Cinévoqué: Design of a Passively Responsive Framework for Seamless Evolution of Experiences in Immersive Live-Action Movies

Jayesh S. Pillai, Amarnath Murugan, Amal Dev

Cinévoqué: Design of a Passively Responsive Framework for Seamless Evolution of Experiences in Immersive Live-Action Movies

Jayesh S. Pillai[1[✉]], Amarnath Murugan[2] and Amal Dev[3]

[1] IDC School of Design, IIT Bombay
[2] Next Tech Lab, SRM Institute of Science and Technology
[3] Design Discipline, IIITDM Jabalpur
jay@iitb.ac.in

Abstract: In this paper, we present a passively responsive framework for immersive movies, called Cinévoqué. The framework seamlessly alters the narrative and visual elements within an immersive live-action movie, based on real-time passive data, such as gaze direction and system time, obtained during the experience. The paper primarily focuses on the design and storytelling aspects of Cinévoqué, such as possible narrative structures and the design challenges involved in creating responsive experiences. We further examine the potential of this framework through two prototypes of varying complexity and responsive features, and the insights from them are used to suggest approaches that can lead to effective seamless narrative experiences.

Keywords: Virtual Reality, VR Storytelling, Responsive Narrative, Presence.

1 Introduction

Traditional non-immersive movies guide the viewers through the story by displaying events within a fixed rectangular frame determined by the filmmaker, thereby making them focus on the intended points-of-interest (POIs). In contrast, within an *immersive movie*, experienced in Virtual Reality, the viewers (users) are completely surrounded by the environment of the narrative. Due to its nature of pushing the user beyond being merely an observer, it gives them complete freedom to change the point-of-view (POV) throughout the experience. Immersive movies being relatively new and less explored than frame-bound movies, the storytelling and filmmaking techniques are still evolving [1,2]. Studies on the grammar of VR cinema point towards the potential of new storytelling opportunities [3]. The idea of a responsive narrative that we explore here stems from the very basic properties of an immersive movie experience.

In this study, we present a novel framework based on the concept of seamlessly evolving narratives in live-action immersive movies, that we call Cinévoqué. The term Cinévoqué is derived from Cinema Évoqué (evoked cinema) to represent a form of responsive narrative that evolves based on the user's passive interactions, such as changes in POIs throughout the experience. Although our study explores storytelling as well as technical challenges in creating a Cinévoqué experience, in this paper, we focus mainly on the storytelling aspects, i.e. narrative structures and approaches. In a spectrum of immersive experiences arranged in the order of their interactivity, ranging from a 360° immersive movie to a high-end VR experience (Fig. 1), we believe that a Cinévoqué experience falls between an immersive movie and an interactive immersive movie experience which allows the user to make conscious choices about the progression of the story.

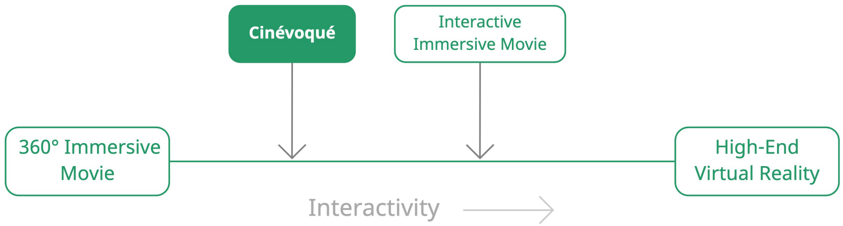

Fig. 1. Interactivity Spectrum

2 Background

Within an immersive movie, the users' ability to choose their own POV makes it difficult to ensure that they look at the important events in the experience and are following the story consistently. VR filmmakers and researchers have tried to address this problem through visual & auditory cues [3,4] and through specific editing techniques. For instance, the intended POIs across consecutive scenes are aligned to be in the same viewing direction so that the users are more likely to not miss them [5]. Though these approaches may be effective in directing the attention of the majority of users, it may still not guarantee that all of the users would follow the intended POIs consistently. Another drawback of these techniques is that they become less effective when there are large number of POIs within a scene [6].

Some techniques try to force the user to look in a specific direction by limiting the information in the other locations within the immersive space, or by displaying UI elements within the scene that point to the intended direction. However, sometimes these techniques were observed to cause discomfort to the users [7,8]. Cinévoqué takes a different approach by using real-time data such as the user's gaze direction, to seamlessly adapt the experience to the user. It dynamically modifies the narrative and other elements within the environment based on the POIs the user has focused on.

