Grammar of VR Storytelling: Visual Cues

Jayesh S. Pillai
IDC School of Design, IIT Bombay
Powai, Mumbai 400076, India.
+91 2225767804
jay@iitb.ac.in

Azif Ismail
IDC School of Design, IIT Bombay
Powai, Mumbai 400076, India.
+91 2225767801
azifismail@iitb.ac.in

Herold P. Charles
IDC School of Design, IIT Bombay
Powai, Mumbai 400076, India.
+91 2225767801
heroldpc@iitb.ac.in

ABSTRACT
Visual storytelling is an integral part of the experience in media such as comic books, movies, video games and virtual reality. The grammar of storytelling that has been explored abundantly in virtual reality has close relationship with that of video games. However, while designing content for virtual reality video experiences one must understand and consider the fundamental distinctions between 360° panoramic storytelling and a traditional frame-bound visual narrative. In this study, we attempt to understand the visual cues in a virtual reality space and propose guidelines that would guide the viewer in orienting towards the intended point-of-view.

CCS Concepts
• **Human-centered computing ~ Virtual reality** • **Human-centered computing ~ HCI theory, concepts and models** • *Human-centered computing ~ Mixed / augmented reality* • *Human-centered computing ~ Scenario-based design*

Keywords
Virtual Reality; Visual Storytelling; 360° Narrative; Presence.

1. INTRODUCTION
In the recent few years the cameras and gears to record 360° panoramic videos as well as platforms to experience them have become increasingly accessible. This easy access has quickly populated the internet with numerous 360° videos explored in wide range of genres. Moreover, currently the availability of inexpensive 360° video cameras has opened up these opportunities for even more people, leading to a large amount of 360° experiences that are neither planned nor designed with specific narrative structure. In many such videos that are non-structured, the experiencer is left clueless in the space surrounding him/her with no specific purpose or point-of-view.

Thus storytelling in Virtual Reality (VR) has become a primary area of interest for many researchers and filmmaking professionals [1,3,4].

VRIC '17, March 22–24, 2017, Laval, France

ACM ISBN 978-1-4503-4858-4/17/03…$15.00
http://dx.doi.org/10.1145/3110292.3110300

2. BACKGROUND
Although 360° panoramic images have been explored for many years, 360° videos became popular in the recent years. Since the technology to record and experience 360° videos advanced rapidly, research on grammar and content was left unexplored.

We believe that the grammar for 360° videos will develop to be stronger in the coming years with respect to the content; both conceptual and perceptual.

2.1 Storytelling in VR
Although there have been countless studies on the experience of 'Presence' in VR [9,10], the research on grammar and storytelling in VR was gradual until 360° video related technologies became more accessible. However, to create effective 360° video experiences one must also understand the grammar of traditional frame-bound visual narratives.

2.1.1 VR Cinema
Sensing its potential for storytelling, many professional filmmakers previously working on frame-bound experiences have already taken the leap into this medium [7,8]. They are now in the process of learning the art of creating compelling and memorable experiences in such 'evoked reality' [10] that surrounds us.

We posit that storytelling through this new medium of 360° video falls right in between a movie and a VR experience.

A VR experience is interactive, with the user having full control over the virtual environment [9]. In movies, we may be 'passive observers' while in VR we become 'active participants'. If so, in a 360° video space we would be 'active observers' with a limited amount of control - the freedom to change our point-of-view (POV) at any point in time. This being an experience in between VR and movies, from here on we would refer to it as *VR Cinema*.

2.2 Beyond the Frame
The primary difference between a traditional movie and a VR Cinema is concerning the perceptual boundary. Movies have a fixed rectangular 'frame' within which the audience attention is focused. The film-grammar that evolved over a century takes into account various aspects of this visual frame. For the frame-bound cinema, guidelines and techniques have been well-developed to efficiently create and sustain audience attention within the frame (and even within different parts of the frame).

2.2.1 The Visual Grammar
A frame-bound movie is one's 'window' to the world where the story occurs, while a VR Cinema is one's own 'perspective' of that world, with the freedom to look in any direction. So, in VR Cinema one feels present inside that world and perhaps be part of the story itself.

