



Adaptive Playback Control: A Framework for Cinematic VR Creators to Embrace Viewer Interaction

Lingwei Tong^{1*}, Robert W. Lindeman¹ and Holger Regenbrecht²

¹HIT Lab NZ, University of Canterbury, Christchurch, New Zealand, ²Department of Information Science, University of Otago, Dunedin, New Zealand

OPEN ACCESS

Edited by:

Dirk Reiners,
University of Central Florida,
United States

Reviewed by:

Kath Dooley,
Curtin University, Australia
Lei Zhang,
Arkansas Tech University,
United States

*Correspondence:

Lingwei Tong
kris.tong@pg.canterbury.ac.nz

Specialty section:

This article was submitted to
Technologies for VR,
a section of the journal
Frontiers in Virtual Reality

Received: 20 October 2021

Accepted: 09 December 2021

Published: 14 January 2022

Citation:

Tong L, Lindeman RW and
Regenbrecht H (2022) Adaptive
Playback Control: A Framework for
Cinematic VR Creators to Embrace
Viewer Interaction.
Front. Virtual Real. 2:798306.
doi: 10.3389/frvir.2021.798306

Content creators have been trying to produce engaging and enjoyable Cinematic Virtual Reality (CVR) experiences using immersive media such as 360-degree videos. However, a complete and flexible framework, like the filmmaking grammar toolbox for film directors, is missing for creators working on CVR, especially those working on CVR storytelling with viewer interactions. Researchers and creators widely acknowledge that a *viewer-centered story design* and a *viewer's intention to interact* are two intrinsic characteristics of CVR storytelling. In this paper, we stand on that common ground and propose Adaptive Playback Control (APC) as a set of guidelines to assist content creators in making design decisions about the story structure and viewer interaction implementation during production. Instead of looking at everything CVR covers, we set constraints to focus only at cultural heritage oriented content using a guided-tour style. We further choose two vital elements for interactive CVR: the narrative progression (director vs. viewer control) and visibility of viewer interaction (implicit vs. explicit) as the main topics at this stage. We conducted a user study to evaluate four variants by combining these two elements, and measured the levels of engagement, enjoyment, usability, and memory performance. One of our findings is that there were no differences in the objective results. Combining objective data with observations of the participants' behavior we provide guidelines as a starting point for the application of the APC framework. Creators need to choose *if the viewer will have control over narrative progression* and *the visibility of interaction* based on whether the purpose of a piece is to invoke emotional resonance or promote efficient transfer of knowledge. Also, creators need to *consider the viewer's natural tendency to explore* and *provide extra incentives to invoke exploratory behaviors in viewers* when adding interactive elements. We recommend more viewer control for projects aiming at viewer's participation and agency, but more director control for projects focusing on education and training. Explicit (vs. implicit) control will also yield higher levels of engagement and enjoyment if the viewer's uncertainty of interaction consequences can be relieved.

Keywords: cinematic VR, digital storytelling, interaction, 360-degree video, framework, authoring tools, virtual reality

1 INTRODUCTION

As 360-degree videos become widely popular, content creators are now trying to produce engaging narratives using this immersive medium. They aim for contents with complete narrative arc and let the viewers feel immersed in the story world, rather than short and simple footage for brief excitement. We use the prefix “cinematic” to define those narrative-based VR experiences, including story-based dramas, documentaries, or hybrid productions, that feature a beginning, middle, and end. The term Cinematic Virtual Reality (CVR) then emerged and is defined as a type of experience where the viewer watches cinematic content as omnidirectional movies, using a head-mounted display (HMD) or other Virtual Reality (VR) devices (Mateer, 2017). The physical viewing experience may vary from simple 360-degree videos, where the viewer only has the freedom to look around, to a complex computer-generated experience that allows the viewer to walk around, interact with objects and characters in the scene, and even altering the narrative progression (Dooley, 2017). However, we have noticed a lack of a standard production framework as we experienced a series of CVR works, and discovered the ways of constructing a story and implementing viewer interaction varies widely between them, leading to high variability in the viewer experience. In filmmaking, directors rely on a series of cinematic techniques to compose the story, guide viewer attention, and invoke different emotional responses (Bordwell and Thompson, 2013). These cinematic techniques form a framework for staging, camera grammar, cutting, and editing. They work because film directors know that the cinema audience will be sitting in a fixed chair and look directly towards the rectangular screen without any interaction with the onscreen story. In contrast, researchers and practitioners focusing on CVR have, for example, realized the necessity for treating the viewer as a character in the story scene, and that viewers will be expecting a certain amount of agency. They have been producing exploratory projects, gaining experience, and developing design principles (“rules of thumb”) for supporting both the story and viewer interaction for CVR (Pillai and Verma, 2019; Yu, 2019; Pagano et al., 2020). However, a robust framework that is flexible and accessible for non-VR experts to create CVR stories is still lacking.

We are addressing this gap with our prototype framework Adaptive Playback Control (APC) presented here. First, instead of looking at everything CVR could potentially cover, we set constraints to narrow our focus, looking at cultural heritage oriented content using a guided-tour style. This focus allows us to illustrate our approach in a concrete way while also maintaining generalisability. We review previous work on the same theme, which provided frameworks in various media with viewer interaction, trying to find any typical patterns and necessary elements to be included. Then we present the workflow of creating a storytelling experience with viewer interaction by summarizing these works. The construction of the APC framework parallels this multi-step workflow and will produce guidelines at each step, which are story structure guidelines, content preparation guidelines, and content assembly guideline with viewer interaction at its core.

Since there are various story structures and viewer interaction techniques to cover and evaluate, we also set constraints to limit the number of combinations we will look at for now, and extracted two key components essential to design decisions for CVR with viewer interaction: 1) narrative progression (director vs. viewer control) and 2) visibility of interaction (implicit viewer control, vs. a control method with explicit input by the viewer). We report on a formal user study to evaluate four conditions combined from these key components: director control, viewer control, implicit control, and explicit control. We measured viewer engagement and enjoyment towards the content, their general user experience, and their performance on memory tests. A semi-structured interview was also added to collect subjective feedback from the immersive storytelling experience. Although the objective data did not reveal any significant differences between the conditions, the participant behaviors and responses to interview questions helped us deriving several guidelines for creators who are working on CVR projects embedded with viewer interaction. We recommend that creators should choose between increasing the viewer’s feelings of engagement and enjoyment, or enhancing the efficient transfer of knowledge, depending on the purpose of their projects. Also, creators need to consider the viewer’s natural tendency to explore and provide extra incentives to invoke exploratory behaviors in viewers when adding interactive elements. We recommend more viewer control for projects aiming at viewer’s participation and agency, but more director control for projects focusing on education and training. Explicit (vs. implicit) control will also yield higher levels of engagement and enjoyment if the viewer’s uncertainty of interaction consequences can be relieved.

The remainder of this paper is organized as follows: we first introduce related background information to build a framework for creators of CVR experiences. Then we describe how we construct the APC framework and what key elements we have chosen to focus on with higher priority here. After that, we describe the preparation and execution of the user study, followed by the presentation of results, analysis, and discussion. The conclusion and future work are presented in the last section.

2 BACKGROUND

2.1 Interactive Cinematic Virtual Reality

Filmmakers have established practices and guidelines for effective storytelling with movies (Bordwell and Thompson, 2013), such as the “Mise-en-scène” about using the set and arrangement of elements to present a proper perspective towards the story, and cinematography grammars including various shots to compose the picture and direct the viewer’s attention to specific elements. When 360-degree cameras became widely available, practitioners and researchers also explored and came up with guidelines for creators to capture engaging footage and effectively tell stories. Those guidelines include the principles of arranging story elements in the scene, the placement of the camera and characters, and the gestures and body language for human actors to use with the purpose of direct the viewer’s attention in the immersive media without using non-diegetic objects (Pope

et al., 2017; Syrett et al., 2017; Gødde et al., 2018; Bender, 2019; Tong et al., 2020). Later, as 360-degree video became more widely available and easily accessible for the general public, it was no longer a simple medium for short-term excitement. Creators started to treat it as a more serious form of media, and use it to create long and complete stories, aiming to immerse the viewer into the story scene and invoke more intensive emotional resonance with the characters and plot (Bevan et al., 2019; Hassan, 2020). At this stage, the term CVR emerged to define such experiences (Mateer, 2017). Thus, viewers could develop a feeling of “being there” within the scenes and could freely choose the viewing direction (Rothe et al., 2019a). Researchers also started to notice in CVR, compared to traditional flat videos, that both the viewer’s role and expectation of interaction had changed; viewers were no longer passive spectators like in cinemas, but characters inside the story scene, expecting a certain amount of agency within the virtual world (Syrett et al., 2017; Bender, 2019).

Unlike early days where guidelines only focused on directors, new works have moved the viewers into the spotlight (Cavazza and Charles, 2005; Mateas and Stern, 2006; Sharaha and Al Dweik, 2016). Although the discussion is viewer-centered, those researchers were mainly thinking about migrating frameworks from high-level interactive media, such as games, to narrative-oriented media such as 360-degree videos. Because they see the viewer experience has been extensively studied in highly-interactive media and regard these findings as well constructed for CVR. Gradually creators realize a direct migration may not work as both interactive freedom and the mechanisms for interacting with the narrative are different between those two media. On one side, we see creators producing CVR works in a trial-and-error mode, experimenting with prototypes, and gathering design references from filmmaking projects (Ibanez et al., 2003; Brewster, 2017). On the other side, the theories of “co-construction of the story” and “ludo-narrative” have emerged (Verdugo et al., 2011; Koenitz, 2015), highlighting the viewer’s necessary contribution towards the progress of delivering a complete story and providing a circle of experience. In these works, they put forward models of story construction regarding the viewer also as an author of the story (Roth et al., 2018). Nevertheless, they have not designed models specifically for immersive media such as 360-degree videos. We still lack a well-constructed set of guidelines for creators moving from filmmaking or regular 360-degree video to creating immersive storytelling with viewer interaction.

