UNIVERSITY OF CANTERBURY

DOCTORAL THESIS

Exploring Effective Storytelling Guidelines for Cinematic Virtual Reality

Author:
Lingwei TONG

Supervisor:
Prof. Robert W. LINDEMAN
Prof. Holger REGENBRECHT
Dr. Rory CLIFFORD

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

The Human Interface Technology Laboratory New Zealand

August 31, 2022

Declaration of Authorship

I, Lingwei TONG, declare that this thesis titled, Exploring Effective Storytelling Guidelines for Cinematic Virtual Reality and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: LINGWEI TONG

Date: August 31st, 2022

Abstract

Lingwei TONG

Exploring Effective Storytelling Guidelines for Cinematic Virtual Reality

Content creators have been exploring ways to use virtual reality (VR) as an effective storytelling tool. The term cinematic virtual reality (CVR) was then created to describe the kind of VR experience that is produced using pre-rendered content with lengthier and complete story structures, with the interaction design that enables viewers to actively choose where to look. Initially, creators of CVR content began by transferring storytelling grammars and techniques from mature media, such as cinema and theater. However, specific challenges for CVR followed, including the narrative paradox (NP) (which is the conflict and tension arising between authorial control and viewer agency), the fear of missing out (FOMO), and the discrepancy between viewer expectations on agency and the system's interactive capacity. Because CVR is a type of immersive experience, viewers are also inclined to interact with the story world freely. To achieve a final product that is a successful and engrossing storytelling experience, creators must address the NP and FOMO issues and establish a design balance between authorial control and viewer participation in terms of narrative progression.

To investigate the issues raised above and assess potential solutions, several user studies were undertaken in this thesis. A human body-language-based attention guidance cue set called Action Units (AUs) was created to address the FOMO issue. It was then compared with two other commonly used synthetic cues for user experiences. According to the findings, the use of AUs in CVR content can boost viewer enjoyment and engagement with the story. The AUs were also favored by viewers for their diegetic qualities and by creators for the simplicity of use. Moving on to the NP issue, the second user study sought to identify the upper limit of a viewer's desire to actively interact and participate in the narration. Results indicated that viewer control is advised for CVR projects. To handle viewer curiosity and motivate them to interact freely, creators must carefully set up the interactors. Based on the findings, a coherent framework was researched and developed by tying together previously acquired knowledge and rules that were dispersed

to various components of producing CVR with the workflow that a creator uses to build the experience. The procedure resulted in a formalized framework called the Adaptive Playback Control (APC) for CVR. The APC starts by guiding content preparation by highlighting the need for applying diegetic attention guidance cues. It also includes guidelines for interactive design by emphasizing the need for design considerations regarding the harmony between viewer and creator roles in directing the narrative development, and raising the visibility of interaction affordances in the immersive storytelling experience. Then, a real-world case study of applying the APC to an immersive Māori (New Zealand indigenous people) storytelling experience was presented. The case study examined whether viewer-participatory design, including profiling viewers and the strategies to introduce narrative variations, was culturally appropriate. In this case study, personalized variations were added to CVR by taking into account both the unique demands of each viewer and their participation in the storytelling process. Insights from the case study showed that for creators to safely guarantee that experiences will live up to viewer expectations and be entertaining and diverse, individual users must also be taken into account from the very beginning of content design.

Finally, this thesis offers the Adaptive Playback Control (APC), a novel framework for those who create CVR experiences. They can follow the framework's instructions to create materials specifically designed for an immersive experience utilizing pre-rendered content, such as 360-degree videos. It intends to address the FOMO issue and help creators produce CVR experiences with correct viewer interaction and integrated viewer personalization, resolving the problem of NP and improving the overall experience. This thesis also employed a case study to show how adaptable the framework is and how it may be used in a larger context, in and beyond the cultural heritage sector.

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Sections 2.3 and 2.4 in Chapter 2 of the thesis is a reproduction of the work conducted and published in collaboration with Robert W. Lindeman and Holger Regenbrecht. The result of the work was published in a peer reviewed journal as:

L. Tong, R. W. Lindeman, and H. Regenbrecht, "Viewer's Role and Viewer Interaction in Cinematic Virtual Reality," Computers, vol. 10, no. 5, Art. no. 5, May 2021, doi: 10.3390/computers10050066.

Please detail the nature and extent (%) of contribution by the candidate:

Robert W. Lindeman and Holger Regenbrecht helped on research conceptualization, methodology design, review and feedback on the final manuscript. The candidate did all the writing and editing of the paper. The candidate's contribution is more than 80%.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all

The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: **Robert W. Lindeman** Signature: Date: **2022-08-16**

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 3 of the thesis is a reproduction of the work conducted and published in collaboration with Robert W. Lindeman, Sungchul Jung, Richard Chen Li and Holger Regenbrecht. The result of the work was presented in the following conference and then published in its proceedings:

32nd Australian Conference on Human-Computer Interaction, Sydney NSW Australia, Dec. 2020;

L. Tong, S. Jung, R. C. Li, R. W. Lindeman, and H. Regenbrecht, "Action Units: Exploring the Use of Directorial Cues for Effective Storytelling with Swivel-chair Virtual Reality," in 32nd Australian Conference on Human-Computer Interaction, Sydney NSW Australia, Dec. 2020, pp. 45–54. doi: 10.1145/3441000.3441063.

Please detail the nature and extent (%) of contribution by the candidate:
Sungchul Jung and Richard Chen Li both provided suggestions about user study design and data analysis strategies. Robert W. Lindeman also provided feedback on user study designs. Robert W. Lindeman and Holger Regenbrecht helped on reviewing the draft manuscript. The candidate carried out the user study and did all the writing and editing of the paper. The candidate's contribution is more than 75%.

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: **Robert W. Lindeman** Signature: Date: **2022-08-16**

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4 of the thesis is a reproduction of the work conducted and published in collaboration with Rob Lindeman and Holger Regenbrecht. The result of the work was published in a peer reviewed journal as:

*L. Tong, R. W. Lindeman, and H. Regenbrecht, "Adaptive Playback Control: A Framework for Cinematic VR Creators to Embrace Viewer Interaction," Frontiers in Virtual Reality, vol. 2, 2022, *
Available: https://www.frontiersin.org/article/10.3389/frvir.2021.798306

Please detail the nature and extent (%) of contribution by the candidate:

Robert W. Lindeman and Holger Regenbrecht helped on methodology idealization, and provided feedback on the final manuscript. Robert W. Lindeman also helped on content preparation and feedback on user study data analysis. The candidate carried out the user study and did all the writing and editing of the paper. The candidate's contribution is more than 80%.

Certification by Co-authors:

If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: **Robert W. Lindeman** Signature: Date: **2022-08-16**

Acknowledgements

This study, spanning from mid-2018 to the end of 2022, is one experience in my life that nothing compares to, or even comes close. It was a choice, a challenge, a one-way ticket, a journey, and a quest where I hesitated at the beginning, doubted, hated, liked, and cringed unlimited times during the path, but did not regret it at the end. It was beyond what texts can describe. But I am lucky that I was not alone. On this path, many walked with me and many have been there since the beginning all the way to the end. And I want to say thank you, thank you 3000, thank you 5000 times, to all of them.

First I would like to wholeheartedly and deeply thank my senior supervisor Prof. Rob Lindeman, who helped me choose and shape the research topic, assisted me on everything of conducting research, offered valuable advice when I felt uncertain, and provided support and love when I got lost (especially during the COVID years, when I went on to a suspension).

Then I would like to sincerely thank my associate-supervisor Prof. Holger Regenbrecht who provided support and advice on conducting research, on using scientific mindsets on problem solving and turning results into publications. I also want to thank associate-supervisor Dr. Rory Clifford, who assisted me with immersive storytelling for Mātauranga Māori in Bluff. I could not have reached this point today without the precious help, love, and guidance from these amazing mentors.

My appreciation also goes to my once-supervisors Dr.Sungchul Jung, and Dr. Richard Chen Li. They both offered much help and especially hands-on experience about user study design. Those were valuable and necessary during the early stages when I was unfamiliar and always getting messy with conducting research (still messy though). I also want to say thank you to tohunga Prof. Hēmi Whaanga, Dean Whaanga, Bubba Thompson, and more from Te Rūnanga o Awarua, Te Rau Aroha Marae in Bluff, and other research members in the Ātea project, who offered help during my studies. The same gratitude also goes to the New Zealand Science for Technological Innovation National Science Challenge (NSC SfTI) project *Ātea*, which funded my PhD research.

But life is more than just doing research. Thanks to my good friend at the lab, Yuanjie Wu who offered tips and sparks on everything; lab veteran Huidong Bai for every now and then reminding me it is totally fine to try something crazy, and other friends including Le, 7, Zac, and many more who shared memorable moments and stories, making my life full of color, laughter and tears. Most importantly, great great thanks to my parents, who always support me and gave selfless, unconditional, unshakable, strongest-in-the-universe love when their son chose to do a PhD and went in an opposite direction from most of the kids in other Chinese families. Great thanks to all, and I will remember every moment.