For the same content in VR there could be various possibilities of POV with each one of us having purely subjective experience. As rightly expressed by Jessica Brillhart [4] "In VR, a "frame" is a relative window of experience derived from the visitor's field of vision. This makes everything a potential frame, but also makes a premeditated frame based on my own interests presumptuous and, well, wrong most of the time". Thus the very freedom of shifting POV can as well defeat the intended storytelling experience if not well-designed [6]. It has led designers and researchers to look deeper into the visual grammar of VR Cinema [1,3].

2.3 Research Questions

A good understanding of grammar is required for efficient storytelling [2,5]. What are the fundamental differences between a 360° panoramic and a traditional frame-bound visual narrative? What aspects can we learn and adopt from the well-evolved grammar of the frame-bound movies?

We propose that appropriate visual, audio and other narrative cues will contribute to a better experience in VR Cinema. As a first step in the process of understanding the potential grammar of VR Cinema with various possibilities of POV, we attempted to study the visual cues. We hypothesize that appropriate visual cues in a virtual reality space can guide the viewer towards the intended point-of-view.

3. THE EXPERIMENT

For the experiment to study visual cues, a 3-minute music video was visualized and developed as a 360° panoramic experience.

3.1 The VR Cinema

3.1.1 Content

The music video explores the emotions of a person relishing the time he spent by a lake side. The narrative was planned to provide the essence of the space surrounding the lake along with the ambience at different times of a day, complimenting the lyrics of the song.

3.1.2 Designing the Experience-Intended

The basic narrative structure of the video followed the sequence: A logo, establishing shot of the lake, title of the video, morning scene, night scene, noon scene, concluding shots shifting between different times of the day and end credits. Figure 1 illustrates the narrative structure and the potential visual cues along with a sequence of the "intended POVs" ('intended' here refers to 'as envisioned by the director or creator'). We identified 11 instances of POVs and corresponding frames (F1, F2, …F11) at different points in time during the experience, to be significant for the analysis of the effect of visual cues.

The video was shot with a panoramic camera at the premises of a lake where the protagonist is seen taking a stroll, signing the song.

3.2 Experimental Setup

3.2.1 Subjects

The video was experienced by 20 participants within the age group of 18-50 years. All the participants had basic knowledge of VR but had seldom used a head mounted display (HMD) device.

3.2.2 Methodology

The participants were informed beforehand that the experience would be a 360° music video. However, to avoid observer effects they were not informed until after the session that the entire experience was screen-recoded for the study. They were provided with a HMD attached with headphones. As the song was in the Indian language Tamil, the video also had the the option of viewing subtitles if required.

3.3 Analysis & Results

The recording of each participant's experience (E1, E2, …E20) were then analyzed for its closeness to the intended experience so as to understand whether the visual cues worked; and the reason why they worked or not. Using a grid of 10x10 units overlaid on every video, the participants' POVs were examined for each of the 11 points (F1, F2, …F11) where responses to visual cues were expected. Figure 2 shows an example of frame F9 of all 20 participants' POVs with the intended POV at the center.