In reviewing this literature, we surveyed previous works that provided frameworks for other types of storytelling, such as tabletop games, video games, and interactive TV programs, trying to find common patterns across them to describe the necessary elements within a framework to support content creation. Carstensdottir et al. (2019) examined interaction design in interactive narrative games, specifically the structure and progression mechanisms, from the perspective of establishing common ground between the designer and player; Ursu et al. (2008) listed a series of existing programs and summarized a system structure for an effective software to

support both authoring and delivery of TV programs. This work has led us to an important conclusion: we believe that to assist in the creation of an effective storytelling experience for CVR, such a framework should provide support for both *how the story structure should be scripted for interaction* and *which mechanisms viewers (or players in this context) will use to interact and move the narrative forward*. The following section describes our proposed framework, Adaptive Playback Control (APC), for interactive storytelling in immersive media. We will start from its motivation to the process we followed to build it, and then the evaluation we conducted.

2.2 The APC Framework

The motivation for creating the APC is to provide a framework for making CVR (e.g., 360-degree video, computer-generated immersive movies) a *more* interactive experience, while not pushing interactivity to the point where it becomes more of a gaming experience. Since consuming film is more of a “lean-back” experience for many, striking this balance is a key to its appeal.

We expect that the APC will give creators a familiarity akin to scripting for conventional videos, and provide them with the confidence of authorial control, combined with satisfactory outcomes throughout production, delivery and later consumption. Thus, resulting works will on the one hand have a pre-scripted narrative as their backbone, but an interactive and immersive experience at the front face. This will ensure the narrative arc remains under the control of the director, while the freedom of interaction is placed in the viewers’ hands. However, there are various common storytelling methods for different user scenarios, defined by factors including the content, the emotional consequences that the director wants to invoke in the audience, and the type of interaction viewers are going to use (Tong et al., 2021). In order to ensure the exploration will be effective and can yield practical results for APC users (creators and consumers), we set the first constraint, that in this study *applicable contents* we will apply the APC onto are those which:

- (1) are cultural heritage oriented;
- (2) use a guided-tour style, meaning there will always be an embodied host in the scene, visible to the viewers, whether it is an actual human or synthetic character; and
- (3) use content that is prerecorded.

Within this realm, we review existing storytelling frameworks for various media, trying to locate a common pattern. We noticed a trend that viewer participation is regarded as a key component if immersion and story comprehension is the aim of the entire experience, such as those ones (Mulholland et al., 2013; Habgood et al., 2018; Lyk et al., 2020). Although covering different topics, they all used immersive content as a base and constructed the storytelling experience on top of that, regarding the viewer as a part of the story, emphasizing viewer’s interaction (gazing, gesture, selection, etc.) as a tool to drive the narrative forward. We can extract similarities from those examples, that they usually start by laying out the non-linear structure, prepare the content for each node of the structure, then implement the system, filling

in the content and adding viewer interactions. Since those works have been proved effective and others have been using them as references for design (Ferguson et al., 2020), we decide to construct the APC framework in a pattern matching this workflow, and divide it into three parts.

- **Structure Guidelines** for designing an appropriate non-linear structure for the story for interactive CVR;
- **Content Creation Guidelines** for content preparation for immersive media (mainly 360-degree videos);
- **Assembly Guidelines** for assembling content by combining the structure with viewer-interaction design.

We expect that content creators will use this APC framework as a reference for their production process, choosing the story structure to support their purpose with storytelling, preparing content segments, and assembling them, with interactive elements designed for viewers, into an engaging and satisfying immersive storytelling experience with interactivity.

2.3 Foundation of the Framework

To build a prototype of the APC framework for evaluation, we constructed each of its components in different ways. For the Structure Guidelines, we looked at structures summarized by researchers from other immersive media, such as video games and VR games, and migrated them to narrative CVRs with proper modifications. The Content Creation Guidelines were distilled from previous projects and assembled them from verified film techniques, including the *Mise-en-scène*, camera manipulation, and guidance techniques such as the Action Units (Tong et al., 2020), to cover the preparation and capturing stage. For Assembly Guidelines and interaction design, we evaluate and choose interactive methods for each of the structures and see which one fits. Since the guidelines cover a series of non-linear structures and viewer progressing mechanisms, the abundance of combined variation means evaluation work will be time and resource intensive. We do not cover all possibilities in one study, instead we filter out impractical combinations and focus on those are closely related to the content type we mentioned in last section, mainly for reasons of required sample size and evaluation effort. First instead of cover all structure types, we focus on one commonly used non-linear form, known as “hub-and-spoke,” which is the base of other derived types (Carstensdottir et al., 2019) and has been widely used in both tours (Mulholland et al., 2013; Sharples et al., 2013) and games (Moser and Fang, 2015).

Then we acknowledge that involving viewers in the control of narrative progression is an outstanding topic in interactive storytelling research (Verdugo et al., 2011) on the one hand, moving from traditional screen-based media to immersive media, a shift in the viewer’s perspective naturally calls for giving the viewer the freedom of choice and the agency of impacting narrative progression (Tong et al., 2021). On the other hand, researchers have also discovered that for culture- or education-related content, a clear structure helps in audience understanding and retention (Lorch et al., 1993). Researchers who focus on games have also found that, in particular, explicit narrative

progressions designated by the creators have a positive effect on declarative knowledge acquisition (Gustafsson et al., 2010).

Viewer awareness of interaction also has an impact on the user experience of storytelling (Rezk and Haahr, 2020). We know that if the viewer is regarded as a character in the story scene in immersive storytelling, they expect a certain level of interaction to be involved in narrative progression (Tong et al., 2021). A system designed to be responsive to viewer input will also increase the level of viewer involvement and immersion (Ryan, 2008; Roth et al., 2018). Concern has also been raised by some researchers, pointing towards the design around consequences of choices. Rezk and Haahr, (2020) cautioned that if viewers are given explicit choices during storytelling, they are likely to hesitate when faced with too many options, as they will evaluate every potential consequence of each option, therefore be unable to make confident choices, thus shattering the feeling of “being there” in the story world. Realizing this, some creators turned to a new design style known as “invisible control,” where the viewer still participates in the progression of the narrative, but is not *explicitly aware* of making choices. To achieve so, one implemented a system monitoring viewer’s behavior during virtual museum tours at predefined spots, and respond to it by making unannounced narrative choices over branches (Ibanez et al., 2003), another recorded the player’s actions in the game and used the data to determine his/her overall contextual intention, then presented a matching ending from several parallels (Sengun, 2013).

Thus we narrow down the focus of this study to a two-element combination, the narrative progression (director vs. viewer control), and the visibility of interaction affordances (implicit vs. explicit interaction). We also explicitly prioritize them in the APC framework.

3 METHODS

The main focus of this study is set on how viewer interaction can be enabled in interactive cinematic virtual reality, such as 360-degree videos. The structure of the story we employ has a pattern of hub-and-spoke, i.e., the viewer starts from a central location (the hub) and all alternatives (the spokes) start from here. Since the prerecorded content of the story will not change, and essentially every participant will watch the same content, we conducted a between-subjects experiment to compare the user experiences between design variants. The content we used for this study was an in-house guided tour through a series of 360-degree videos. We will describe the material and production process in detail in **section 3.3**.

In the experiment, we set up four conditions of the interactive and immersive storytelling experience.

- (1) Director control (DC);
- (2) Director control, but with randomly rearranged segments (DR);
- (3) Viewer control, but with implicit input (VI); and
- (4) Viewer control, but with explicit input (VE).

The structure of the content remained unchanged between conditions and will be illustrated in detail in **section 3.3**. However, the control models and viewer interaction methods varied. The exact choices of parameters, including the control of temporal sequence and the type of involvement, are listed in **section 3.4**. We set up the condition DR to observe, in this specific story, if the viewer's experience will change when the order of the segments was different from what the director initially intended. We measured each condition's effect on the viewer's level of engagement with the content, enjoyment from the experience, and the general user experience. We also designed a series of content-related questions, both directly asking about one of the elements from the scene and conclusions derived from what the host introduced, combined with information visible in the scene. Those content-related questions were used to assess how the conditions affect the viewer's memory performance and the system's performance on the transfer of knowledge.

3.1 Research Questions and Hypotheses

In this study, the research questions we would like to ask are:

- **RQ₁**: For the hub-and-spoke structure, which type of control pattern will bring a higher level of engagement and enjoyment, director control or viewer control?
- **RQ₂**: When viewers have control over the order of the segments in a hub-and-spoke structure, which one will yield a better usability experience, implicit involvement or explicit control?
- **RQ₃**: Between implicit involvement and explicit control, which one will lead to better memorization of the content?

We formulated hypotheses corresponding to the research questions based on previous research related to the viewer's role and behavior in narrative CVRs. Firstly, we expected the viewer-controlled modes (VI and VE) will bring higher levels of engagement and enjoyment, as in immersive environments, agency increases the level of presence and brings a deeper feeling of being directly involved in the story scene, as well as a stronger feelings of fun (Ferguson et al., 2020). As has already been verified by previous research, the viewer's role in immersive media is different from the one in a regular movie. The viewer feels they are a character in the scene and form the expectation that they have some influence over how the guided tour will progress (Ryan, 2008; Roth et al., 2018). Secondly, as mentioned in the previous section, explicit choices for the user will make them think about the potential consequence of the action of choice, thus breaking immersion and deteriorating the general experience (Rezk and Haahr, 2020). We also assumed the invisible control method (VI) will impact less on the viewer's general experience towards the system, because explicitly making selections adds extra workload during the experience. Thinking over the choices will also distract viewers from focusing on the narrative from the host. Thirdly, we expected the condition VI will also lead to a better result in the viewer's memory test of the content, because in the viewer's perception, they are *passively* watching and more focused on the content. In summary, we formulated the following hypothesis:

- **H₁**: compared to director control, viewer control will lead to higher levels of engagement and enjoyment with the experience;
- **H₂**: compared to the explicit control-based method, the invisible involvement will lead to a more positive user experience;
- **H₃**: the invisible involvement will lead to better performance on memory tests.