最后我将最真挚真切的感激致我的父亲母亲；他们在我的留学期间给予了无数的支持；在我选择和绝大多数中国家庭的孩子相向而行，只身踏上这条路时，亦无反对，而是继续给出他们无私的、无条件的、从不犹豫的、全世界最深的爱和支持。若没有我父母的支持，我绝无法一路坚持走到今天和现在。如今一段旅途已经结束，新的未来即将开启，求索路漫漫，记忆远流长，感情真无价，已无须多言。

Preface

The research described in this thesis has generated the following publications, which have appeared in the indicated peer-reviewed journals and conference proceedings:

Chapter 2 Background

Section 2.3 Viewer Interaction in Immersive Media

L. Tong, R. W. Lindeman, and H. Regenbrecht, "Viewer's Role and Viewer Interaction in Cinematic Virtual Reality," Computers, vol. 10, no. 5, Art. no. 5, May 2021, doi: 10.3390/computers10050066.

Chapter 3 Producing CVR Content with Attention Guidance

Section 3.3 User Evaluation of AUs

L. Tong, S. Jung, R. C. Li, R. W. Lindeman, and H. Regenbrecht, "Action Units: Exploring the Use of Directorial Cues for Effective Storytelling with Swivel-chair Virtual Reality," in 32nd Australian Conference on Human-Computer Interaction, Sydney NSW Australia, Dec. 2020, pp. 45–54. doi: 10.1145/3441000.3441063.

Chapter 4 Viewer Participation and Interaction in CVR

Section 4.2 Two Interaction Elements - Control and Visibility

Section 4.3 The Experiment

L. Tong, R. W. Lindeman, and H. Regenbrecht, "Adaptive Playback Control: A Framework for Cinematic VR Creators to Embrace Viewer Interaction," Frontiers in Virtual Reality, vol. 2, 2022,
Available: https://www.frontiersin.org/article/10.3389/frvir.2021.798306

Contents

List of Figures

List of Tables

Acronyms

APC Adaptive Playback Control. iv, x, xi, xiv, 7–12, 90, 91, 93–100, 102–108, 110, 112, 114, 116, 118, 120, 122, 124, 126–128, 130, 132, 134, 136–158

AR augmented reality. 30

AUs Action Units. iii, vii, ix, xiii, 7, 10, 11, 48–53, 55, 58, 64–66, 93, 98, 99, 110, 134, 137–139, 142, 143, 147, 150, 151, 155

CVR cinematic virtual reality. iii, iv, vii–xi, xiii, 2–13, 15, 16, 24–27, 29–37, 39–43, 45–48, 50, 52, 54, 56, 58, 60, 62, 64–68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90–108, 110, 125, 133, 137–139, 142–152, 154–158

DOF degree of freedom. 145, 150, 155–157

FOMO fear of missing out. iii, iv, 2, 5, 6, 19, 25, 27, 28, 45, 47, 53, 65, 74, 93, 95, 97, 98, 146–149, 156

FOV field of view. 48, 54, 56, 73, 74, 137, 143

HMD head-mounted display. 17, 24, 29, 52, 54–57, 70, 77, 100, 101, 130, 132, 133, 135, 143

HUD heads-up display. 74, 75, 77

MS&T Matrix of Segments and Takes. xiv, 108–110, 126, 140

NP narrative paradox. iii, iv, 1, 5, 6, 16, 25, 27, 28, 45, 47, 65, 66, 146, 147, 149, 156

POV point of view. 1, 14, 17, 20, 37, 38, 45, 47, 95, 143

ROI region of interest. xiv, 14, 21, 27, 28, 34, 38, 39, 47, 48, 50, 52–56, 60–64, 74, 87, 89, 93, 110, 137, 150, 151, 157

SME subject-matter expert. xiii, 50–53, 106, 111

UI user interface. xiii, 36–38, 129

VR virtual reality. iii, viii, xiii, 1, 2, 13, 14, 16–24, 30, 32, 39, 45, 47, 50–53, 56, 57, 62, 66, 71, 72, 77, 78, 83, 87, 100, 102, 120, 130, 132, 147, 151

Chapter 1

Introduction

Virtual reality (VR) is regarded as a storytelling medium somewhat different from cinema, theater, television, and literature as it completely blocks out the real world, immersing the viewer in a virtual environment of visual and audio sensory stimuli [98]. Because of this novelty, content creators have become very attracted to this new medium, finding ways to adopt VR into their storytelling routines to harness its main ability of fully immersing its viewers. However, being vastly different from the mature media of cinema and theater, VR has brought new challenges to the delivery of engaging and memorable storytelling experiences. In cinema, directors use filmmaking grammar such as mise-en-scène[1] and cinematography to invoke suspense, drive curiosity and guide the viewers through the plot, thereby creating expressive stories [17]. In theater, directors, set designers and other practitioners rely on stage setup and spatial visual elements to tell their story and guide viewers. Since theatrical performances cannot rely on the same visual elements as those commonly found in cinema, such as frames and edits, theater performances employ lighting, stage setup, sound, and blocking (movements of actors) to navigate the flow of storytelling and lead the viewers [5, 94].

Able to draw from the well-developed storytelling toolkits used in cinema and theater upfront, the first VR content creators naturally applied them to this medium. Early work mainly experimented with 360-degree video, which is an easy and low-cost means of creating and consuming VR content. From a series of 360-degree video projects, creators noticed that there was a change in users' point of view (POV) compared to both cinema and theater, as the viewer was now inside the scene instead of watching it from outside the screen or stage [39]. Two new challenges also emerged in storytelling through VR. The first is the narrative paradox (NP), which is the conflict between authorial control and the viewer's free will and

[1]Mise-en-scène is a French term which means "placing on stage" or "what is put into the scene." It describes the stage design and arrangement of actors in scenes for a theater or film production.

agency. The second is the fear of missing out (FOMO), a side effect that appears when the viewer has free will to interact in the storytelling scene. Those challenges do not exist in cinema and theater [161, 86]. As 360-degree videos became more popular, creators also tried to produce more engaging experiences with complete narratives. Here the term cinematic virtual reality (CVR) is employed to define these experiences from the perspective of content format, content plot and viewer interaction, and to differentiate them from other types of VR experiences. In terms of content format, unlike traditional VR which uses real time generated graphics, CVR uses pre-rendered visual and sound elements that are either captured from the real-world or are synthetic [93]. From the perspective of plot, CVR is a type of narrative-focused fiction story or documentary, or a hybrid of both. Rather than short, non-interactive clips for temporary stimulation, CVR creators aimed for prolonged content with a narrative arc that could lead the viewers through a well-prepared emotional engagement [39]. In terms of interactivity, a CVR viewer can look around inside the 360-degree spherical scene, freely choosing what to look at [93]. The CVR experiences produced and used throughout this thesis all follow this definition.

Evolved from simple 360-degree videos, CVR allowed for expansion of the abundance and depth of content delivered. It places the viewer into a rich environment and gradually leads them from being a passive spectator to being a character inside the story world [158]. First, users of CVR start to expect some agency as they find themselves now inside the story scene, with the potential to interact with objects and characters in the scene and able to intervene in or control the unfolding of the narrative instead of watching it passively from behind the fourth wall [2], isolated from the progression of the story [4]. Second, creators and directors also realized that their absolute authorial control in CVR storytelling was also diminishing, since users were stepping into their roles and influencing the storytelling activity [157]. Evidently, the definitive separation between the role of *creator* (those who create the content and control the narrative experience) and *viewer* (those who watch/consume the content) is blurred in CVR, leading to creators' realisation of the necessity of having *viewer interaction* within CVR. This change has expanded the design space around CVR storytelling from the linear form used by filmmakers and theater practitioners [17, 31] to new territories more similar to those found in video games, live streaming, and remote collaboration [33, 141, 144]. Inspired by

[2] *Fourth wall* is a conventional term for the invisible wall that separates actors from the audience. While the audience can see through this wall, the actors act as if they cannot.

Ens et al. [44], the larger design space can be visualized as having two dimensions. The first dimension is a system's design features in terms of whether the creator and the viewer are sharing the same space (a shared space) and whether they are communicating with each other in real time (a synchronized experience). These two characteristics are then used to set up a 2×2 matrix to allocate various interaction and communication scenarios between the creator and the viewer in different types of media, as shown in Part A of Figure 1.1. The second dimension pertains to the details of the creator's authorial control and the viewer's free interaction in each scenario. Specific examples from people's daily lives were chosen in this chapter for easy visualization and understanding. It is noteworthy that categories 1 and 2 apply to both online (virtual) and offline (physical) scenarios. For example, the features of a face-to-face classic lecture are equivalent to an online zoom lecture, and similarly, an online discussion board to a physical billboard. CVR is listed in category 4 of the design space, similar to online videos, but has its own attributes when it comes to *roles of control*. These sectors and examples are illustrated in Part B of Figure 1.1.