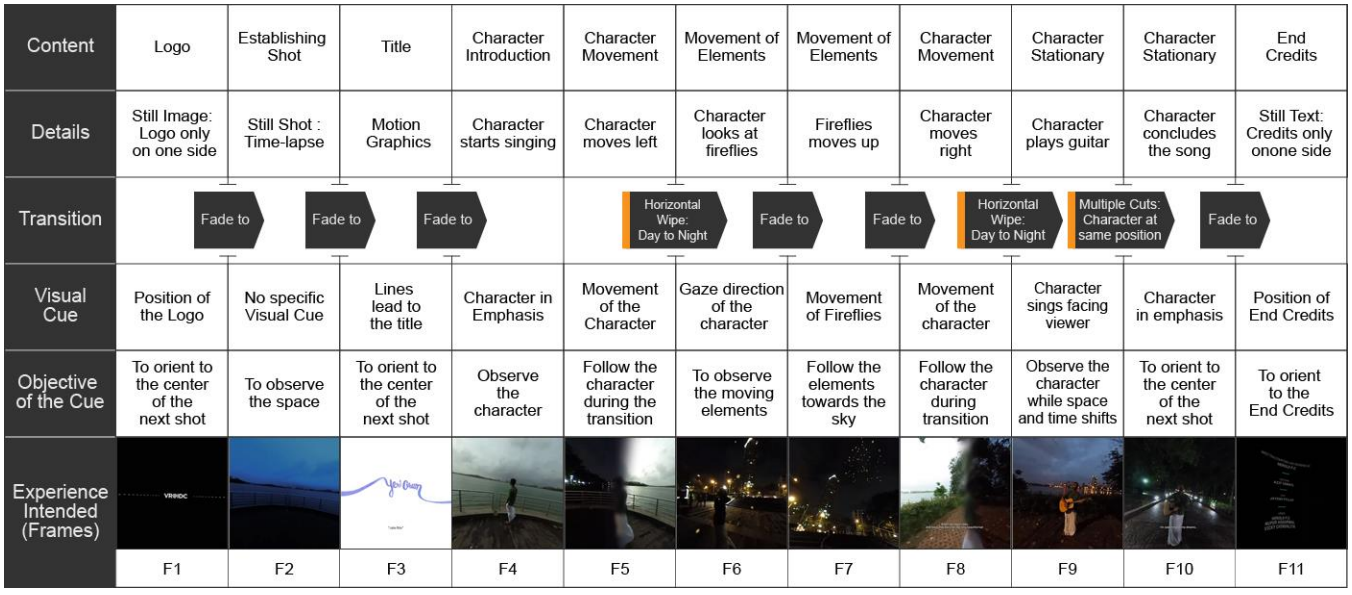

Figure 1. The Narrative Structure of the 360° Video

Figure 2. All POVs with intended POV at the center for F9

Primarily two methods were employed for the analysis - to understand (1) how close the POVs of the participants were to the originally intended POVs and (2) whether the POVs of the participants successfully contained the intended Center-of-Interest.