3.2 Apparatus

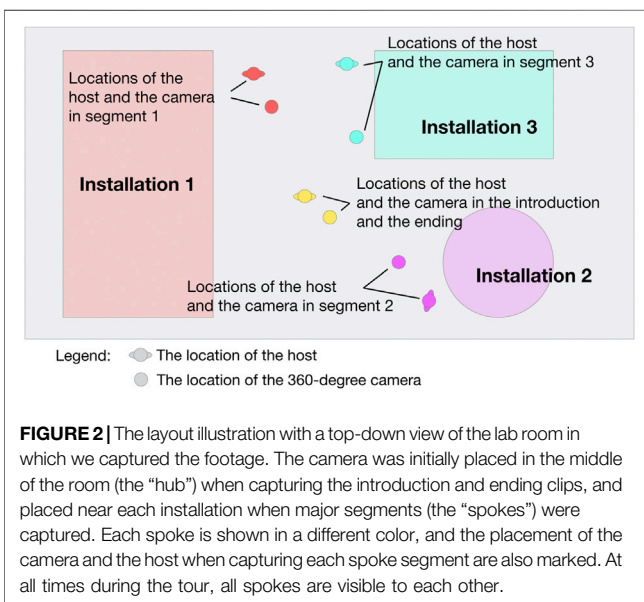
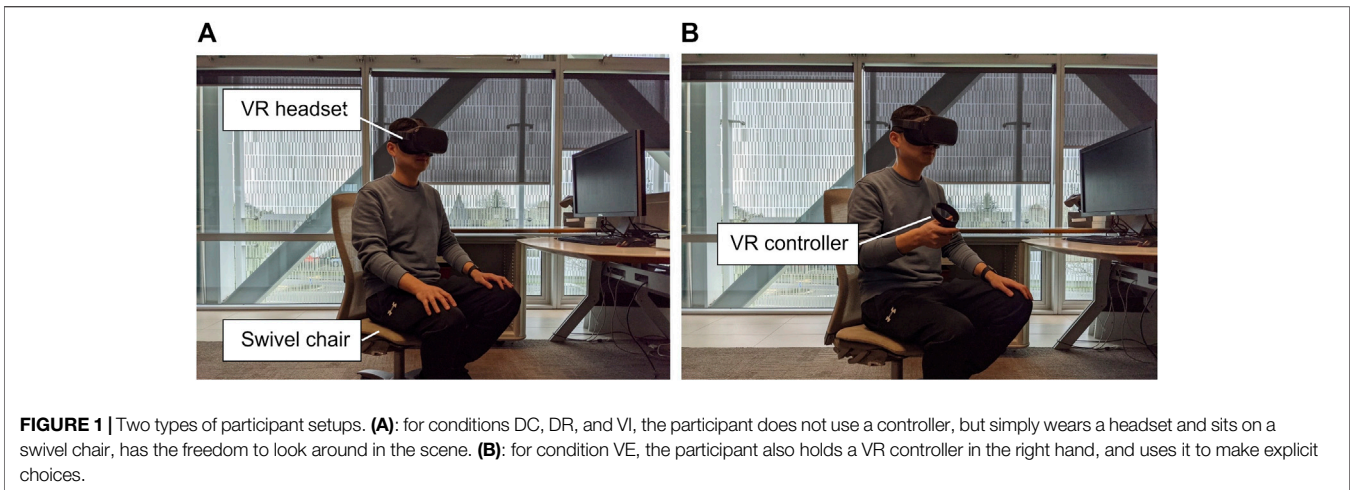
We used a computer running 64-bit Windows 10 Professional with a 3.2 GHz i7 processor and a GeForce RTX 2080 graphics card, to implement the APC system, record viewer behavior data, and ensure the smooth playback of the 5.7 k videos. During the experiment, participants viewed the 360-degree videos wearing an Oculus Quest 1 HMD with or without using its controllers, depending on the conditions, as shown in **Figures 1A,B**.

3.3 Materials

For this experiment, we captured a series of 360-degree videos in a laboratory room we selected for the content, then assembled them into a virtual guided tour. We first scouted the location and designed the story. In the space we chose, several large experiment installations are placed against each wall, as shown in the top-down view of the space in **Figure 2**. In the real world, to a newcomer to this space, a host would first introduce the purpose and daily activity in that room, then move on to each of the installations. We duplicated such a visit in this study by capturing 360-degree videos of it, as the spatial layout of the *space* and the *installations* matches the hub-and-spoke story structure. It is also noteworthy that in this space the viewer could see all the spokes as they were all "open" and "equivalent", imposing no hierarchical relationship between them.

We then captured the footage when the host was introducing the installations at each designated spot, using an Insta360 ONE R 360-degree camera at the resolution of 5.7 k (5760 × 2880). The camera was mounted on a tripod so the viewpoint was fixed in each of the segment. The positions of the camera and storyteller (host) for each segment were carefully chosen so they are not blocking views, shown in **Figure 2**. The distance from the host to the camera was kept the same (2 m) at each location, so the viewer would always feel as having a similar space between themselves and the host when watching. A wireless mic was also used and directly plugged into the camera, so the audio quality was maintained at the same satisfactory level no matter the distance between the camera and the host.

Following the script, we first captured the Introduction clip at the hub where the host gave an overview of the lab room and introduced all the installations covered in the tour. Then at each spoke, the host talked about the experiment installation, including its features, applications, and research projects running on them. The narrative structures of each segment at the spokes were also scripted to be similar in terms of running time. We captured a total of five clips, including one Introduction clip at the hub, three major segments at each spoke and one Ending clip, which contained no key information, but



only wrapped up the tour experience, rather than having an abrupt cut at the end. The clips were then processed using Adobe Premiere Pro to add fading to black transitions at the beginning and end of each clip, and to adjust the volume levels and narration pace to minimize differences between the recordings.

3.4 Implementation of the APC System

In this study, the APC system consisted of two components: 1) a video player to present via the VR headset to the participants the 360-degree video clips we captured, and 2) a mediator component to deliver the designated control from the director (stored in advance) or to respond to viewer interaction during playback. Both components were implemented using Unity3D 2020.3.13f1. The implementation details of each condition are described below.

DC: in this condition, the five clips were loaded into the library of APC following a specific order (the Introduction, three major segments, then the Ending). The APC played those 360-degree video clips one after another. There was no viewer input in this condition.

DR: similar to DC, the APC played those 360-degree videos without viewer input involved. The only difference was that the system randomly rearranged the three major segments every time the experimenter initiated the experience for a new participant. The Introduction and the Ending were permanently fixed at the beginning and the end.

VI: in this condition, the viewer’s head orientation was monitored and recorded in real time by the moderator component as she was watching the 360-degree **Supplementary Videos**. A series of invisible gates were set up in the scene, overlapping with the area of subject-matter objects from the perspective of the viewer in the center of the spherical scene, as shown in **Figure 3**. Thus, when the viewer looked at one of the objects, her head orientation fell within the corresponding gate. The moderator then recorded and calculated how long the viewer had been dwelling on the object shown behind the gates. If the viewer dwells on a gate longer than a threshold (set to 3 s in the system we used for the study), the moderator will decide that the viewer might be interested in this object and pulls the corresponding segment to the top of the playlist as the next one to play after the current clip. The gate threshold can be triggered multiple times along the playback experience until the current clip plays 95% of the way through. The gates were positioned manually in each segment scene to match the viewer’s head location in that scene. The entire process was executed by the moderator in the background, and the viewers were not aware of it at any moment.

VE: we implemented a hand-held laser-pointer method in this condition to enable the viewer’s explicit interaction. It used a simple point-and-activate mechanisms (Rothe et al., 2019b) and helped reducing the unnecessary learning workload of the viewers. Based on the gate design from VI, we replaced those invisible gates with visible cards, as shown in the screenshot in **Figure 4**. Each card had text showing the segment it represents,

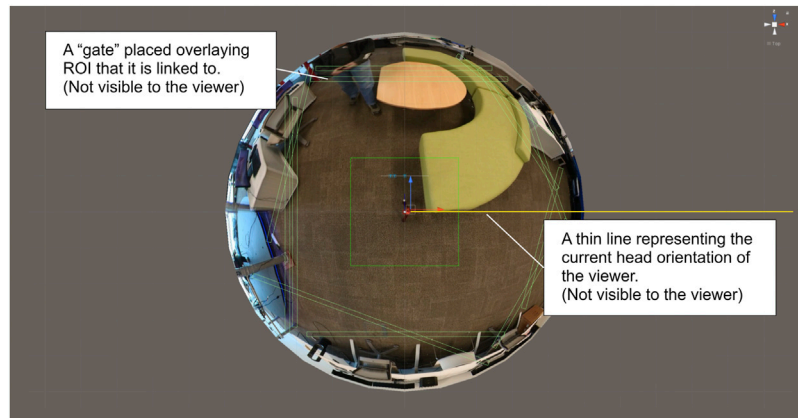


FIGURE 3 | A top-down view of the 360-degree spherical screen showing the gates scattered around covering the Regions of Interest (ROIs) behind them, waiting to detect the viewer's head orientation vector (which is represented as a thin yellow line in this screenshot) to collide with them, registering gaze dwelling over the ROIs. The image is a top-down view so the gates are the green rectangles. In the 3D scene they are actually thin boards standing in front of the ROIs, and serve only as detection mechanisms and are not visible to the viewers.

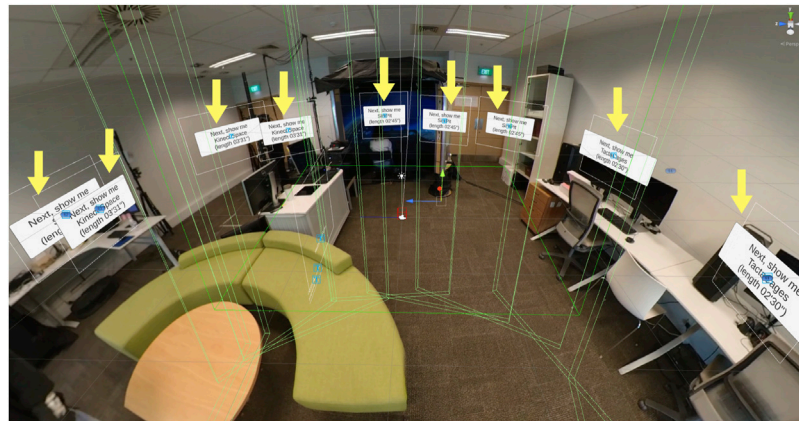


FIGURE 4 | A screen shot from the Unity editor showing the cards distributed around the scene. They are interactable with the virtual controller in the hand of the viewer. In this figure, all of the cards are visible, and pointed out with yellow arrows for illustration purposes only. In the actual scene the yellow arrows are not visible. Also the cards will not show in an unanimous fashion like in this figure. They are programmed to show at certain locations and times according to the progress of the current segment.

plus the time length of that video segment as an extra aid of the viewer's decision-making process. Viewers who were holding the controller can then point at the cards with the laser pointer and pull the trigger to make a selection, as shown in **Figure 5B**. A popup message would temporarily appear in the viewer's Field of View (FOV) to confirm the choice. We set the cards to appear when the progress of the current video clip reaches 70%, so they were not always shown to cause distractions, nor appearing too late to leave the viewers with a very short time window to make selections. Since the interactable cards overlaid on top of topic-related objects and were scattered around the scene, popup messages were programmed to appear in the center of the viewer's FOV to remind her when cards were available for selection, reducing the fear of missing out. If the viewer does not make any choices

before the video segment ends, the default order from DC will be used.