In CVR, the aforementioned changes in both the relationship between creator and viewer and the capabilities of each party lead to new challenges. Creators are now required to consider more carefully questions that are specific to CVR, especially CVR with viewer interaction. These unresolved challenges are derived from the dimensions illustrated in Figure 1.1, including the viewer's role, the viewer's high expectation of interacting, the viewer's control over the narrative progression, and the creator's authorial control over the plot and story arc. Various answers to these questions lead to an evolution of the interactive perspective of the CVR experience, facilitating exploration towards newer and experimental CVR content and user experience. To summarise, the creators are responsible for maintaining a delicate balance among all factors to ensure an engaging and effective storytelling experience. At time of writing these questions remain open, and the definition of CVR is also evolving. Many creators and researchers are still exploring and experimenting with CVR. A definitive, conclusive so-called golden rule or all-purpose formula for CVR experiences does not yet exist (and indeed may never exist).

1.1 Motivation and Methodology

Tools, grammars, and skills for CVR creators have been explored since the introduction of 360-degree video into the storytelling realm. Some creators are totally

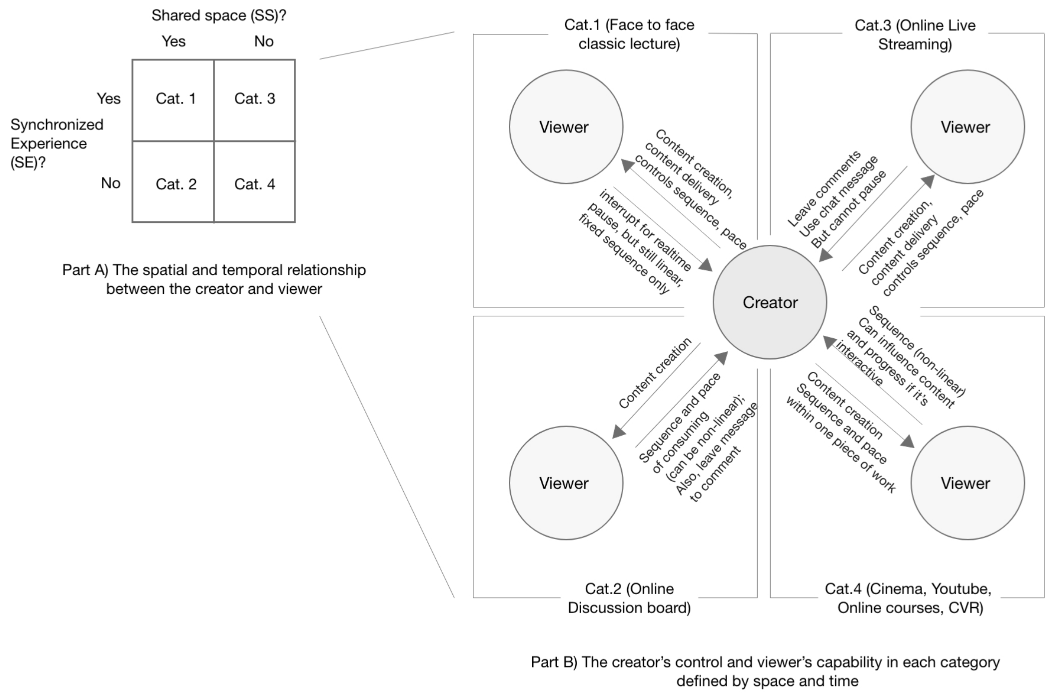


FIGURE 1.1: A series of examples illustrating the design space around CVR storytelling with viewer interaction. Part A shows a 2×2 matrix of interaction and communication scenarios between creator and viewer in different types of media. In Part B, several examples of each category are filled into the matrix. Each category is further explained by looking at what is in the creator's control and what is in the viewer's control.

new to CVR storytelling while others come from existing realms, especially filmmaking, theater practice, or the gaming industry [51]. Different creators emphasize different types of content, with some focusing on fictional content [8, 79, 105] and others on real documentary content [10, 9, 32]. The issues they are trying to resolve also vary. Since it is not practical to try to resolve all CVR-related issues and apply solutions to all types of CVR content, an *hourglass approach* was applied in this thesis (see Figure 1.2). Among various types of CVR experiences and issues creators face, the focus of this research was narrowed to begin with one specific type of CVR content, the guided tour. The exploration began with an attempt to apply 360-degree video to scientific outreach for scientists working in Antarctica under AntarcticaNZ's expedition programs. In the early stage of the study, several Antarctica experts were trained in the use of 360-degree video to capture their work on the ice. The content was presented in a *guided-tour style* as the experts usually appeared in the scene and introduced their work orally. From this experience, core challenges for CVR content creation emerged: that CVR has a FOMO issue and an NP issue. Both were found to negatively affect the viewer experience, causing loss of the track of the story, confusion, and frustration. Mitigating NP and FOMO is necessary for CVR creators to effectively deliver complex and viewer-engaging content. After exploring and discussing solutions to these problems, we then evaluated them to assess their suitability for application to other types of CVR content in different realms and also how they might be used to solve related challenges and problems. The two challenges of FOMO and NP were identified as the study moved downwards along the "hourglass" and expanded to adjacent CVR realms (such as indigenous storytelling in New Zealand).

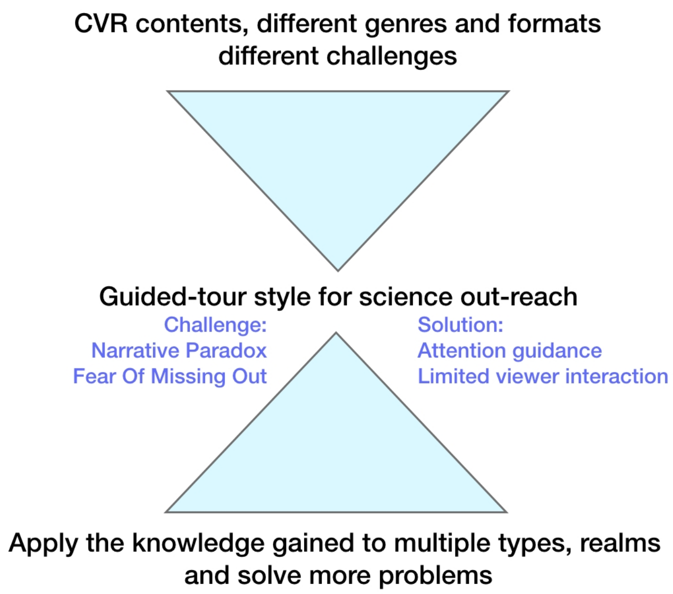


FIGURE 1.2: An illustration of the hourglass approach, showing how this study focused on one type of CVR content and the two most relevent challenges this posed. This thesis then endeavored to apply the knowledge gained to a broader range of CVR content types and issues, like moving down along an *hourglass*.

1. In most CVR projects, creators still hold a filmmaking mindset and aim to produce a passive experience. The viewer's role, expectation of interaction, and agency are not considered. The viewer's intention and their actual experience thus do not match.

2. Even with some interactive CVR projects where the viewer is considered, viewer differences are not taken into consideration and the digital experience is not

tailored to each viewer's own knowledge, features, characteristics, and tendencies. This means that despite each viewer's unique background and interests, their attributes are not related to the disclosure of the story they face, and thus the experience and outcome will be same for all viewers.

Based on the information learned during the first stage of the study, this work additionally explored and looked for solutions to the NP and FOMO problems. Overall, the vision of this work is to improve the CVR experience for both creators and viewers, whereby well-established guidelines are first produced for the guided-tour style CVR and then expanded to other kinds of CVR content more broadly. The guidelines can help creators to reach a point where they can easily capture content with CVR drawing on a well-versed narrative language to express what they want to capture and show, just as film directors, cinematographers, and actors are able to effectively communicate using their common screen grammar. Creators can also develop and update guidelines based on the findings reported in this thesis. For the viewers, their experience will be more engaging and immersive because they will be able to participation in the narrative progress and forge their personal connections to the content, thus leading to a more emotional connection to the experience afterward and unforgettable memories. With well-established guidelines, the CVR production process can be more complete and satisfactory for all involved parties.

1.2 Research Questions

From the motivations and methodology discussed in section 1.1, the following research questions were identified:

- RQ_1: For *guided-tour*-style CVR content, can social cues from the characters be used as an attention-guiding technique to resolve NP and FOMO issues in CVR?
  - Will social cues used by human characters in the scene be as effective as other artificial cues on guiding the viewer's attention?
  - Will the use of human characters' social cues, on the other hand, reduce the risk of breaking presence, compared to existing artificial cues?
- RQ_2: When the viewer is acknowledged as not fully passive but rather as requiring a participatory role in the CVR experience, what design considerations need to be added?

  - What level of control should the viewer be exposed to so they are aware of their agency but not too distracted or overwhelmed by the interaction task?
  - What is the proper type of interaction (mechanism, complexity, input methods) for a CVR viewer?

- RQ_3: Given that viewers all carry different personal attributes in terms of engaging with the content, how can a creator appropriately ensure a personalized CVR experience, especially in the cultural heritage sector?

  - What individual viewer attributes and variation strategies will be used to shape the cultural storytelling experience?
  - Does a digital version of personalized storytelling give viewers a similar impression of storytelling as a physical, face-to-face one?
  - Can personalized CVR be used as an appropriate tool by indigenous communities for storytelling?