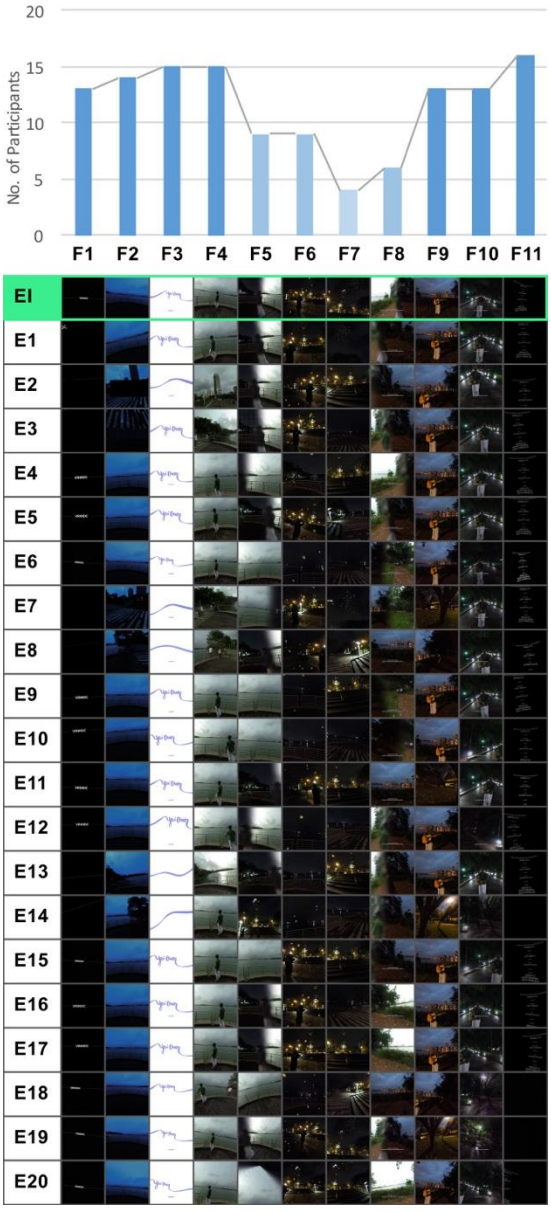

Figure 3. Closeness to Intended POV

3.3.1 Closeness to Intended Point-of-View

The corresponding frames (F1, F2, …) of POVs were analyzed for each participant's experience (E1, E2, …) with respect to the experience-intended (EI). Each frame was checked whether it shared a common area of at least 50% or above with the respective frame of the intended POV. Figure 3 shows the closeness of the POVs of all experiences to the originally intended POV. On evaluation, the POVs of more than half of the participants were found to be close to the intended POVs. Accordingly, inferences were derived (see 4.1 below).

3.3.2 Closeness of Centre-of-Interest

Additionally, corresponding frame of each intended POV was compared in detail with the respective frames of the experiences to understand whether the visual cues helped in orienting towards the intended center-of-interests. For this study, the central 25% area of each frame of the intended POV was considered as the center-of-interest. Each frame of the participant's experience was checked whether they contained the intended center-of-interest entirely. Figure 4 shows the closeness of the center-of-interests of all the experiences to the intended center-of-interest. It was noticed that for the frames F5, F6, F7 and F8 the intended center-of-interest was not noticed by most participants. The insights derived are discussed in section 4.1 below.

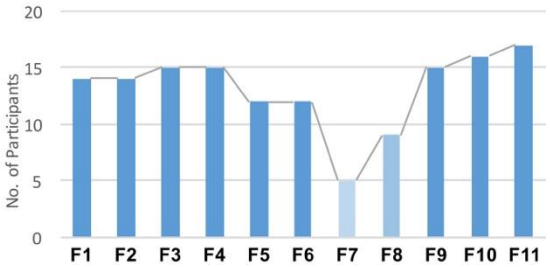

Figure 4. Closeness to Intended Centre-of-Interest

The experience being a retrospective music video may have constrained our inferences to certain related genres. However, we believe that it would help us understand visual cues from a broader perspective as well. Apart from the direct inferences, few subjective issues were also noted. For instance, the animation of fireflies spanned for about a minute with no specific visual cue till the last few seconds, which led to distraction from the intended POV. Certain parts recorded in low-light led to underexposed shots where the character movements were uncertain to few participants. The resolution of the 360° video was non-stereo 1920x960 pixels, which if higher would have presented with more definition to the visual quality. Hence the effect of resolution, field-of-view and depth perception with respect to stereo 3D were not explored.

4. INFERENCE AND DISCUSSION

4.1 Insights

4.1.1 Effective Visual Cues

Position of a considerably contrasting graphics or logo within a static environment enabled participants to orient towards it (F1 to F2). Motion within static environment helped participants to orient towards the intended POV (F3 to F4). Character addressing the viewer directly created and sustained interest for a longer time (F9). Continuity of a common element irrespective of the change

in space helped sustain the interest at that POV (F10). Visual cues that led to intended POV at the end of a shot were helpful in transition to the next shot orienting intuitively towards the intended POV (F10 to F11).

4.1.2 Ineffective Visual Cues

More deviation from the intended POV was observed during the time without much action (F6) or during the time with prolonged similar action (F5 & F7). Multiple elements within a space distracted the attention (F7). Those who were incorrectly oriented at the end of a shot, missed the intended center-of-interest of the next shot especially when the duration of the latter shot was less (F7 to F8).

4.2 Guidelines

With respect to the insights gathered, we propose certain guidelines with respect to visual cues in a VR Cinema.

1. Visual Focus: As in traditional frame-bound experiences, specific points in both time and space require visual centers-of-interest for guiding one's attention effortlessly.
2. Elements of Interest: Multiple elements within any potential POV must be presented with a hierarchy of importance for intuitive shifts in attention. They help bring the viewer's focus back to the story, if drifted away in the 360° virtual space.
3. Character addressing the viewer: The subject facing directly the viewer creates and sustains interest.
4. Action and Duration: A balance between action (or no action) and how long it spans is crucial.
5. Motion to Orient: Appropriate motion of elements would help orient towards the intended POV.
6. Cuts: The POV at the end of a shot must help orient to the intended POV of the next shot.
7. Transitions: A transition from one shot to another would be more effective when certain elements or subjects within the two shots are associated visually.
8. Orient with Cut: Irrespective of the observer's POV at the end of a shot, the following shot could be programmed to orient itself such that the intended POV is achieved at the beginning of that shot.