In all four conditions, a Helper Head-Up Display (HUD) was set up and constantly visible at the lower left corner of the viewer's FOV, as shown in **Figure 5**. It aided the viewer in being aware of the progress of the current segment and the progress of the entire tour. Thus the viewers would not lost in the tour and experience difficulty recalling the content of a segment. Instead, they maintained an awareness of the progress and pace.

3.5 Measures

When designing the experiment, we noticed two important facts. First, in the VI condition, the viewer was not aware that the system was monitoring her head orientation and making alternations to the playback progress. Second, since a

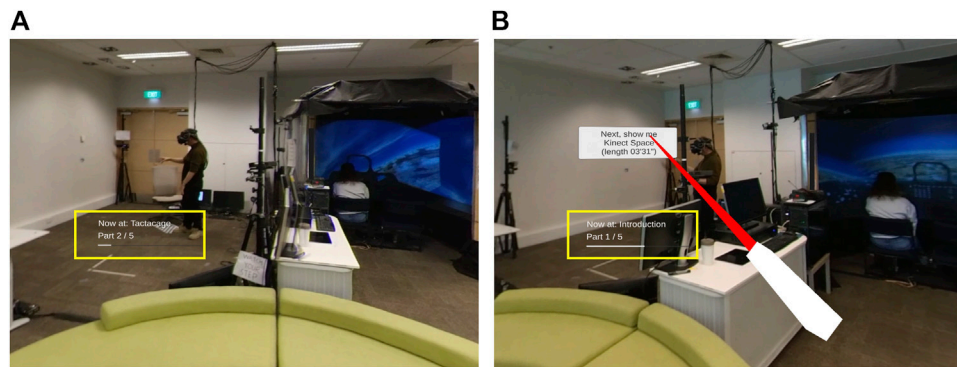


FIGURE 5 | Screenshots of what participants see in four conditions. **(A):** For conditions DC, DR, and VI, the participants simply sit and watch the 360-degree videos. The Helper HUD is always visible in the scene and fixed to the lower left corner of the view. It is highlighted with the yellow box. The yellow box is not visible in the experiment. **(B):** For condition VE, the participant will also see a virtual controller with a red laser pointer in her right hand. There are also interactive cards with texts visible in the scene at certain times (like the one in the screenshot). The participant then uses the virtual controller to point and make explicit choices.

participant only experienced one of the four conditions, across DC, DR, and VI, participants were also not aware that some conditions might impact the narrative progression, while others will not. Taken together, we realized that the viewer's subjective experience across the conditions would not be equivalent unless they are told afterwards which one they experience; we concluded that we cannot assess and compare the user experience by explicitly asking about their preferences. Instead, we measured their subjective feelings by regarding all conditions as generic immersive storytelling experiences. Then we derived design reference by comparing the participant's overall experience across the conditions.

For subjective measures, we assessed each participant's level of engagement with the content, enjoyment of the experience, and general usability experience. The engagement was measured by the widely-applied User Engagement Survey Short Form (UES-SF) (O'Brien et al., 2018). We also used the operational guidelines provided by Schmitz et al. when applying this measurement (Schmitz et al., 2020). The enjoyment was measured using the 12-item scale provided by Ip et al. (2019). They used this scale to measure learner enjoyment after watching a series of Massive Open Online Courses (MOOCs) in the form of 360-degree videos, similar to our study setup. A simple three-part evaluation form measured the general user experience towards the system, provided by Shah et al. (2020) in their study evaluating a 360-degree video playback system. We removed one non-relevant item, then applied it to see if the participants find it easy to use the system, both the invisible interaction and the traditional point-and-activate method.

Content-related questions were used to assess participants' memory performance from each condition. Three multi-choice questions required the participant to combine and extract several elements, either directly from the host's speech or visually from the scene, all around one specific topic. Three single-choice questions then asked the participant to verify some facts visible in the scene. The seventh question asked the participant to make a design choice for a "simulated situation" based on the higher-level overview given by the host from the

entire virtual tour. These questions provided insights into how much the participant remembered from the tour and understood from the introductions. The questions did not require any reasoning or deduction, and so did not require any background knowledge.

We also added a plugin to the APC system to automatically record the viewer's head orientation in real time and to generate a heat map of the viewer's attention and dwelling among the entire scene, accumulated for each segment. We then used the heat maps to identify whether the viewer showed a relatively active exploratory behavior or mostly passive, static watching.

3.6 Experiment Procedure

We compared the four conditions (DC, DR, VI, and VE) by applying them to the same five-segment virtual guided tour we captured. For this study, we recruited 44 participants (22 females, 21 males, and 1 who chose not to specify) from the university. All were between 19 and 39 years old ($M = 26.68$, $SD = 4.978$). They self-reported having different levels of VR experience and 360-degree video experience. Among the 44 participants, 28 had never, or only a few times in the past year, tried a VR experience, 14 had used VR at least once per month, and two reported using VR on a daily basis. A total of 35 out of 44 reported never having watched a 360-degree video before, and nine had watched a few times in the past few months.

Before the session started, we obtained consent from each participant. They were then introduced to the "virtual guided tour" and explained how he/she would experience it with the HMD and the Swivel-Chair VR setup (Tong et al., 2020). Then, a sample image of the Helper HUD was given to the participant for him/her to understand its purpose before the tour started. If the participant was assigned to the VE condition, a brief introduction of how to use the controller and interact with the selection cards was also added. The participant used only the right-hand controller in this user study. When the session started, each participant went through five 360-degree video segments. The first and last segments were always fixed, while the three in the middle varied by condition and each viewer's interaction or

TABLE 1 | The mean values of the results for Engagement, Enjoyment, General user experience and Memory performance, for each condition.

Measure	DC	DR	VI	VE
Engagement	15.97	16.30	16.52	15.58
Enjoyment	10.79	11.59	11.80	11.89
General user experience	11.52	12.39	12.95	12.32
Memory	17.27	16.23	17.14	16.41

The highest values for each measure are highlighted in bold.

behavior during the experiment. When the tour ended, we asked the participant to remove the HMD and move on to complete the questionnaires.

After the questionnaire, we conducted a semi-structured interview using the following prompts:

- (1) Please give a subjective description of the entire tour.
- (2) Please talk about the most impressive part of the tour.

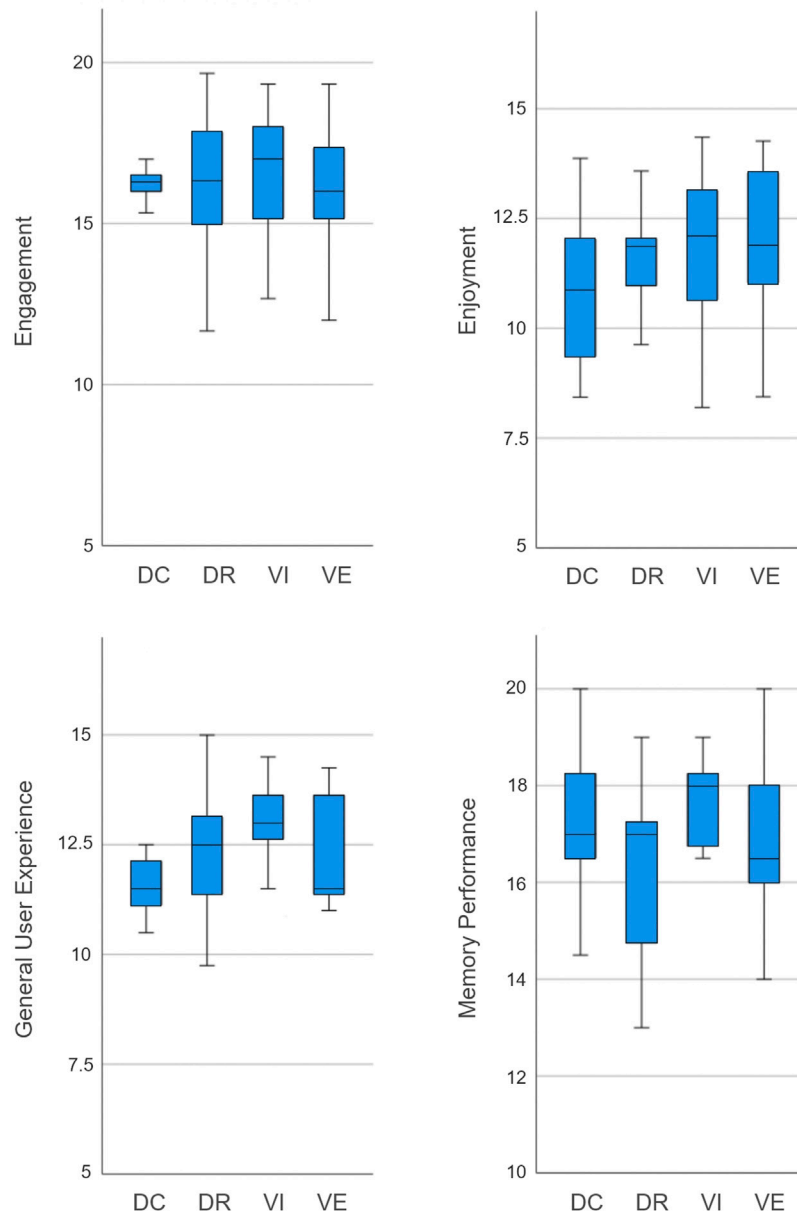


FIGURE 6 | Boxplots summarizing the results of levels of engagement (converted from the UES-SF items), levels of enjoyment, general user experience scores towards the system, and the memory performance scores from the content-related questions, for each condition, in terms of medians, inter-quartile ranges, minimum and maximum ratings. Top row, from left to right: Engagement, Enjoyment. Bottom row: General user experience and Memory performance.