1.3 Findings and Contributions

This thesis presents my work around creating compelling storytelling with CVR. It primarily focuses on discovering and expanding the toolkits for creators to capture and create content dedicated for CVR and embraces viewer interaction as an intrinsic and essential feature of CVR. The Adaptive Playback Control (APC) framework that I have developed combines my findings and insights, highlighting considerations around the viewer's role, the balance between viewer action and storyworld possibilities, and personalization according to individual viewer traits. In particular, my work makes the following contributions to the field:

- A collection of guidelines for CVR creators, pointing towards different aspects in the CVR production workflow, especially when seeing CVR as a type of viewer-centered interactive experience distinct from cinema and theater. It includes:

  - A content creation guideline for CVR creators to capture immersive content, including *triangular rules* for story element setup on scene, and then Action Units (AUs) as attention-guidance techniques to utilize human actor performances instead of complex post-processing work. Compared to other capture techniques, this proposal requires no technical knowledge background or complex work, and thus can be easily used by a wide range of creators and integrated directly into their workflows;

- A structured design guideline for creators to bring viewer interaction into the CVR experience. Unlike simple omnidirectional videos or fully interactive video games, the interactive CVR guideline system stands on the balance point by matching the viewer's expectation of agency with their interaction technique provided by the CVR system and the viewer's depth of participation into the actual narrative progression. This applies in particular to points like the viewer's role permeating the author's realm and the visibility of interactors in CVR. It helps creators prepare a story knowing they are aiming for interactive, immersive media from the earliest design stage.

• Second is the APC framework, consolidating all discoveries and proposals from the research and presented to CVR content creators as a comprehensive guideline for creating an effective storytelling experience. The APC provides support for both CVR creation with viewer participation in mind and viewer interaction design, by involving the viewer as a character in the storytelling scene and also in the design of the narrative progression mechanism. It covers the full workflow from content capturing to system implementation and presentation. The APC's exploration of the roles and control levels of both the creator and the viewer applies to any immersive media where the viewer occupies centre stage, regardless of interaction methods and levels. The project explored design decisions about the system's participatory capability provided to the viewer, taking into account how the system can give the viewer a vital role in the narrative, even incorporating some of the viewer's personal traits to personalize the results. This interplay between the viewer and the storytelling system also applies to other types of experiences where a story is contained and told.

This thesis also presents a case study, where the APC was used in a culture-oriented storytelling project conducted with Māori research partners and *kaikōrero* (Māori storyteller/s). The case study had two phases. Phase one was a co-design workshop where Māori research partners and I worked together around the technology (CVR) and the rules (APC) to create a storytelling experience. Phase two included capturing content at the Māori partners' *marae*[3] in Bluff, New Zealand (Figure 1.3) and evaluating the storytelling experience with the kaikōrero at the same marae. The case study combined the guidelines from APC with *tikanga Māori* (Māori customs and traditions) in oral storytelling. Based on APC, the viewer's role in CVR storytelling was further emphasized. In this case, viewers were treated as individuals holding different attributes and applied narrative variation to the content in response to those attributes. Cultural appropriateness was considered as an important factor by the creators when producing the experience, ensuring the digital version remained in line with its original physical form. By applying APC in an actual storytelling project, practices including how research can be conducted with a focus both on storytelling experiences and the preservation of cultural heritage, led by indigenous people, are also demonstrated. The entire process observed *tikanga Māori* to guide us towards what the technology represents and the appropriate factors to address and deliver the desired knowledge. Instead of serving as a checklist for collaborating with indigenous people, this research project shares the process that I, as a Pākehā (a Māori term to refer to non-Māori people), adopting western-view-oriented storytelling guidelines (APC) and digital tools to create tikanga Māori–guided storytelling experiences. Evaluation results indicated that a CVR experience with personalization designed into it and APC guidelines can be a viable tool for storytelling and sharing of knowledge in the context of Māori culture. The use of personalization when adopting APC in the cultural realm proved effective. Further, during the evaluation, additional awareness of the *kaikōrero*'s thoughts on the purpose of using digital tools and digital experiences in their *marae* setting was also raised. The evaluation also served as a review of and reflection on employing *kaupapa Māori* (placing Māori practices and values at the centre of the process), thereby conferring agency and considering what it means to the Māori community when creating content and experiences.

[3]In Māori culture, the **marae** is a gathering place where formal greetings, important events, and discussions take place.

FIGURE 1.3: An aerial shot of Te Rau Aroha Marae in Bluff, New Zealand, showing its *wharenui* (literally “big house”, with its carved entranceway), *wharekai* (dining house, left) and the complex of buildings around it. Source: Te Rau Aroha Marae Facebook page.

The other important discovery was that the use of AUs for attention guidance in CVR is independent of the work’s narrative structure and storytelling style. Such independence was observed during the content production phase for the user studies 1 and 2, presented in Chapters 3 and 4, and the co-design project in Chapter 6.

Another attention-worthy item for content creators is how universally applicable the findings of this thesis are to all types of storytelling styles. In this thesis, all content was produced in a guided-tour/documentary style, with a storyteller always visible in the scene. This visibility enhanced the AUs’ effectiveness. However, the design of the story elements, e.g., their temporal-spatial configuration, visibility, and priorities, including the storyteller, were unique to each CVR experience. When a creator chooses to apply AUs and the APC guidelines for viewer interaction in their work, there is the possibility that their capacity to create an engaging story and deliver an enjoyable narrative will be impeded. While the details and effects are not yet fully explored, I would like to make the case for consideration of CVR by creators when employing the APC and other similar guidelines.

1.4 Overview of the Structure of This Thesis

This thesis starts with a general introduction, including relevant background information, the state of the art of the research field and design space, my motivations for doing this research, and the main research questions. In Chapter 2 it then examines the background to the research topics and existing literature, with a summary of previous research and what is still missing. All relevant background information is discussed in this chapter to make clear the scope of my work and the structured steps I employed. Following this, Chapters 3 and 4 present the two user studies aiming to answer the research questions.

The user study presented in Chapter 3 addresses RQ_1 by exploring the use of human actors performing as diegetic[4] attention-guidance cues in the simple form of immersive media, the 360-degree video. Results suggested that the guiding cues (AUs) can effectively guide the viewer's attention to key story elements in the scene while maintaining low cybersickness, and an acceptable level of engagement and enjoyment compared to commonly used synthetic cues such as a pointing arrow or scene rotation.

Chapter 4 aims to answer RQ_2 by examining two fundamental questions when viewer interaction is involved in CVR storytelling and when that interaction is linked to narrative progression: *the viewer's role in the control of the narrative progression*, and *the visibility of interaction affordances*. Although no significant quantitative effects were found, a conclusion was drawn that creators must first acknowledge that viewer interaction comes as a natural tendency in CVR; both the level of control given to the viewer and the visibility of the interactors depend on the content and its purpose, derived from observations and qualitative measures. Note that in Chapters 3 and 4, descriptions of background knowledge are simplified, having already been extensively explained in Chapter 2. Starting from insight gained from each user study, Chapter 5, covers a higher level of understanding about CVR storytelling, viewer experience, and appropriate production guidelines by gathering the knowledge from the three studies into a more comprehensive framework, which I have termed as the Adaptive Playback Control framework. This decision was made as the issues are closely related and more appropriately combined into a single chapter rather than scattered across various parts of the thesis.

After the concept of the APC framework is advanced, a case study is presented in Chapter 6. The case study introduces a real-world use of the APC framework,

[4]Diegetic is a Greek word, originating in theater, meaning "being part of the story or the narrative"

applying it to a culture-oriented storytelling project created with Māori storyteller partners in Bluff, New Zealand. The project aims to enable a physical-world storytelling experience using CVR as its chosen digital tool. In this chapter the workshops conducted with Māori partners around building culturally appropriate immersive experiences are first described. The process of bringing APC into the storytelling project is also explained. Several insights into using APC in real-world projects learned from this case study are discussed after a description of the workshops and evaluation tasks.

The discussion about the application of this APC, not only to CVR but also to other related immersive media, the limitations of my study, and my future plans are covered in Chapter 7, the final chapter of this thesis.

Chapter 2

Background

To explore effective storytelling techniques for CVR, well-developed storytelling media such as cinema, theater, and video games cannot be overlooked. This chapter provides a background to storytelling in CVR and reviews previous work in that area. It then moves to early explorations in storytelling using VR, and finally the emergence of CVR. The migration of skills employed in cinema and theater is evident in early VR attempts. Practices used in the well-established media provided guidelines for creators in producing new content and experimenting with innovative skills and technologies. Later, creators tried to enable viewers to be active participants in immersive storytelling rather than simply invisible and passive spectators. New types of media have emerged with this shift in the viewer's role, such as immersive theater [4], improvisational theater [133], interactive TV drama [155], and gamified cinematic narrative [156]. The way the story is constructed, the logic of the presentation, the way the content is expressed, and the rhythm have all changed accordingly as well. Absolute authorial control has also started to diminish, with part of the role inclining toward the viewer's hand. Additionally, the idea of seeing viewers as individual people instead of a single-form archetype has come into play, and thus the concept of personalized storytelling. However, it should be noted that while the experimental media act as testbeds for creators to explore uncharted realms such as viewer participation and interactions in the storytelling activity, the well-known and well-established expressive media forms–cinema, theater, and video games–remain the cornerstone and the starting point for creators in their experimentation with novel storytelling approaches. Thus this chapter begins by introducing the three cornerstone media, then moves to various stages, concepts, and representative examples of creator exploration. The final section of this chapter concludes with a summary of the state of the art, setting up the main theme of this work.