2.1 Responsive Narratives

The Cinévoqué experience is conceptualized as a narrative that dynamically responds to passive interactions from the user, unbeknownst to them, as opposed to actively interactive movies that allow the users to make conscious choices about the progression of the narrative [9]. Thus to the user, a Cinévoqué experience would seem to be the same as watching a normal VR movie.

The concept of seamlessly switching storylines based on passively collected data has been explored in traditional movies through the use of eye tracking to measure the user's interest to alter the narrative [10], or by using data from sensors that measure bio-signals of the viewers [11,12]. Creating a responsive narrative within traditional movies require specialized hardware on top of the system running the experience. In contrast, all immersive VR experiences implicitly generate the core data used by Cinévoqué, which makes it relatively more accessible to end users.

Researchers have previously proposed the application of a passively responsive narrative approach to live-action VR movies [3] and some have explored the idea in parallel to a limited extent [13]. Cinévoqué explores a similar direction while investigating the fundamental design and development challenges that exist in such a system.

3 Cinévoqué

Similar to interactive movies, the narrative structure of a Cinévoqué experience would also have a branched construct, that allows for multiple possible storylines. The storylines within Cinévoqué are structured as directed graphs i.e. the overall story is divided into smaller segments which point to one or more segments. This structure is broadly composed of *Levels* and *Nodes* which contain certain condition-based entities called *Hotspots* that influence the evolution of the storylines.

3.1 Narrative Structure - Levels & Nodes

A node is a segment/snippet of the larger possible experience and a level is a collection of multiple nodes which are alternate possibilities at the same point of the story. The story always begins from a *start level* and concludes on a *terminal level*. If the start level contains multiple nodes, one node is chosen based on data collected prior to the beginning of the experience or at random.

Fig. 2. shows the structure of a simple Cinévoqué experience, here L1 is the start level, followed by the terminal level L2. Each node is denoted by the nomenclature *L(level number)N(node number)*. The lines diverging from a node lead to the possible nodes that it can transition to. Each node has a predetermined default *transition time* and *transition duration* associated with it. When the node reaches the default transition time it transitions to a node in the succeeding level over the given transition time. The transition to a particular node would primarily be influenced by the user's interest in certain events or objects within the immersive environment. These events/objects are associated with Hotspots.

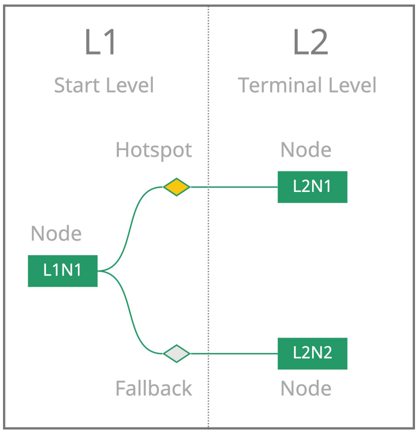

Fig. 2. Levels and Nodes in a basic Narrative Structure

3.2 Hotspots

A hotspot is a region (predefined or dynamically marked) within the immersive environment of a node that has a set of conditions, that when met (triggered), helps the system make changes to the user's experience dynamically. A hotspot could exist throughout a node or for a specific duration depending on the event/object it's associated with. Each hotspot can lead to one or more target nodes, of which one would be played after the current node if that particular hotspot is triggered. The target node of the hotspot could exist in any level, so the storyline could skip levels or go back if required. If multiple target nodes are given, one final target is chosen based on the conditions set by the storyteller or at random (if the narrative requires it). A hotspot could also be used to alter the dynamic elements within its own node or in the nodes that follow. A node may contain one or more hotspots that influence the experience. In Fig 2., the experience would start from the node L1N1 and would conclude at L2N1 if the hotspot leading to it is triggered. If none of the hotspots in a node are triggered, the experience would transition to a predetermined node called *fallback node*. In Fig 2., the node L2N2 is the fallback. Cinévoqué consists of multiple types of hotspots that operate primarily based on the user's POV and gaze. In an experience where gaze-time is used to derive the user's interest, the hotspots could be broadly classified as *timer-based* and *immediate*. If other factors are used to derive user's interest (such as heart rate, skin conductance, Brain-Computer Interface, etc.), the classification of the hotspots would vary accordingly.