- (3) Please talk about your preference between using a VR headset or a regular TV for this type of content in the future.
- (4) Do you have any other comments or thoughts?

For VE, we added a question to ask about their motivations for making choices using the controller via the interactive cards. The entire session took approximately 35 min for each participant.

4 RESULTS

We analyzed the data participants' reported in questionnaires by using a one-way ANOVA for fixed effects ($\alpha < 0.05$), with Bonferroni post hoc and Tukey post hoc comparisons. The mean values for levels of engagement (converted from the UES-SF items), levels of enjoyment, general user experience scores towards the system, and the memory performance scores from the content-related questions, for each condition, are shown in **Table 1**. The highest values for each measure are also highlighted. The overall results are summarized and plotted in **Figure 6**. The calculations of levels of engagement, enjoyment, and general user experience scores were conducted following the operational guidelines provided by the original creators of the instruments. For the content-related questions, since there are both multiple-choice and single-choice questions, we calculated the score using the following rules: for a single-choice question, the participant scored one point if she chose the correct answer; otherwise, she scored zero points. For a multiple-choice question, the participant scored a full mark (five points) if she selected all the correct items and only those correct items, lost 0.5 points if she missed one of the correct items, and lost another point if one of the wrong items is selected.

4.1 Level of Engagement with the Content

The levels of engagement with the content were higher on average from the participants who experienced the VI condition ($M = 16.52$, $SD = 0.516$), compared to the other three (DC: $M = 15.97$, $SD = 0.327$, DR: $M = 16$, $SD = 0.608$, VE: $M = 15.58$, $SD = 0.777$). However, the ANOVA test indicated no significant differences among the conditions ($F = 0.344$, $p = 0.794$).

4.2 Enjoyment of the Experience

Participants who experienced the VE condition reported the highest levels of enjoyment of the experience (VE: $M = 11.89$, $SD = 1.684$), which is slightly higher than the levels from participants of the other three conditions (DC: $M = 10.78$, $SD = 1.813$, DR: $M = 11.59$, $SD = 1.211$, VI: $M = 11.79$, $SD = 1.774$). However, the ANOVA test indicated no significant differences among the conditions ($F = 0.973$, $p = 0.415$).

4.3 General User Experience

The levels of user experience with the system indicate that participants who experienced the VI condition showed more positive user experience with the system (VI: $M = 12.95$, $SD = 1.145$), compared with the other three conditions (DC: $M = 11.52$, $SD = 1.207$, DR: $M = 12.38$, $SD = 1.514$, VE: $M = 12.31$, $SD = 1.260$). However, the ANOVA test indicated no

significant differences among the conditions ($F = 2.269$; $p = 0.092$).

4.4 Memory Performance

We looked at the memory performance scores from participants of the four conditions and noticed that those who experienced the DC condition scored higher in the memory test (DC: $M = 17.27$, $SD = 1.664$), compared to the other three conditions (DR: $M = 16.22$, $SD = 2.029$, VI: $M = 17.13$, $SD = 1.818$, VE: $M = 16.41$, $SD = 2.791$). However, the ANOVA test indicated no significant differences among the conditions ($F = 0.662$; $p = 0.580$).

4.5 Behaviors of Attention and Focus

We also recorded each participant's attention over the entire scene by accumulating the dwelling time over the 360-degree sphere as a canvas using a plugin installed in Unity. The final results were generated as heat maps. A total of 176 maps were collected, grouped into four sections, corresponding to the four conditions, shown in **Figure 7**.

Comparing across the conditions, we notice that in the heat maps from the VE group, the scanned area recorded on the canvas is larger than those of the other groups, indicating that participants who experienced the VE condition showed a higher tendency towards exploration (actively looking around).

Comparing across the segments under the same condition, we also notice that participants showed more exploratory behavior in the first segment (Introduction) than the other three main segments where the host introduced details of each installation. We will discuss the participant behaviors in **section 5.2.2** by looking into the heat maps and participant answers during the interviews.

4.6 Subjective Feedback from the Interviews

We transcribed and analyzed the interview results using a thematic approach. Topics were identified and generalized into two high-level themes, with five sub-level topics. Below we present the themes from the interviews alongside our interpretation. The further analysis and discussion by connecting the themes with the conditions participants experienced are presented in **section 5.2.1**.

4.6.1 Theme 1: The Impression of the Virtual Tour Experience

The first theme is about the participants' impression of the virtual tour as a general experience. We learned the viewers were mainly paying attention to either the experience itself, the content presented, or the system, during the virtual tour. Generally speaking, being immersive helped participants' engagement with the experience.

4.6.1.1 Being Immersed in the Virtual Environment

Many participants showed an emphasis on the general user experience itself and the novelty of the 360-degree videos, instead of the content (or story) of the tour, such as: "It feels

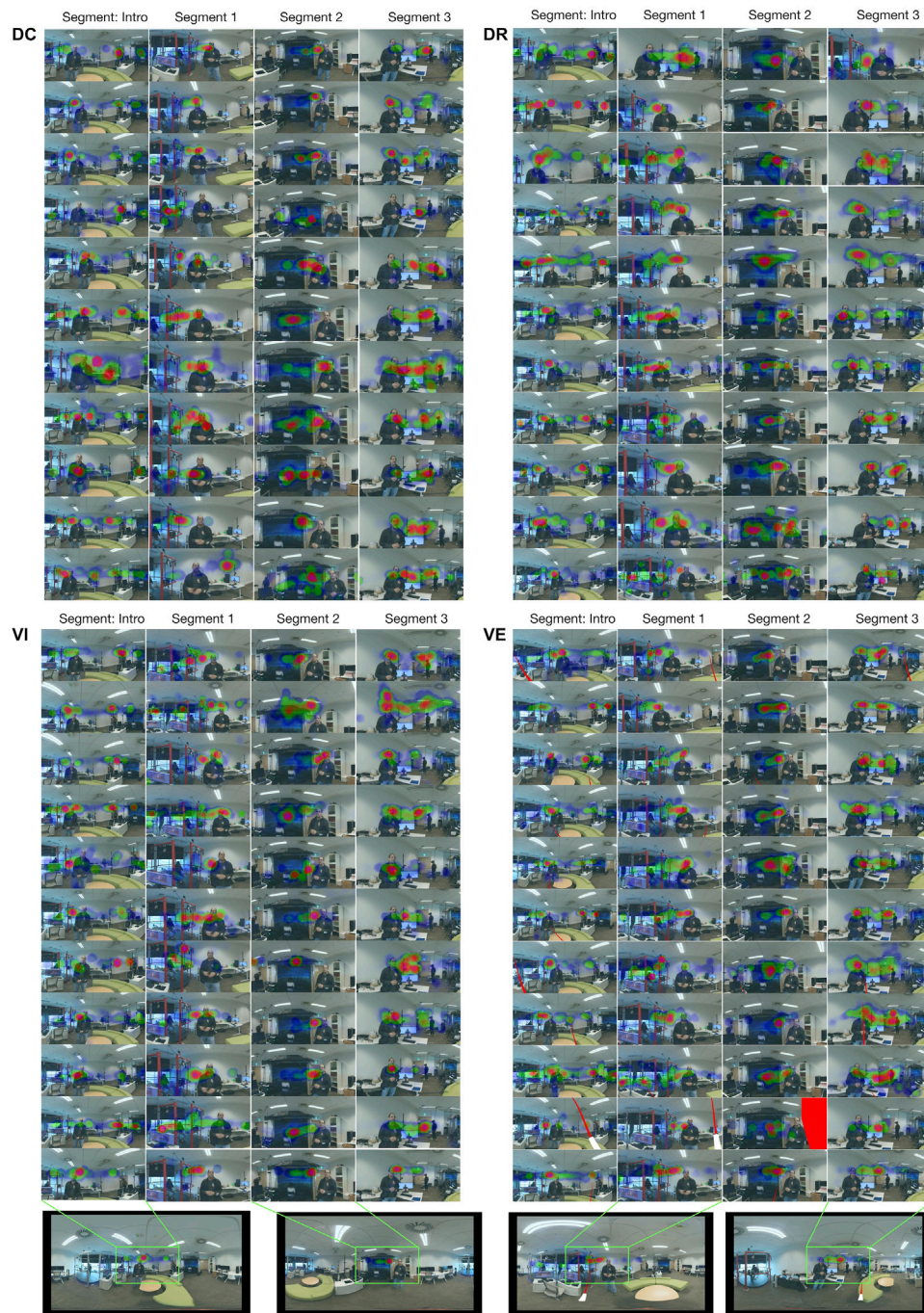


FIGURE 7 | A grid summarizing the generated heat maps, grouping by the conditions. The images are zoomed into the central part to clearly show the dwelling of gaze, from the original equirectangular projections. Four samples are shown at the bottom. In each sector, there are 4×11 (or 44) heat maps collected from one condition, 4 segments. In each row, there are four heat maps representing one participant's behaviors. Each column represents the data from all 11 participants from this condition, on this specific segment. In DR, VI, and VE, although participants watched segment 1, 2, and 3 in random orders, the images here are arranged in DC order for easy comparison.

like a 360-degree experience, an immersive tour ... I can look around". The characteristic being immersive is also preferred by participants: "I feel like as if being transported to the space and being there with the tour guide ... I will prefer to use it."

4.6.1.2 Feel Like Having a Real Guided Tour

Instead of regarding it as 360-degree videos, other participants directly described it as a guided tour, emphasizing its content by mentioning the name of the space or specific installations they

virtually visited (whether using the correct name or their own words): “It was a guided tour of the [name of the space] and I was introduced to three things, [names of the installations]. . .” We also noticed that participants recalled the host’s name and title: “It was a guided tour given by [name]” All of their heat maps show their attention mainly fell on the host. Participants also directly listed the names of the installations and their applications as their impression of the tour: “It was a tour about three installations . . .”, indicating they were mainly attracted by the installations during the tour.