2.1 Storytelling and the Well-Developed Storytelling Media

Storytelling is an essential part of our daily life. The everyday stories people tell each other reconstruct their experiences in narrative form. Stories are fine-tuned to become the content of works in theater, literature, and film [34]. The activity of storytelling is defined as a means of expressing experiences, emotions, and ideas in different forms of transfer [41].

Researchers who study the activity of storytelling primarily focus on how the content is structured and how it provokes an intended emotional response and cognitive change in the audience. They look at various types of media used to tell stories, such as words, images, and motion pictures, and now also VR [164]. Since this research focuses on storytelling using VR as the medium, this background introduction begins with three storytelling forms people generally know and experience: cinema, theater, and video games.

2.1.1 Cinema: One-Way Delivery of Content

In filmmaking, directors rely on a series of cinematic techniques to guide the viewers's attention, invoke suspension of disbelief and curiosity, deliver expressive content, and effectively tell a story. Among the main techniques, *mise-en-scène* and *cinematography* are two that occur in the production phase [17]. Mise-en-scène, a French term meaning "putting content into the scene", is how the film director controls what is happening in the screen frame, while cinematography is the director's decision about how the camera will capture the scene when shooting, including frame composition and camera movement. These comprise the screen grammar used by the director and cinematographer to deliver the narrative and provoke the audience's emotional attachment to the region of interest (ROI) [94]. In some scenes, the director will decide to capture the story with a moving camera, defined as mobile framing, another screen language set piece. Commonly used mobile framing techniques are the camera pan, dolly move, tilt, crane, and zoom, which may be combined [3]. They extend the space beyond the frame, show more content, and sometimes mimic the cue of a character's first-person POV, so that the viewer is now looking out to the world through that character's eyes.

2.1.2 Theater: The Removal of the Frame

In theater, the cinema's screen and frame no longer exist. Additionally, in theater whole segments of a play are run without breaking the chronological sequence in most cases, while some (mainly modernist) employ flashbacks or flashforwards (jumps back or forward in the timeline) for narrative purposes [4]. In theater, some filmmaking methods, such as cuts and edits, are no longer available. Thus, while the film creator renders directly on the screen what they want the audiences to see, the theater playwright and creator use other ways to draw audiences' attention to the vital elements [94]. In a standard theater practice, the viewer's attention is usually controlled by lighting, stage setup, sound cues, and exaggerated moments of action [82], while the actor's body language, such as an open body position, prolonged orientation toward the audience, and exaggerated body movements and gestures, become critical components [40].

Besides the stage elements, the other factor a theater creator considers is the location of the viewers. Cohen lists three widely used configurations of the theater stage in terms of the viewers' location vis-à-vis the actors: the proscenium, thrust, and arena [31]. This change in the actor–audience configuration has inspired researchers such as Cho et al. [29] to explore how this can be applied to viewer engagement in CVR stories, for example by manipulating the viewer's location in the scene from a first-person perspective (where a story character directly addresses the viewer, who thus becomes a part of the plot line) to a third-person perspective (where the viewer acts purely as an observer).

2.1.3 Video Games: Storytelling Through Playing

In cinemas and theaters, the viewers sit in front of the screen or the stage. The director tends to deliver the story to the viewers via a sensory experience including visual and aural stimuli [20]. Video games differ from these methods as they employ interactivity [75]. The role-playing game (RPG), as an example, first moves the user's perspective from that of a passive viewer to that of a player who acts as a character and sees the world through their eyes. The game then enables the player to control the protagonist, hear and see what his/her avatar experiences, interact with the game world via that avatar, and take on the protagonist's motivations and emotions. Instead of communicating narrative and immersion via visuals and sounds, video games do so via interactivity [49]. The term *agency* [98] is used to describe the process whereby a game player interacts with a character or object in the scene and reacts as a consequence of that interaction. In the process of interacting and having agency, the player immerses themselves in the game's virtual world and picks up the story elements hidden within the various game elements, such as dialogue with the characters, their decisions and motivations, events, and the development of player–character relationships [126, 75]. Via agency the game player is deeply involved in the emergence and progression of events in the virtual world. Cinema and theater audiences, on the other hand, sit separate to the events taking place and have no participation in their progression.

Storytelling via interactivity in video games, however, can also be fragile [87]. Since the game designer cannot predict specific actions each individual player will take at a given moment in the progress of gameplay, he/she cannot fully ensure that the narrative will be presented in the designed way since the player's behavior may render it incomplete or incomprehensible. Many game designers chose to embed film-like cut scenes into their games to deliver critical narratives. However, as players lose their ability to interact during these scenes, they lose their agency. Because the cut scene brings about a broken presentation of the game world, they feel a diminished sense of immersion and experience the game as having been interrupted [76].

2.2 Immersive Storytelling using Virtual Reality

When VR emerged into the public eye around 2016, VR headsets and consumer-targeted 360-degree cameras also became more widely available. Creators see 360-degree video as a simple version of VR, and soon adopted it as an immersive storytelling medium [21]. In that period, storytelling projects, experiments, and explorations mainly used 360-degree video as their mainstream form of content [18]. This section begins with 360-degree video, then introduces a more advanced form of VR for storytelling, the cinematic virtual reality (CVR). Issues raised by CVR, such as the narrative paradox (NP) encountered by creators during production and solutions that have been proposed to address it, will also be reviewed in this section.

2.2.1 VR as a Storytelling Medium

Content creators including filmmakers, TV producers, and game designers began looking at methods they could apply to tell engaging and enjoyable stories in VR taking into account its capability of immersing the viewer in the scene. Initially they naturally turned to filmmaking techniques and skills, attempting to migrate tried-and-true methods to storytelling using VR (mainly 360-degree video). Compared to cinema, the change immediately noticeable in the VR experience is in relation to point of view (POV). In a VR experience, such as watching a 360-degree video, instead of looking at a rectangular flat screen the viewer is inside the scene, surrounded by the images, (usually watching via a head-mounted display (HMD), as shown in Figure 2.1). Syrett et al. [146] stated that with this new POV the viewer becomes the narrator and now has control over what to see and what to understand. In contrast, in traditional cinema, the creator has complete authorial control. The change in the camera system also contributed to the necessity of developing new storytelling techniques. Unlike the cinema camera used in traditional filmmaking, which a lot of cinematography skills and screen grammar are a function of, the cameras used to film 360-degree video (for VR) are built with two or more fixed focal length (no zooming) wide-angle lenses to cover the entire scene (no panning) (see Figure 2.1 for details). Thus commonly used cinematic terms such as *pull-in* to focus in on an object or *pan* to reveal an object outside the frame space are no longer relevant in this new scenario [27]. In short, due to the change in the POV of the viewer when watching, the specificity of 360-degree camera features, and the disappearance of the screen frame, many of the grammar elements of cinema, such as various types of shots, camera movement, or time-compression editing, are not applicable in VR content [37].

FIGURE 2.1: One of the 360-degree cameras available on the market (left); a viewer watching a 360-degree video (right).

Prototype projects appeared during this stage as creators and practitioners tried to embrace these changes. Oculus Story Studio published several short omnidirectional videos (I avoid the term *VR film* as VR and film are two independent storytelling media), including *Henry* [105] and *Lost* [106], experimenting with new storytelling methods in VR (for screenshots of these see Figure 2.2). In both of these examples, the creator designed the characters (a hedgehog and a robot arm, respectively; see Figure 2.2a, b) to explicitly acknowledge the existence of the viewer in the story world, as the viewer is not immersed in the scene, which is uncommon in cinema (when used in cinema, it is referred to as “breaking the fourth wall”, as the actor is directly addressing the audience from the other side of the screen [17]).

Another example is the 360-degree video *Dyskinetic* by Oscar Bastiaens [9]. In this video, the first-person view is applied as the camera view (the viewer plays as one of the major characters). The viewer sits behind the protagonist’s eye, experiencing the protagonist’s paralysis, speechlessness, and inability to move, while seeing other characters talking and interacting, expressing their opinions (for a screenshot see Figure 2.2d). In the 360-degree video *Never Bout Us* by Cameron Grey [23], the creator was aware of the differences between storytelling with a camera for VR and traditional cinema. He borrowed techniques from theater, such as having the actors move dynamically toward the viewer (acting towards the camera and moving in terms of the camera). In this video, the story starts with the main focus in front of the viewer, with every other direction darkened to further enhance the front section as the primary focus zone. The memory phantom ghosts then played first to the left to catch the viewer’s attention, followed by the protagonist also moving to that location to complete the shift in the viewer’s main focus, as shown in Figure 2.3. This two-step shift was done multiple times to gradually draw the viewer’s attention towards the left side instead of panning the camera in that direction.

