).

Mel Slater and Martin Usoh. 1993. Presence in immersive virtual environments. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 90–96.

Mel Slater, Martin Usoh, and Anthony Steed. 1994. Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 3, 2 (1994), 130–144.

Mirjam Vosmeer and Ben Schouten. 2017. Project Orpheus a research study into 360 cinematic VR. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. ACM, 85–90.

Yanyu Xu, Yanbing Dong, Junru Wu, Zhengzhong Sun, Zhiru Shi, Jingyi Yu, and Shenghua Gao. 2018. Gaze prediction in dynamic 360 immersive videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 5333–5342.

Yucheng Zhu, Guangtao Zhai, and Xiongkuo Min. 2018. The prediction of head and eye movement for 360 degree images. *Signal Processing: Image Communication* 69 (2018), 15–25.