4.6.1.3 Impression of the Technology

Participants also talked about the VR system itself. They pointed out the usability of the headset, comfort, and how it blocks the peripheral view to help being immersed in the virtual scene. Those are also related to the other theme we will discuss later. But a few also pointed out their expectation of design improvement: “I do not want to use the VR headset frequently . . . not or for long time use because it is uncomfortable . . . I was unable to walk around in it”.

4.6.2 Theme 2: As a Medium for Transfer of Knowledge

The previous theme explained how the participants saw the virtual tour and the impression they had. With this theme, we look at if the virtual tour helped the participants as an immersive tool to assist learning about the lab and the installation presented in the materials.

4.6.2.1 Actual Knowledge Gained

Participants did reply with the essential information presented in the tour, such as “I felt like learning about the lab is very interesting”, or one specific item from the tour that the participant personally felt interested with, such as “The [name of the installation] is quite impressive because I think it is interesting/cool . . .”. Since the participants are providing actual information gained from the virtual tour, instead of seeing it as a plain 360-degree video experience, the immersive experience did help the participants to learn about the lab and the installations from the host’s presentation.

4.6.2.2 Immersive as an Advantage for Learning

Instead of specific information, some participants put their impression mark on the system itself, including its intrinsic characteristic that the viewer is immersed in the virtual scene and has the freedom to look around or choose what to watch next, reporting it as an advantage, when compared with other forms of learning: “I felt I was in the lab all the time and I did not get distracted by other things . . . Seeing myself actually standing in the lab is better than seeing it through a screen or as a video.”

5 DISCUSSION

In this section we will discuss the main findings and implications from the results of the experiments.

5.1 Objective Measures

Since there were no significant differences detected between the objective results from the four conditions, the hypotheses (H_1 and H_2) were not supported; we do not find higher levels of engagement and enjoyment in those who experienced the conditions with viewer control (VI and VE). Data analysis detected no significant differences among the memory performance scores from four conditions, so H_3 is also not supported. These indicate that if we look at the objective responses, participant did not react to the four conditions differently. Alternatively, they did not objectively become aware or care about all four conditions applied to the virtual guided tour experience.

We assume these findings are due to three factors. The first one is that participants were unaware of any alternative other than director control, and so assumed that was in play. Since they were not told about the different mechanisms running in the background, nor did they know the original arrangement of the segments, in their eyes, the experience was just a series of 360-degree videos. Even for the VE condition, participants regarded controller input as a separate task to perform (and several of them did not even use it), while the virtual tour itself was still treated as “a series of 360-degree videos I need to watch.” This assumption is also supported by the subjective observations we are going to discuss in the next section.

The second factor is related to the content. As stated in previously, we set constraints in advance and focused only on one type of story structure, “hub-and-spoke.” Also, none of the experimenters involved in this work was a professional filmmaker or scriptwriter, so narrative intensity and creativity were limited in the content we made. We expect participants’ feelings of enjoyment and engagement would have been higher if the content was well-prepared and creatively made by professionals.

The third factor is the amount and granularity of viewer interaction. In conditions VI and VE, viewer interaction was only used to drive the narrative progression. More specifically, among various elements within a storytelling experience, what we allowed the viewers to control was “which segment will I go to next after this one” instead of any specific visual elements or action choices with consequences in the scene. Compared to this virtual tour, we have seen in other interactive, immersive storytelling experiences, that viewers can interact in a much richer manner than simply picking segments on a playlist. One might interact with an object within the scene or even interfere with a character’s behavior. It is possible that, compared to only narrative progression, viewers might show noticeable changes in their feelings and memory performance if they are able to interact with more specific and visible elements in the scene (and maybe also linked to narrative progression/branching).

5.2 Subjective Feedback and Observations

In this part, we will discuss several topics related to viewer behaviors that we collected from three sources: 1) the answers from the interviews with the participants, 2) observations of participant behaviors while they were watching the 360-degree

TABLE 2 | The number of participants' responses grouped by themes and sub-level topics, with their distribution in the different conditions experienced.

Themes and sub-level topics	Total	DC and DR	VI and VE
Theme 1: The impression of the virtual tour experience			
Being immersed in the virtual environment	31	13	18
Feel like having a real guided tour	18	13	6
Impression of the technology	37	/	/
Theme 2: As a medium for transfer of knowledge			
Actual knowledge gained	24	14	10
Immersive as an advantage for learning	20	8	12

videos, and 3) the heat maps showing the distribution of viewer attention across each segment of the tour.

5.2.1 Viewer's Main Impression and the Role in Control

In the previous section, we summarized the answers from the interviews and analyzed them for themes and sub-level topics. We matched the transcription of each participant's responses to individual topics and counted the number of times each topic was mentioned. We then investigated the conditions that each participant experienced and separated the numbers into two groups, director control (DC and DR) and viewer control (VI and VE). The final results are shown in **Table 2**. The total count of sub-level topic 3 in theme 1 is not further divided as it is not condition-dependent.

For theme 1, we believe the viewer's main impression of the immersive experience changes according to whether the viewer has been given control or not. When a viewer is given control over the narrative progression and notices her agency within the experience, her main impression/recall will be mainly the feeling of "interactivity" or "being immersed," instead of focusing on the content. This conclusion is supported by the number distribution between DC and DR and VI and VE of the two sub-level topics under theme 1). This observation is further strengthened by the distribution of condition origins between the two sub-level topics under theme 2). From that distribution we can also conclude that "viewer control" leads users to remember more about the *experience*, rather than the *content*.

5.2.2 Exploratory Behaviors

On the heat maps in **Figure 7**, we examined where the red clouds were gathered (the area a viewer focused on for the longest time) and the size of the cluster (a small cluster if the viewer almost never moved head when looking at that area). We notice that most viewers' focus point fell on the host during the virtual tour because they were reminded to pay attention to the content as there will be questions asking about them at the end. Most of the viewers took it as a "test" and showed a tendency to carefully listen to the host because of this statement. Several participants also indicated that they did it out of politeness as "*the guide was visible in the scene, I naturally felt I should look at him while listening.*"

A difference between the intentions of the exploratory behaviors for the "Introduction" clip and those in the major content segments was also discovered. Many viewers showed an increased frequency of looking around and drifting away from the host during the Introduction, while the same behavior was much less observed during the content segments. In the latter case, the viewers were

mainly dwelling only on the host and the object being introduced. We also noticed that several participants chose to look at the ROIs. The reason suggested from the interviews were mainly 1) they noticed some topic that they were interested in and would like to know more about, 2) they were attracted by something in the video and wanted to have a closer look (which was not supported in the 360-degree videos). Another group of viewers showed different behavior patterns from the previous two. They kept actively looking around during the whole tour. They stated that they were new to this form of 360-degree video and new to the place shown in the video (the VR lab), thus were driven by curiosity to look around actively. It was also verified when we looked at their answers in the background questionnaires. Most of the participants who showed "curiosity" behaviors answered "never" or "only a few times a year" when asked about previous VR use or watching 360-degree videos.

We want to conclude that the participant tendencies of watching passively or actively looking around could have been affected by their preconceived impression of the experience. During the experiment, before starting the viewing sessions, we described the experience as "a virtual guided tour made from a series of 360-degree videos," and we did not mention the actual mechanism behind the scene because we did not want the participants to be aware of the differences between conditions. We also emphasized that there would be "content-related questions at the end," and that they "might want to pay attention to the content and the details." These descriptions led participants to form expectations before the tour, that these are "videos" and "they should pay attention to the content." Then they showed less exploratory behaviors in the main segments. But first impression is not only the case, personal experience and interests also contribute to exploratory behaviors.

5.2.3 The Non-Stop Flow of Time and the Control Over Pacing

The need for pace control is a factor we noticed in viewer behaviors. Several participants of the VE condition reported in the interview that they felt there was not enough time for them to make selections. Some of them saw the cards appearing in the scene, but did not have enough time to consider the consequences, and make a selection.

This indicates that pace control could also be a vital aspect of the user experience in immersive storytelling. Alongside control over story direction and progression, viewers might also desire control of the *speed* of the progression. Unlike movies, where the viewers are just passively watching and the director has total control over the pace of progression, in interactive storytelling,

the viewer will carry out active input and bear the task of determining the consequences of each option. Thus, proper pace control put in the viewer's hands could be substantial. In other words, when the storytelling process encompasses viewer interaction with the narrative, enabling the viewer to perceive herself in a role involved in the story, the control of story progression might also be rightly handed over to the viewer, at least partially, in order to match the agency a viewer now has.

5.2.4 Making Choices, or not Making Choices

We added an extra question during the interview for participants in the VE condition to ask their motivations for making their selections. Since in this condition, we found both those who actively made selections when the cards appeared and those who did not make any choices along the tour, even though the controller was in their hands, we discuss them as separate groups.

For those who made selections, the responses show that their motivations mainly fell into two categories: 1) making choices based on personal interest, and 2) making choices by comparing the running-times of the options, and choosing the shortest one. Both of these were linked to information presented on the cards (Name of the ROI, and running time of the segment). The text on the cards became an important reference for viewers to support their choices.

For the participants from VE who did not make choices, many reported they were not sure if by making the selection the system would directly cut to the next scene (like in music players), or just queue it up. They wanted to finish the current segment before proceeding, so they hesitated and eventually gave up on making selections. Once they noticed the system made default selections, they switched to a fully-passive attitude and let the defaults run while simply enjoying whatever was coming up.

Three participants can be viewed as outliers from these two groups. Their perception of explicit control differed from what we designed, or did not follow what we explained before the experiment. One indicated that he used the cards to force "the host to present the segments in a counter-clockwise fashion," as this was his personal preference when visiting museums. Two other participants reported that they regarded the cards as a method to skip and jump forward to the next segment, which is an intention we did not foresee in the design. They indicated in the interview that they wanted to skip forward once they felt bored and thought "the cards were a button to skip the current segment so I pressed it." However, the system did not allow for that, making them feel frustrated. These participants also stated that if they had been more sure of what the consequences of the selection mechanism were, they would have chosen according to whichever segment interested them most.