The 360-degree video *HELP* from the Google Spotlight Stories studio [79] made an attempt in a different direction. While also acknowledging the viewer as a character in the centre POV, the creator placed several key characters (a policeman, a pedestrian, a monster) around the camera (i.e., the viewer). They then handed the freedom of choosing where to look completely over to the viewer, giving away authorial control (as shown in Figure 2.2c). In this video, the viewer can freely follow one of the characters throughout the story or frequently jump between the characters but still maintain a rough understanding of the plot. Eventually, the viewer will need to re-watch the video several times if s/he wants to grasp all the details.

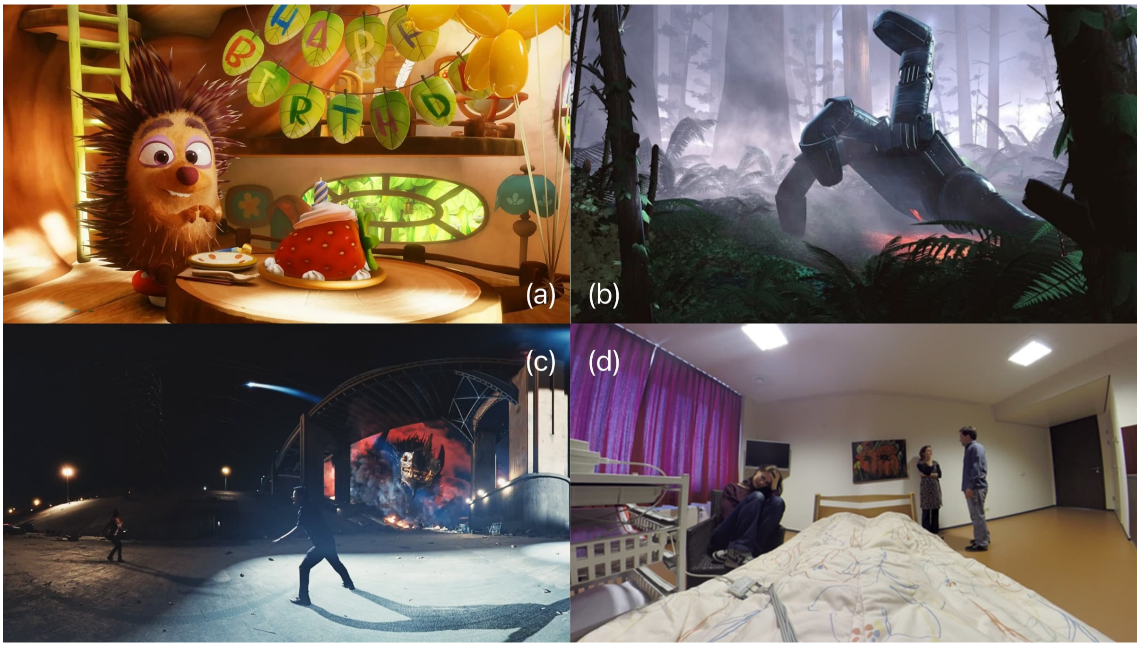


FIGURE 2.2: Screenshots from VR artworks: (a) *Henry* [105]; (b) *LOST* [106]; (c) *HELP* [79]; (d) *Dyskinetic* [9].
(a) and (b) are examples of the protagonist in the scene acknowledging the existence of the viewer; (c) shows multiple events and characters placed in the scene simultaneously (the pedestrian on the left, the policeman in the middle, the alien on the right); (d) shows the viewer taking the first-person view of a patient on a bed.

An updated cinematography language was also proposed to help VR content makers gain confidence in incorporating compelling storytelling within this new medium while resolving the FOMO issue for a lengthy immersive viewing experience [28, 93]. However, in the works stated above, creators migrated the storytelling grammar and toolkit of filmmaking to immersive storytelling in 360-degree video. The mindset behind the migration implies that they still see VR as a form of cinema, only differing in the spatial configuration of the elements of the story. Evidence includes the absence of the frame and the shift in the viewer's location, from the seated bystander of cinema to the first-person view in the scene in a VR experience. The change in the viewer's role in the narrative and the intrinsic characteristics of the immersive environment were much less considered.

FIGURE 2.3: Screenshots of the 360-degree video "Never Bout Us" [23]. The images are taken chronologically from the video, showing how the "memory phantom ghosts" (shown in red boxes) and the protagonist (shown in green boxes) alternately moved towards the left of the view, from a) through f), gradually directing the viewer's attention to the exit on the left side. The changing location of the bookshelf from a) to f) indicates the shift in the viewer's POV towards the left.

Researchers have found that when moving from flat 2D video to an immersive medium, not surprisingly, the most noticeable change is in terms of point of view (POV). In an immersive medium, the viewer sits in the centre of the scene instead of looking at a rectangular flat screen. Syrett et al. [146] stated that with this new POV, the viewer becomes the narrator since s/he can choose what to look at and understand. The new POV represents a change in the viewer's role. One commonly used method to cope with this change is to define the viewer as either a spectator (not part of the story) or a character (either invisible or acknowledged) in the story. Bender [12] compared the effects of two camera positions in VR, the character view and the immersive passive view, considering their effects on attention to essential spots. The character's first-person view did help the viewer to establish a fixation on the ROI faster than a third-person view.

In her research, Dooley [38] also proposed a less coarse division. She categorized the roles of a character as those of silent witness, participant, and protagonist. The difference between participant and protagonist is determined by whether other characters in the story give the viewer social acknowledgment. It also relates to how often and with what priority this is given. Furthermore, acknowledging the viewer as a person or a character is a VR approach borrowed from theater practitioners. In staging and performing 360-degree video using principles from theater, Pope et al. [116] found that actors had lines directly spoken to the viewer/camera, treating it as another person in the scene. The viewer were treated as an invisible ghost by the actors, since in 360-degree video, the viewer cannot see his/her own body when looking down. Brewster [18] also stated that the viewers could alternatively experience a "morphing" of identity, where they find themselves being something in the story or the scene other than a human being. One example is the 360-degree film *Miyubi* from Oculus [46]. In the film, the viewer finds that s/he has become a toy robot newly acquired by a child and witnesses how other family members treat the child. Figure 2.4 shows a screenshot from the film, with the toy robot's arms visible in the lower part.

FIGURE 2.4: A screenshot from the 360-degree film *Miyubi* [46] (source: Oculus.com). The viewer is embodied as a toy robot, with the robot arms visible if the viewer looks down. The image is slightly distorted because of the conversion from 360-degree video to a flat image.

Besides defining the viewer's role from a screenwriting or theater staging perspective, the pose a viewer takes to watch the narrative content will also influence how his/her role is perceived, thus affecting his/her behavior when watching. Gödde et al. [57] observed viewer behavior using the same content, but in a seated versus a standing pose. They point out that when seated (the most common way to view 360-degree videos), viewers spent more time looking towards the front and showed less exploratory behavior. Tong et al. [153] also explored the preferred user scenario for 360-degree video and coined the term *swivel-chair VR* to specify the preferred way to consume 360-degree video content (seated rather than standing). Swivel-chair VR describes a scenario where the viewer watches a 360-degree video while sitting in a swivel chair, wearing a VR headset. The swivel chair allows the viewer to rotate around 360 degrees by turning the body together with the chair. It is easier and more comfortable than only turning with one's neck, as the swivel chair serves both as a cue to imply the affordance of rotation and an anchor point to assist rotation. It also provides greater comfort and safety compared to standing while viewing.

Mise-en-scène is a powerful tool for filmmakers to design story elements to help the viewer discover where they are located and what to focus on in the scene. This was also explored in the context of VR for storytelling. Dividing zones around the viewer was proposed because viewers were free to look around and decide where the current view and focus would be in an immersive scene. However, researchers discovered that not all directions around a viewer had equal weight, given a choice. Gödde et al. [57] divided the entire 360-degree area around the viewer into the front area, the rear area, and a blind spot directly behind, as shown in Figure 2.5. They point out that viewers mainly focus only on elements within the front zone (180-degree front-facing) and are less likely to look for elements within the rear zone, where significant head turning is needed. The blind spot is where elements will most likely be ignored or missed by the viewer, even though they may be relevant to the narrative. With proper staging and directing cues, viewers can be encouraged to explore the rear zone. However, the placement of essential elements in the front zone is still recommended, as their study demonstrated. They also found that if the viewer takes a seated pose, the preference for the front will be intensified.