4.3 Potential of the Grammar

Apart from visual cues, the study led us to understand related facets that may influence the grammar of of VR Cinema and would help enhance the experience in future.

1. Interactive VR Cinema: Seamlessly integrating interactivity within a film experience has been attempted numerous times with experimental storytelling techniques and must be explored as part of research on VR Cinema as well.
2. Primary & Secondary Experience-Intended: VR Cinema has the possibility to offer viewers a unique experience for every new viewing of the same story or content, which presents with opportunities to further explore novel storytelling techniques. So multiple intended experiences could be planned, especially if the content deals with parallel stories and subplots.
3. Storyboarding and Visualization: The potential of unique experience on multiple viewing also demands innovative storyboarding techniques to evolve.
4. Language & Subtitles: How we intergrade text to eliminate language barriers must also evolve in accordance with 360° panoramic experiences.
5. Audio Cues: Audio cues, similar to its usage in traditional films, must become an integral part of VR Cinema as they definitely help efficiently orient towards intended POVs.
6. Narrative Cues: Story related cues would evolve that can utilize the potential of immersive 360° experiences.
7. Physical Comfort: Logistics of physical comfort and the usage of HMDs are still progressing and perhaps need to be customized for specific experiences.
8. Genre and Style: Currently VR cinema has already begun exploring new frontiers, boldly venturing into new genres beyond the clichéd.

5. CONCLUSION

In this study we proposed that appropriate visual, audio or narrative cues will contribute to a better POV enhancing the VR Cinema experience. Accordingly, we probed into the possibilities of visual cues. From this study we found that appropriate visual cues can lead the viewer intuitively towards the intended POVs. Visual cues if well-planned would help storytellers design compelling experiences for VR Cinema. This is one of the many steps towards understanding the potential of storytelling in VR. We believe that similar to the evolution of grammar of traditional frame-bound movies, the grammar for VR Cinema will develop to be more efficient in the coming years.

6. REFERENCES

[1] Mike Alger. 2015. Visual Design Methods for Virtual Reality. Retrieved September 1, 2015 from http://mikealger.com/professional

[2] Daniel Arijon. 1976. *Grammar of the Film Language*. Focal Press, London.

[3] Jessica Brillhart. 2015. The Language of VR: Concepts and Ideas. Retrieved December 7, 2015 from https://medium.com/the-language-of-vr/tagged/storytelling

[4] Jessica Brillhart. 2016. The Language of VR: In the Blink of a Mind. Retrieved July 6, 2016 from https://medium.com/the-language-of-vr/tagged/editing

[5] Neil Cohn. 2013. Visual Narrative Structure. *Cognitive Science* 37, 3: 413–452. https://doi.org/10.1111/cogs.12016

[6] Joel Gunz. 2016. Just Because You Can Make 360 Cinema Doesn't Mean You Always Should. Retrieved November 8, 2016 from https://medium.cinematicvr.org/360-cinema-and-a-case-for-limited-view-virtual-reality-immersion-839d7f7b5593

[7] Chris Milk. 2015. How Virtual Reality can Create the Ultimate Empathy Machine. Retrieved May 20, 2003 from https://www.ted.com/talks/chris_milk_how_virtual_reality_can_create_the_ultimate_empathy_machine

[8] MindShow. 2016. Mindshow: First Look. Retrieved August 5, 2016 from https://youtu.be/KdK_C0Kr8Wo

[9] Jayesh S Pillai. 2014. Réalit evoquée, des rêves aux simulations : un cadre conceptuel de la réalité au regard de la présence. Arts et Métiers ParisTech.

[10] Jayesh S. Pillai, Colin Schmidt, and Simon Richir. 2013. Achieving presence through evoked reality. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00086