We assume that this confusion is related to the wording on the cards, as we wrote "Next up: [the name of the ROI]," which may have confused some participants. However, there could be a mismatch between the user's perception of the consequence of control, and the actual consequences. As mentioned by Carstensdottir et al. (Carstensdottir et al., 2019), the creator of an interactive experience needs to consider how the viewers establish a certain mental model in their minds when a control or input method is put in their hands. We need to further investigate and consider our design choices for interactive elements.

5.3 Implications for CVR Creators

We derived a first set of guidelines by comparing the interview responses of participants who went through the different conditions. We learned that the viewer's tendency to recall the experience can be mediated by whether the viewer is given control over the immersive experience. This means a CVR creator can guide the viewer's general impression of the experience towards an emotional feeling of interactive/playable progress (projects designed for fun or suspense) or towards the content itself (projects for education or training) by giving viewers different levels of control, such as solely director control, implicit viewer control, or explicit viewer control. This guideline is also a part of the APC framework we plan to provide, which will assist CVR creators in choosing proper viewer interaction designs for projects with different purposes.

Another observation is the need for balance between *the viewer's natural tendency to explore* and *incentives to invoke exploratory behaviors in viewers*. The former was observed as a result of the shift of the viewer's role in CVR. The viewer naturally has an increased tendency to actively explore the virtual environment when s/he is immersed in it, as verified by our own observations from the participant behaviors during our experiments. The latter we derive from the combination of both quantitative data and qualitative observations. We believe that the viewer's tendency to actively explore in an immersive environment can be mediated by external elements, in both top-down and bottom-up forms.

Top-down mediation is related to the tasks a viewer carries out within the immersive experience. As we saw during the experiment, while carrying out the task of "I need to pay careful attention to the speech from the tour guide," viewers showed fewer exploratory behaviors. On the other hand, when they were told to "look for interactive cards in the scene and choose which segment you want to watch next," viewers' exploratory activities increased (as seen from the heat maps generated from the VE condition).

Bottom-up mediation seems more delicate and complex. We have not collected enough data in this user study to fully understand this mechanism. Thus, we only provide our preliminary insights here. On the one hand, when viewers were provided with a controller, they treated it as an entry point for potential interaction, and tried to interact with many things. However, their *willingness to explore and interact* hit a roadblock after several tries as there are no *interactables*. This made viewers unclear of the consequences behind their actions, and made them step back into a passive-viewing mode. To sum up, we believe CVR content creators need to be aware that viewer intentions to explore actively has a firm grounding in immersive storytelling, and that viewers must strike a balance between willing to act and being afraid to act, mediated by several factors, some under creator control, and some not.

Returning to the two central questions raised near the beginning, the decision of who should control narrative progression and the visibility of the viewer's interaction option, we sum up the following guidelines for CVR creators as a starting point of applying the APC framework:

- (1) We recommend viewer control for immersive storytelling projects that focus on providing an experience of participation and feeling of agency. For other projects where learning and transfer of knowledge is the priority, the creator might want to consider still leaning more heavily on director control. While this might potentially also apply to learning experiences, it needs further investigation.
- (2) If the viewer interaction is enabled as a significant component of the storytelling experience, the creator will need to consider how to best utilize the viewer's natural tendency to explore actively. Invisible interaction is a more balanced choice, as it helps the viewer be immersed in the content while removing distractions by interactive elements.
- (3) Explicit interaction requires more careful design and increases the amount of work needed to construct the entire experience. However, if the creator can find a proper design to relieve the viewer's uncertainty of interaction consequences and properly guide viewers to use this explicit interaction method, it will yield a higher level of engagement and enjoyment in the experience.

6 CONCLUSION AND FUTURE WORK

In this paper, we first proposed Adaptive Playback Control (APC) as a prototype framework for content creators working on CVR projects with viewer interaction. We acknowledge the necessity of this framework as CVR creators have realized the shift of the viewer's roles when migrating from flat-screen media, with more emphasis on the viewer's intention to interact. We first reviewed existing frameworks for interactive non-CVR media, and identified a common three-step workflow of creating storytelling experience. Realizing it is not practical to use one framework to cover all scenarios, we first set constraints to limit the applicable realm of APC for our current work. We then constructed the APC framework as a series of guidelines for each of the steps in that workflow.

In the experiment we conducted, we further constrained the section down to a two-element combination, the narrative progression, and interaction visibility. Our formal user study evaluated four combinations from those two elements, looking at the viewer's level of engagement and enjoyment, general usability experience, and memory performance with the content of each condition. We also conducted semi-structured interviews to assess their subjective feedback. The objective data did not reveal any significant differences. However, by combining both objective and subjective data, we provided guidelines to CVR creators as the first part of the APC framework. We suggest that creators need to design control modes and interaction visibility based on the original purpose of the storytelling experience.

We also understand that the content we used was a self-made "virtual guided tour" based on one possible type of story structure, "hub-and-spoke." Compared to the broad design space that the APC framework covers, this is only a starting point. In the user study, each participant only experienced one of the four conditions, and might have overlooked the differences between the DC, DR, and VI conditions. This might be the reason for the lack of significant differences we found among the objective measures.

We also did not foresee the potential relationship between the viewer's tendency toward exploratory behavior and how the experience and tasks were introduced before they started. In this user study, we mainly looked at the visibility of viewer interaction methods while limiting the form of viewer input to controller-based "laser pointer" only, and the destination of viewer's interaction only to narrative progression. The effects of other parameters with possible link to viewer control, such as the temporal sequencing, pace, and speed of progression, still need further exploration. In the future, we will explore them and add them as pieces forming a comprehensive APC framework to cover various demands in effective CVR content creation and production.

We will also expand our exploration to other story structures suitable for CVR, including "detours", "string-of-pearls" and the "gauntlet" (unlike "hub-and-spoke", they are linear based and only contain sub-branches, cf. (Carstensdottir et al., 2019)). Another dimension related to story structure is the "alternative takes" of the same segment in content. The term "takes" is borrowed from the filmmaking industry to describe a cluster of multiple captured clips of the same story segment. Content-wise they are the same story, but in detail, each of them is unique, because just like an oral storyteller will use different words, sentences, or expressions when telling the same story at different times, characters performing the same story will also bring in variations to their acting, making one take always slightly different from another. In the future, we can tweak one segment into several variations with different takes, to match different viewers with different profiles. Since different viewers will have different expectations towards the story in their minds, if the content is somehow adjusted to match the viewer's profile, everyone will find the content they experience fitting their expectations.

We also plan to further investigate the "embodiment of host" in guided-tour style experiences. During the study, some participants provided feedback that they preferred a narration-only style where the host was not in the scene, but was a disembodied voice-over. Because they thought the host himself would attract their attention, and would therefore be less likely to explore, as this might be seen as "inappropriate," voice-over might lead to less-restrictive behavior. Although the voice-over style is out of the scope of current APC framework, it does point out the necessity to consider the opposite side of attention-directing methods we have been discussing. Since the embodiment of the storyteller can be more than just a recorded real human, computer generated avatars and manipulations can be employed to see if they help to relieve the effect of "an attention anchor," and give viewers more freedom to choose what they want to focus on, while not increasing the feeling of "fear of insulting the host." These will all contribute to the construction of an effective APC framework.

DATA AVAILABILITY STATEMENT

The dataset presented in this study is not available for sharing publicly as regulated by the Human Research Ethics Committee of University of Canterbury, which does not allow sharing data to a party other than the researchers within this research project.

However, any concern or question about the dataset should be directed to the authors.

ETHICS STATEMENT

The study involving human participants was reviewed and approved by the Human Research Ethics Committee, University of Canterbury. The participants provided their written consent to participate in this study. Written consent forms were obtained from the individual(s) for the publication of any potentially identifiable images, transcribed audio recordings, and data included in this article.

AUTHOR CONTRIBUTIONS

In this article the contributions of authors are as follows. LT contributed to conceptualization, methodology, investigation, implementation, data collection, visualization and writing; RL

contributed to conceptualization, methodology, reviewing, editing, supervision, and funding acquisition; HR contributed to reviewing, editing, project administration and funding acquisition.

FUNDING

The authors would like to thank the support of other researchers from HIT Lab NZ, University of Canterbury, and research members from the New Zealand Science for Technological Innovation National Science Challenge (NSC) project Atea.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/frvir.2021.798306/full#supplementary-material>