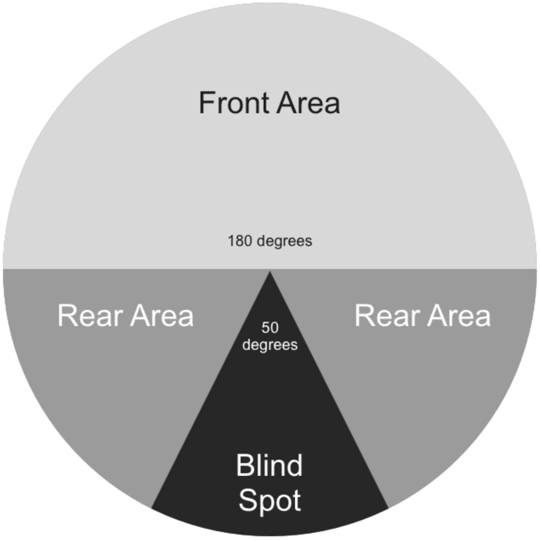


FIGURE 2.5: Gödde et al. [57]'s division of staging zones in the full 360-degree area around the viewer. Note the a blind spot of about 50 degrees directly behind the viewer.

Distance is another crucial factor to be taken into consideration. Dooley [38] pointed out that in traditional filmmaking, within the border of a frame, directors can use various types of shots to control what story element to focus on, such as close-ups or wide view, to direct the viewer's gaze and attention. Likewise, in VR storytelling, while there is no frame to define a shot, defining "experience" based on "distance between the viewers and actors" can be essential and viable. Dooley proposed a theory that "in an immersive scene the distance can influence the viewer's emotional engagement with the characters" and analyzed three scenes in a sample narrative 360-degree video, *Dinner Party* [88], to see if distance variations changed how viewers perceive their relationships with the characters around them. In their study, Pope et al. [116] asked actors to stage and perform a short drama for either a viewer sitting on a swivel chair or a 360-degree camera taking the place of the viewer, as shown in Figure 2.6. They discovered that when the camera replaced the actual viewer, actors still performed based on principles regularly used in theater. The actors tend to group on one side of the camera so the viewer can easily see all the action without having to turn their head frequently. In another test, Bailenson et al. [6] found that in an immersive environment, users tend to maintain their personal space and keep their distance from other human characters, just as they do in real life, and that if other human characters make eye contact with the viewer, this tendency of distance-keeping will be more obvious. It should be noted that their work was not conducted on 360-degree video but rather mainly in a highly interactive setting where the viewer could freely move around.

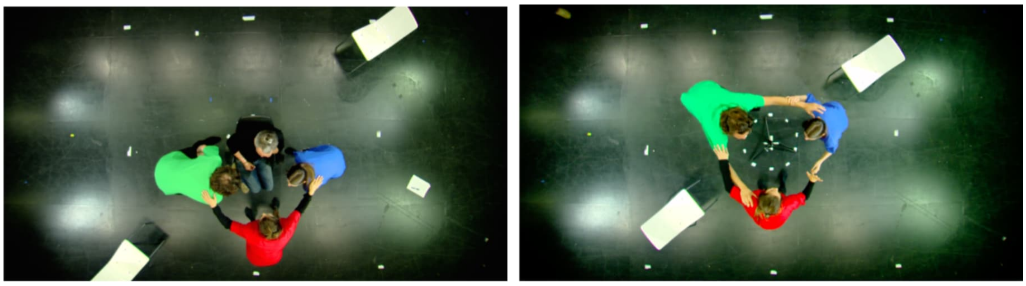


FIGURE 2.6: Two top-down shots from the study conducted by Pope et al. [116]. A group of three actors "huddle" around a real viewer (left) and around a 360-degree camera (right). Source: Pope et al. [116]

.

Besides controlling scene elements, directors have tried to manipulate the camera's parameters directly to affect how the viewer perceives his/her role. One example is the camera's height. Researchers discovered that the viewers are more accepting of vertical differences between the camera and the eyes if the camera position is lower than the viewer's eye height [129]. Another example is the viewer's pose. A seated pose is preferred for watching 360-degree videos and can be easily adapted compared to a standing pose [69].

2.2.2 Cinematic VR and Narrative Paradox

As 360-degree video has become widely popular, content creators are now trying to produce more engaging narratives using this immersive medium. The term CVR is used to describe such VR experiences. CVR creators now aim to create lengthier content with a complete narrative arc (similar to the drama or documentary commonly seen in cinema) and to make viewers feel immersed in the story world, rather than short and simple footage for brief excitement [39]. The CVR viewer watches pre-recorded or pre-rendered content in omnidirectional (360-degree) video format, using an HMD [93] so s/he can freely look around and change viewpoint [93]. It should be noted that the definition of CVR is not definite but still evolving. The physical viewing experience with CVR, especially the viewer interaction element, can vary from simple passive immersive video, where the viewer only has the freedom to look around, to a computer-aided system that allows the viewer to interact with objects and characters in the scene, and even alter the narrative progression [158, 126, 43]. In the production and distribution of CVR content to a broader audience, one of the significant issues creators encounter is the narrative paradox (NP), which is the tension between the viewer having freedom of choice and customizability and the creator controlling how the narrative plays out [161]. The NP poses a challenge for creators in terms of balancing engaging interactivity with dramatic progression. The second major issue linked to viewer control is that it can cause the viewer to miss essential story elements, inducing a fear of missing out (FOMO) [86] and yielding weak narrative comprehension and low emotional engagement.

Because researchers initially sought solutions to these issues by migrating filmmaking techniques to CVR production, terms from the film director's vocabulary are also brought into this realm, namely, mise-en-scène, cinematography, and editing [17], to group the solutions proposed in previous works. Sound is left out as it is outside the scope of this research.

In terms of mise-en-scène, creators tried to maintain both narrative control and the feeling of immersion by changing the viewer's viewpoint in the scene (e.g., through camera placement), by rearranging the placement of the character's action, and through other story elements [116, 129, 38]. CVR practitioners also borrowed skills from theater when they found similarities between CVR and theater, such as increased viewer freedom and continuous playback of the content [4]. They used on-stage elements as cues to attract the viewer's attention, such as lighting, shading, and stage setup with depth changes [82]. From the cinematographic perspective, several researchers discussed how the spatial-temporal density of a story, framing grammar, and editing techniques apply to CVR [57, 39, 71, 45], as well as going beyond this to explore methods to directly guide the viewer's attention to essential story elements while not taking away their freedom to browse around. In terms of editing, researchers and creators reached less solid conclusions compared to the other two. Fearghail et al. [45] and Serrano et al. [136] both looked at 360-degree video editing and came to the conclusion that jump cutting is applicable in CVR but not as frequent and effective in cinema. Relevantly, pauses and transitions were also migrated to CVR, with modifications [71]. Prolonged pauses are preferred in scene changes in CVR because viewers need extra time to explore the scene and understand where they are after the jump. Further discussion on this can be found in section 4.6.1.

One observation is that these solutions all focused on the creator's role, as opposed to viewer agency. Even the terms used to group them are from the film director toolkit. Although the viewer's new role as being *inside* the scene and part of the storytelling activity in CVR is acknowledged (see section 2.2.1), little attention has been given to this aspect of CVR, especially the viewer's role and agency through interaction. One assumption is that creators are most familiar with mature filmmaking techniques used in the industry and choose to migrate these into the newer medium (a direct example is how well-established terms from filmmaking have been borrowed to define CVR rule perspectives). In traditional filmmaking, viewer agency is seldom considered because, in a cinema, the viewers are passively sitting and looking straight at the screen, with zero interaction with the content and no influence on the story. Thus, CVR creators take this as the given viewer scenario [17] and tend to ignore viewer agency when migrating techniques to CVR. However, the viewing experience of swivel-chair CVR is quite different from traditional cinema. Firstly, although both are normally experienced while seated, a CVR viewer can turn his/her head or chair to look around [153, 55]. This freedom of head rotation means the viewer now controls where to look in the scene, changing their perception of their role and extending their interactivity and agency [146]. Secondly, a CVR viewer is not a fully active participant, because CVR is a "lean-back medium" with limited interaction possibilities [130]. The viewer is still passive and mainly wants to watch the story unfold and follow the narrative instead of acting in it [57].

Discussions focused on the properties of interaction design from the viewer experience perspective (such as the viewer's awareness of affordances for interaction and expectations of agency) were also rarely found in the literature. The typical discussion has been around how to enable full viewer interactivity in immersive environments [150, 115]. However, full interactivity would require complex hardware and an overly demanding amount of effort on the viewer's part, and thus does not apply to CVR. This amount of interactivity borders on something other than cinema and tends more towards the immersive video game (see section 2.1.3). Also, in many games the player controls an avatar, which is a character in the story. Representation of the viewer him/herself via an avatar does not exist in CVR [39]. Thus CVR viewers and video-game players are considered to have different primary motivations and as such are likely to comprise different demographics. Design considerations around CVR viewers will be further discussed in the next section.