A *timer-based hotspot* measures the number of seconds it has been gazed at. If a node were to contain multiple timer-based hotspots, then the next node would usually be the target node of the hotspot which was gazed at the most. The hotspots can also be assigned different priorities within a node if a certain storyline is to be given more importance. Hotspots with a specific gaze-time threshold could have the ability to override the conditions met by other hotspots with equal or lower priorities. In this case, the transition to the next node always starts at the predetermined transition start time.

An *immediate hotspot*, as the name implies, would immediately decide the next node, thereby overriding all the other hotspots. In this case, the transition to the next node can start either immediately or after a certain amount of time specified by the storyteller.

4 Prototypes

The framework which was developed in the game engine Unity3D was utilized to create two prototypes in order to test our concept and investigate storytelling possibilities. These prototypes were deployed in both mobile VR and HTC Vive to be experienced by the users.

4.1 Schrödinger's Vada-Pav

Schrödinger's Vada-Pav is the first prototype that was built using Cinévoqué. In this experience, the user is introduced into the first person view of a character (Schrödinger) working in an office, and a waiter walks in and takes an order (a vada-pav) from him. After some time, the waiter returns with the vada-pav and knocks on the office door, noticing that the character is busy. From this point, there are two possible outcomes to this narrative. If the user gazes at the waiter within a certain amount of seconds after the knock, he would walk in and place the vada-pav near the character. If the user were to concentrate on something else, the waiter would leave without delivering the vada-pav.

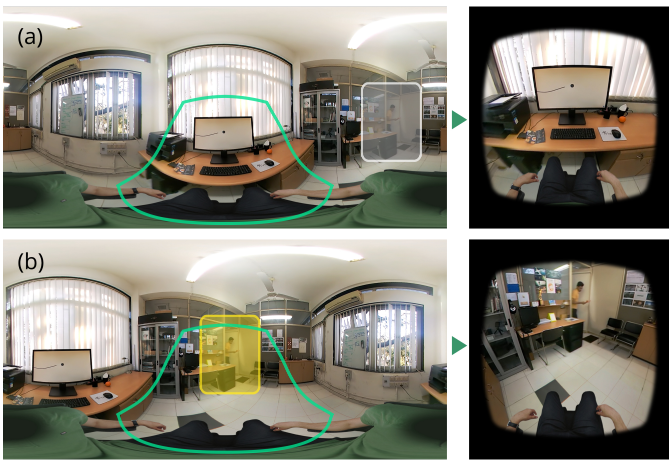

Fig. 3. a) Equirectangular frame with an active hotspot (grey region), the region of user's POV (within green outline) and corresponding user's view (right) in *Schrödinger's Vada-Pav*. b) Hotspot detected (yellow region) within the POV and corresponding user's view (right)

This experience is composed of two levels L1 & L2 (Fig. 2.), where L1 has one node (L1N1) which shows the waiter taking the order and returning with the vada-pav. An immediate hotspot (Fig. 3.) that corresponds to the waiter's position within the immersive environment is activated after the knock. If the user were to gaze on it (Fig. 3. b), the node would transition to L2N2, which shows him walking in and placing the vada-pav near the character. In case the user doesn't gaze at the hotspot (as seen in Fig. 3. a), the experience would transition to L2N1, which shows the waiter leaving with the order. This prototype also has a virtual body on a chair, that rotates to correspond to the user's physical body (as seen in Fig. 3.a & 3.b). This approach could provide a relatively better sense of embodiment while reducing the discomfort caused by placing a fixed virtual body [7].

4.2 Shapeshifter

Shapeshifter is a relatively complex Cinévoqué prototype. The narrative has three levels, seven nodes and four hotspots in total. The story immerses the user into the first-person perspective of a character stuck in an office along with two coworkers during an invasion by a shape-shifting alien race. The experience starts at L1N1 (Fig. 4.) with the user's character and the coworkers watching the news. As the coworkers react to the news, two timer-based hotspots corresponding to each of them are activated. If the user gazes at Coworker-1 the most, the narrative transitions to L2N1 and if they focus on Coworker-2 more, node L2N2 would follow. L2N2 was assigned also as the fallback node, in case none of these two hotspots were gazed at.

Node L2N1 shows both coworkers arguing about leaving the building to seek help. For a small time frame, part of Coworker-2's hand changes colour subtly. During this phase, an immediate hotspot corresponding to Coworker-2 is activated. If the user gazes on this hotspot, the narrative would transition to L3N1 and reveal Coworker-2 as an alien. If the user doesn't gaze at the hotspot, the movie would conclude at N3L2 with Coworker-1 being exposed as an alien.