REFERENCES

- Bender, S. (2019). Headset Attentional Synchrony: Tracking the Gaze of Viewers Watching Narrative Virtual Reality. *Media Pract. Educ.* 20, 277–296. doi:10.1080/25741136.2018.1464743
- Bevan, C., Green, D. P., Farmer, H., Rose, M., Cater, K., Stanton Fraser, D., et al. (2019). “Behind the Curtain of the “Ultimate Empathy Machine”: On the Composition of Virtual Reality Nonfiction Experiences,” in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, May 4–9, 2019 (New York, NY, USA: Association for Computing Machinery), 1–12.
- Bordwell, D., and Thompson, K. (2013). *Film Art: An Introduction*. 10th edn. New York: McGraw-Hill.
- Brewster, S. (2017). Designing for Real Feelings in a Virtual Reality. [Dataset]. Available at: <https://medium.com/s/designing-for-virtual-reality/designing-for-real-feelings-in-a-virtual-reality-41f2a2c7046> (Accessed October 05, 2021).
- Carstendottir, E., Kleinman, E., and El-Nasr, M. S. (2019). “Player Interaction in Narrative Games,” in Proceedings of the 14th International Conference on the Foundations of Digital Games-FDG '19, San Luis Obispo, CA, USA, August 26–30, 2019 (New York, NY, USA: Association for Computing Machinery), 1–9. doi:10.1145/3337722.3337730
- Cavazza, M., and Charles, F. (2005). “Dialogue Generation in Character-Based Interactive Storytelling,” in Proceedings of the First AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment-AIIDE'05, Marina del Rey, CA, USA, June 1–3, 2005 (Marina del Rey, California: AAAI Press), 21–26.
- Dooley, K. (2017). Storytelling with Virtual Reality in 360-degrees: a New Screen Grammar. *Stud. Australas. Cinema* 11, 161–171. doi:10.1080/17503175.2017.1387357
- Ferguson, C., van den Broek, E. L., and van Oostendorp, H. (2020). On the Role of Interaction Mode and Story Structure in Virtual Reality Serious Games. *Comput. Educ.* 143, 103671. doi:10.1016/j.compedu.2019.103671
- Gödde, M., Gabler, F., Siegmund, D., and Braun, A. (2018). “Cinematic Narration in VR - Rethinking Film Conventions for 360 Degrees,” in International Conference on Virtual, Augmented and Mixed Reality, Las Vegas, FL, USA, July 15–20, 2018, 184–201. doi:10.1007/978-3-319-91584-5_15
- Gustafsson, A., Katzeff, C., and Bang, M. (2009). Evaluation of a Pervasive Game for Domestic Energy Engagement Among Teenagers. *Comput. Entertain.* 7, 1–54. doi:10.1145/1658866.1658873
- Haggood, J., Moore, D., Alapont, S., Ferguson, C., and van Oostendorp, H. (2018). “The REVEAL Educational Environmental Narrative Framework for PlayStation VR,” in Proceedings of the 12th European Conference on Games Based Learning, Sophia Antipolis, France, October 4–5, 2018 (Academic Conferences and Publishing International Limited), 175–183.
- Hassan, R. (2020). Digitality, Virtual Reality and the ‘Empathy Machine’. *Digit. Journal.* 8, 195–212. doi:10.1080/21670811.2018.1517604
- Ibanez, J., Aylett, R., and Ruiz-Rodarte, R. (2003). Storytelling in Virtual Environments from a Virtual Guide Perspective. *Virtual Reality* 7, 30–42. doi:10.1007/s10055-003-0112-y
- Ip, H. H. S., Li, C., Leoni, S., Chen, Y., Ma, K.-F., Wong, C. H.-t., et al. (2019). Design and Evaluate Immersive Learning Experience for Massive Open Online Courses (MOOCs). *IEEE Trans. Learn. Technol.* 12, 503–515. doi:10.1109/TLT.2018.2878700
- Koenitz, H. (2015). “Towards a Specific Theory of Interactive Digital Narrative,” in *Interactive Digital Narrative* (New York City, NY, USA: Routledge), 15. doi:10.4324/9781315769189-8
- Lorch, R. F., Lorch, E. P., and Inman, W. E. (1993). Effects of Signaling Topic Structure on Text Recall. *J. Educ. Psychol.* 85, 281–290. doi:10.1037/0022-0663.85.2.281
- Lyk, P. B., Majgaard, G., Vallentin-Holbech, L., Guldager, J. D., Dietrich, T., Rundle-Thiele, S., et al. (2020). Co-Designing and Learning in Virtual Reality: Development of Tool for Alcohol Resistance Training. *Electron. J. e-Learn.* 18, 219–234. doi:10.34190/ejel.20.18.3.002
- Mateas, M., and Stern, A. (2006). “Towards Integrating Plot and Character for Interactive Drama,” in *Socially Intelligent Agents* Boston, Massachusetts, USA: Springer, 221–228. doi:10.1007/0-306-47373-9_27
- Mateer, J. (2017). Directing for Cinematic Virtual Reality: How the Traditional Film Director’s Craft Applies to Immersive Environments and Notions of Presence. *J. Media Pract.* 18, 14–25. doi:10.1080/14682753.2017.1305838
- Moser, C., and Fang, X. (2015). Narrative Structure and Player Experience in Role-Playing Games. *Int. J. Hum. Comput. Int.* 31, 146–156. doi:10.1080/10447318.2014.986639
- Mulholland, P., Wolff, A., Zdrahal, Z., Li, N., and Corneli, J. (2013). “Constructing and Connecting Storylines to Tell Museum Stories,” in *Interactive Storytelling*. Editors H. Koenitz, T. I. Sezen, G. Ferri, M. Haahr, D. Sezen, and G. Catak (Cham: Springer International Publishing), 121–124. doi:10.1007/978-3-319-02756-2_14
- O’Brien, H. L., Cairns, P., and Hall, M. (2018). A Practical Approach to Measuring User Engagement with the Refined User Engagement Scale (UES) and New UES Short Form. *Int. J. Hum. Comput. Stud.* 112, 28–39. doi:10.1016/j.ijhcs.2018.01.004
- Pagano, A., Palombini, A., Bozzelli, G., De Nino, M., Cerato, I., and Ricciardi, S. (2020). ArkaeVision VR Game: User Experience Research between Real and Virtual Paestum. *Appl. Sci.* 10, 3182. doi:10.3390/app10093182

- Pillai, J. S., and Verma, M. (2019). "Grammar of VR Storytelling: Analysis of Perceptual Cues in VR Cinema," in *European Conference on Visual Media Production*, London, UK, December 17–18, 2019 (London United Kingdom: ACM), 1–10. doi:10.1145/3359998.3369402
- Pope, V. C., Dawes, R., Schweiger, F., and Sheikh, A. (2017). "The Geometry of Storytelling," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, Denver, Colorado, USA, May 6–11, 2017 (Denver Colorado USA: ACM), 4468–4478. doi:10.1145/3025453.3025581
- Rezk, A. M., and Haahr, M. (2020). "The Case for Invisibility: Understanding and Improving Agency in Black Mirror's Bandersnatch and Other Interactive Digital Narrative Works," in *Interactive Storytelling*. Editors A.-G. Bossler, D. E. Millard, and C. Hargood (Cham: Springer International Publishing), 178–189. doi:10.1007/978-3-030-62516-0_16
- Roth, C., van Nuenen, T., and Koenitz, H. (2018). "Ludonarrative Hermeneutics: A Way Out and the Narrative Paradox," in *Interactive Storytelling*. Editors R. Rouse, H. Koenitz, and M. Haahr (Cham: Springer International Publishing), 11318, 93–106. doi:10.1007/978-3-030-04028-4_7
- Rothe, S., Buschek, D., and Hußmann, H. (2019a). Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. *Multimodal Technol. Interact.* 3, 19. doi:10.3390/mti3010019
- Rothe, S., Pothmann, P., Drewe, H., and Hussmann, H. (2019b). "Interaction Techniques for Cinematic Virtual Reality," in 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019-Proceedings, Osaka, Japan, March 23rd–27th, 2019, 1733–1737. doi:10.1109/VR.2019.8798189
- Ryan, M.-L. (2008). "Interactive Narrative, Plot Types, and Interpersonal Relations," in *Interactive Storytelling*. Editors U. Spierling and N. Szilas (Berlin, Heidelberg: Springer), 6–13. doi:10.1007/978-3-540-89454-4_2
- Schmitz, A., Macquarrie, A., Julier, S., Binetti, N., and Steed, A. (2020). "Directing versus Attracting Attention: Exploring the Effectiveness of Central and Peripheral Cues in Panoramic Videos," in *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020, Atlanta, Georgia, USA, March 22–26, 2020*, 63–72. doi:10.1109/vr46266.2020.00024
- Sengun, S. (2013). "Silent Hill 2 and the Curious Case of Invisible Agency," in *Interactive Storytelling*. Editors H. Koenitz, T. I. Sezen, G. Ferri, M. Haahr, D. Sezen, and G. Catak (Cham: Springer International Publishing), 180–185. doi:10.1007/978-3-319-02756-2_22
- Shah, S. H. H., Han, K., and Lee, J. W. (2020). Real-Time Application for Generating Multiple Experiences from 360° Panoramic Video by Tracking Arbitrary Objects and Viewer's Orientations. *Appl. Sci.* 10, 2248. doi:10.3390/app10072248
- Sharaha, I., and Al Dweik, A. (2016). "Digital Interactive Storytelling Approaches : A Systematic Review," in *Computer Science & Information Technology (CS & IT) (Chennai, Tamil Nadu, India: Academy & Industry Research Collaboration Center, AIRCC)*, 21–30. doi:10.5121/csit.2016.61002
- Sharples, M., Fitzgerald, E., and Mulholland, P. (2013). "Weaving Location and Narrative for Mobile Guides," in *Museum Communication and Social Media* (New York City, NY, USA: Routledge), 20.
- Syrett, H., Calvi, L., and van Gisbergen, M. (2017). "The Oculus Rift Film Experience: A Case Study on Understanding Films in a Head Mounted Display," in *Intelligent Technologies for Interactive Entertainment: 8th International Conference, INTETAIN 2016, Utrecht, The Netherlands, June 28–30, 2016*. doi:10.1007/978-3-319-49616-0_19
- Tong, L., Jung, S., Li, R. C., Lindeman, R. W., and Regenbrecht, H. (2020). "Action Units: Exploring the Use of Directorial Cues for Effective Storytelling with Swivel-Chair Virtual Reality," in 32nd Australian Conference on Human-Computer Interaction, Virtually online, December 2–4, 2020 (Sydney NSW Australia: ACM), 45–54. doi:10.1145/3441000.3441063
- Tong, L., Lindeman, R. W., and Regenbrecht, H. (2021). Viewer's Role and Viewer Interaction in Cinematic Virtual Reality. *Computers* 10, 66. doi:10.3390/computers10050066
- Ursu, M. F., Thomas, M., Kegel, I., Williams, D., Tuomola, M., Lindstedt, I., et al. (2008). Interactive TV Narratives. *ACM Trans. Multimedia Comput. Commun. Appl.* 4, 1–25. doi:10.1145/1412196.1412198
- Verdugo, R., Nussbaum, M., Corro, P., Nuñez, P., and Navarrete, P. (2011). Interactive Films and Coconstruction. *ACM Trans. Multimedia Comput. Commun. Appl.* 7, 1–39. doi:10.1145/2043612.2043617
- Yu, K. A. (2019). "The Library: A Non-intrusive Gaze Directed Virtual Reality Animation," in 2019 IEEE 2nd Workshop on Animation in Virtual and Augmented Environments (ANIVAE), Osaka, Japan, March 24 2019, 1–4. doi:10.1109/ANIVAE47543.2019.9050930

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Tong, Lindeman and Regenbrecht. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.