In node L2N2, a similar argument takes place. At the same time, a spaceship whizzes past the window near the user, thereby activating an immediate hotspot associated with it. If the user gazes at the hotspot, the narrative would transition to L3N3 and conclude the movie by showing an alien entering the room. If the user focuses elsewhere, it concludes at L3N4 by revealing the user's character as the alien.

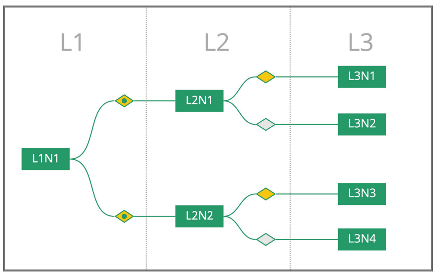

Fig. 4. Narrative Structure of *Shapeshifter*

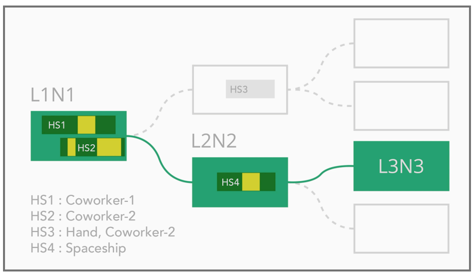

Fig. 5. Example narrative sequence in *Shapeshifter*

Shapeshifter also explored the possibility of varying the genre of the experience across storylines. The tone of Nodes L3N3 and L3N4 is light and if they were to be seen by the user, the interactions between the characters would point towards a sci-fi comedy genre. On the other hand, if the user's experience concludes at L3N1 or L3N2 the genre of the movie would be more of a sci-fi thriller.

The progression of a possible narrative sequence in *Shapeshifter* is shown in Fig. 5. In this case, the rectangular nodes denote the progression of time as we move from left to right. The small rectangles within the nodes show the timeframe within which a hotspot is active. The yellow highlights in the hotspot timeframe indicate the duration for which the user has gazed at the hotspot. In this example sequence, the user has gazed at both hotspots in L1N1, but since the second hotspot was gazed at the most, the narrative transitions to L2N2. At L2N2, as the user has gazed at the immediate hotspot, the narrative concludes at L3N3.

5 Discussion

Shooting live-action immersive narratives that dynamically alter itself imposes new challenges on top of the existing issues that plague production for cinematic VR [14]. In general, care must be taken to ensure that all the objects within the environment aren't altered between successive scenes captured in the same location. Since the user has more visual information that can be observed in an immersive environment, they are more likely to notice any inconsistencies in spatial continuity between consecutive scenes. In a Cinévoqué experience, when multiple nodes across consecutive levels are required to depict the same environment, each node in the succeeding level must start out with an environment that looks exactly the same as that of the preceding node. Achieving such a consistency over all the possibilities becomes increasingly difficult based on the number of succeeding nodes, especially if one of these nodes require changes in existing objects/camera position or if it introduces/removes objects in the surroundings. *Schrödinger's Vada-Pav* and *Shapeshifter* brought these issues to light and provided crucial insights on the ways to address them. These prototypes were also helpful in developing the different storytelling approaches that can be implemented in Cinévoqué and the constraints imposed by each of them.

5.1 Insights from Prototypes

When capturing a narrative for a Cinévoqué experience, the nodes with the same predecessor must be captured in the order of least to most disruptive to the immersive environment i.e. if a particular node requires more changes in the position/orientation of multiple objects or if it adds/removes objects in the environment, it must be captured after its alternative nodes. To further ensure that the environment remains consistent between multiple nodes, some parts of the environment could be replaced with static images or looped videos during post-production similar to approaches in traditional films. It is also advisable to keep the number of objects that are subject to change and in close proximity to the user's POV as low as possible.

In immersive experiences, ambisonic (spatial) audio also helps in directing the user's attention to the necessary event/object [4]. If the nodes were to contain non-ambisonic audio, the audio cues could be spatialized by converting them to an ambisonic format. The transitions between nodes could also be made more seamless by having a global ambient sound that spans across multiple nodes, as the difference in audio would be less noticeable during the transition.

Apart from the storyline, Cinévoqué could also alter other components within the experience, using the data provided by the user before the experience or real-time data. One of the primary uses of this idea is to dynamically recenter the orientation of a new scene or the title and credits. This ensures that the user looks at the intended information within the immersive environment. Another application of this concept is the addition of a virtual body on a chair that rotates with respect to the user's orientation, as shown in *Schrödinger's Vada-Pav* (Fig. 3.a & 3.b).

5.2 Storytelling Approaches

Some of the fundamental approaches of storytelling derived from our prototypes, that could define a narrative structure and the story universe, are as follows:

1. *One story - multiple order*: the narrative structure contains a single node in each level, but the order in which the levels are traversed is determined by the framework. Here different interpretations would be derived by the users based on the order in which the nodes are presented to them.
2. *Unique story - same story-universe*: a user would experience one unique storyline (which is complete in itself) that exists within the same story-universe as other experiences; Each story may or may not be connected to storylines of other users.
3. *Partial story - same story-universe*: based on the POIs gazed at, a user would experience only a part of a larger story within the same story-universe. Each user's experience would be closely connected to experiences of other users, and post-experience discussions among the users will lead to a better closure of the story.
4. *Same premise - parallel story-universes*: although starting with a similar premise each user experiences a unique story, that may exist across parallel universes.

6 Conclusion

Depending on the story content, the branched narrative structures may be diverging, converging or even have circular storylines. One may also create genre-defying narrative structures with the possibilities of each narrative sequence leading to a unique genre of experience. Ultimately it is the storyteller's creativity that will contribute to an effective narrative structure, but these insights could help them create a perfectly seamless experience. For the users, we believe that through the features of Cinévoqué we could maintain coherent storylines more effectively than regular immersive movies, thus giving them further freedom to explore the narrative environment as the story evolves and adapts to them.

References

1. Brillhart, J.: The Language of VR: Concepts and Ideas. https://medium.com/the-language-of-vr, last accessed 2019/04/05
2. Alger, M.: Visual Design Methods for Virtual Reality. http://mikealger.com/professional, last accessed 2019/04/07
3. Pillai, J.S., Ismail, A., Charles, H.P.: Grammar of VR Storytelling: Visual Cues. In Proceedings of the Virtual Reality International Conference-Laval Virtual 2017 (p. 7). ACM. (2017)
4. Bender, S.: Headset attentional synchrony: tracking the gaze of viewers watching narrative virtual reality. Media Practice and Education, pp.1-20. (2018)
5. Brillhart, J.: In the Blink of a Mind - Attention. The Language of VR, https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-attention-1fdff60fa045, last accessed 2019/04/05
6. Brillhart, J.: In the Blink of a Mind - Engagement. The Language of VR, https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-engagement-part-1-eda16ee3c0d8, last accessed 2019/04/05
7. Nielsen, L.T., Møller, M.B., Hartmeyer, S.D., Ljung, T., Nilsson, N.C., Nordahl, R., Serafin, S.: 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 (pp. 229-232). ACM, (2016)
8. Fearghail, C.O., Ozcinar, C., Knorr, S., Smolic, A.: Director's cut-analysis of aspects of interactive storytelling for vr films. In International Conference on Interactive Digital Storytelling (pp. 308-322). Springer, Cham, (2018)
9. Rouse, R., Dionisio, M.: Looking Forward, Looking Back: Interactive Digital Storytelling and Hybrid Art Approaches, pp 93-108, (2018).
10. Vesterby, T., Voss, J.C., Hansen, J.P., Glenstrup, A.J., Hansen, D.W., Rudolph, M.: Gaze-guided viewing of interactive movies. Digital Creativity, 16(04), pp.193-204. (2005).
11. Kirke, A., Williams, D., Miranda, E., Bluglass, A., Whyte, C., Pruthi, R.,Eccleston, A.: Unconsciously interactive Films in a cinema environment—a demonstrative case study. Digital Creativity, 29(2-3), pp.165-181, (2018).
12. Tikka, P., Vuori, R.,Kaipainen, M.: Narrative logic of enactive cinema: Obsession. Digital Creativity, 17(4), pp.205-212, (2006).
13. Rico Garcia, O.D., Tag, B., Ohta, N. and Sugiura, K.: March. Seamless multithread films in virtual reality. In Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (pp. 641-646). ACM, (2017).
14. The Cinematic VR Field Guide, Jaunt Studios, https://www.jauntvr.com/cdn/uploads/jaunt-vr-field-guide.pdf, last accessed 2019/04/07