2.2.3 Guiding the Viewer's Attention in CVR

Directors rely on a series of cinematic techniques to guide the viewer's attention, invoke curiosity and a suspension of disbelief, deliver expressive content, and effectively tell a story [31, 94]. Similarly, researchers have been using cues in the immersive scene to direct viewer attention to ROIs, while maintaining the viewer's freedom to explore the scene. Elements from video games including virtual arrows [22, 50, 54, 114], signs, markers [162], and audio [99, 110] are imported into the scene as cues to direct viewer attention to the expected ROIs (several screenshots are provided as examples in Figure 2.7). However, researchers have also discovered drawbacks. Nielsen et al. [99] pointed out that when a cue was applied, such as an arrow, it was presented in the scene as a non-diegetic object. This reduced the viewer's sense of presence since s/he was constantly looking at an item that was not contextually related to the environment. Lin et al. [80] used scene rotation as the cue to avoid non-diegetic intrusion. They rotated the rendered scene in real time to directly orient the viewer toward the ROI. However, the level of cybersickness spiked when the scene rotated. In addition, production-wise these cues are applied in the studio after the content is captured. Using these approaches, therefore, requires a large investment in time, resources, and skills in sophisticated post-production software.

Instead of using post-processing or non-diegetic cues, others looked into combining attention guidance into the production stage by learning from theater practices. Lighting control was first migrated from theater stages into virtual scenes as an attention-guidance technique [7, 16, 42]. Then Brown et al. [19] conducted an early trial that directly used human actor performance, in which they scripted the actions of human actors to make the viewer aware of specific events happening in the scene. Lindeman [81] also worked on providing a "lexicon of shots" for 360-degree video capturing, providing body language references to actors, allowing them to implement attention guidance while shooting the videos, as shown in Figure 2.8. These all sit at the border between mise-en-scène and cinematography, as some use diegetic story elements to mitigate the NP and FOMO issues to a certain extent. However, they still have shortcomings, such as breaking the feeling of presence as non-diegetic elements or lacking consideration of the viewer's expectation of participating in the story as a character.

FIGURE 2.7: Examples of pointing arrows as directional cues in the VE: a) pointing arrow [22]; b) GPS-style pointing arrow [54]; c) the Bubble Bee as a pointing and directional cue, pointing to the blue dot as a target [162]

FIGURE 2.8: Examples of shot types for a 360-degree video live stream developed by Lindeman [81]: a) the "look there" shot, a hand and head gesture where the presenter points and looks out toward a distant point of interest, then looks back at the camera; b) the "come along" shot, a camera movement that seems to beckon the viewer to move along with the presenter at a measured pace; c) the "look here" shot, a camera movement that brings the camera towards a very close, immovable ROI. The images are screenshots from 360-degree videos from the viewer's POV.

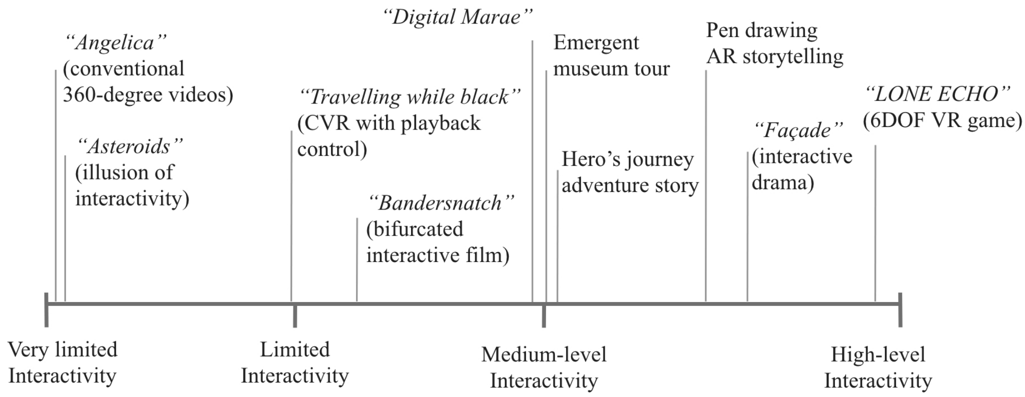
2.3 Viewer Interaction in Immersive Media

As stated in section 2.2.1, many researchers have worked on solutions for issues like NP and FOMO. Some have tried to maintain narrative control and deliver dramatic experiences by changing the viewer's viewpoint in the scene (e.g., camera placement), the placement of action, and the story elements [116, 129, 38]. Others have explored techniques to direct the viewer's attention within the immersive environment to essential story elements, including the use of diegetic story elements, visual cues, and alternations [128, 80, 161, 19]. However, these solutions all focused on the creator's role, not viewer agency. Many emphasize a workflow whereby CVR directors try to manipulate the viewer's focus, compose scripts to balance out spatial and temporal story density, and use framing and editing in both the production and post-production stages, among others. However, none of these put the viewer at the centre of creative decision-making.

As mentioned in the previous section, the viewer's role changes with the move from a flat screen to an immersive medium such as 360-degree video. The shift means the viewer is now part of the story world (as a character or not) and will expect more agency [93, 121]. It can be inferred that the CVR viewer will want to interact with elements in the story world and influence the narrative, similar to what a video game player would normally do (see section 2.1.3). But unlike with gameplay, viewer interaction in immersive storytelling implies a particular control the viewer has over the narrative [138]. This control differs from one system to another (either because of the system's hardware or the purpose of the storytelling). The following sections will first propose the idea of a *continuum of interactivity* to characterize and group experiences and projects by their levels of viewer interactivity, on a continuum ranging from very limited to limited, medium, and high, as shown in Figure 2.9. All groups will be discussed in terms of their interaction design decisions and properties.

2.3.1 Zero Interactivity

The most common and easily accessible form of CVR is the 360-degree video. Like in traditional filmmaking, the viewer takes a seated or standing pose, watching the video either on a flat-screen device or wearing an HMD [153]. In both cases, the only viewer input is the viewer's choice of the direction in which to look at any given time. In a traditional 360-degree video, this input will not affect any narrative progress, and the creator does not predefine any reactions to it [143, 80]. One can conclude that the viewer in 360-degree video playback has zero interactivity with the narrative. Such experiences are located on the left-most part of the continuum, as shown in Figure 2.9.



.

FIGURE 2.9: The continuum of interactivity. The horizontal axis represents the level of interactivity an experience offers, varying from very limited to highly interactive. Various immersive experiences with different levels of interactivity are placed along the continuum, not limited to CVR but also including VR games, interactive theater, and many others. The titles of publicly available experiences are explicitly labeled. The emergent museum tour [60], hero's journey [121], and pen-based AR [36] are described in sections 2.3.1 to 2.3.4.

2.3.2 Illusion of Interactivity

Other than addressing the directors and actors, some researchers also tried to embrace the viewer's first-person view using an "illusion of interactivity" approach, which helps to increase the level of spatial presence and realness, even when no "real" interaction is taking place (see [120]). Brewster from Baobab Studios [18] presented the method they applied in the computer-generated narrative VR short film *Asteroids* [145], where a robotic dog mirrors the viewer's simple head tilting during a dwelling stage before the main story starts. This reinforces the viewer's impression of being part of the scene and that their movement can affect the scene itself, even if general interactivity is not enabled and the experience is remains a linear film. Dooley [39] also points out that strategic VR directors can create the illusion of choice for the viewer when in fact what they are doing is creating a series of audio and visual cues that result in a preconceived narrative experience. This can be done by deliberately laying down a series of content chunks with auxiliary transitions and assisting content in filling in the blanks between important story nodes. Thus, the viewer will feel in control and that the story unfolds as a result of their noticing something remarkable or choosing to look at a specific object first. Yet, system wise, the viewer is not interacting with the environment, nor do they have any impact on how the narrative unfolds. Thus, these experiences also fall in the left-most part of the continuum.

2.3.3 Medium-Level Interactivity

CVR creators realized that in an immersive storytelling experience, the viewer will take the role of a character (at least that of a spectator) and expect to interact with elements in the story world. They also noticed that the fulfillment of viewer agency is closely linkede to the viewer's experience of CVR overall. Thus they began to enable actual interactors in the story scene instead of using illusory ones [130, 43, 53]. However, creators and researchers are facing new challenges because on the one hand, content has always been prerecorded or rendered in advance, and production guidelines are well-established around prerecorded materials (similar to what is widely adopted in the filmmaking industry), and on the other hand, viewers are still broadly seeing the experience as an activity of *watching* rather than one of dedicated *playing*. New challenges here include the tasks creators would like the viewers to carry out, the input viewers will employ in performing the tasks, the environmental and narrative changes the viewers' interaction introduces, and the different levels of interference viewers should have for a CVR experience. Creators and researchers point to different solutions depending on design motivation, context, and creator's purpose for each CVR experience. A detailed discussion of these is presented in section 2.4.1.

2.3.4 High-Level Interactivity

Simply browsing prerecorded clips by pointing and clicking is not the only type of immersive storytelling. On the opposite end of the continuum, researchers have also tried advancing narrative experiences by employing interaction techniques of high-level fidelity similar to those used in video games, discussed in section 2.1.3. Examples include emergent storytelling, interactive drama, and sandbox video games. Sharaha and Dweik [138] reviewed several interactive storytelling approaches with different input systems. These approaches have one aspect in common: the user is highly interactive since they act as one of the virtual characters in the story. In the system they designed, Cavazza et al. [25] focused on auto dialogue generation to drive the narrative based on text input provided by the viewer. A similar approach can also be found in the well-known video game *Façade* [92, 135], in which the story unfolds as the player interacts with two virtual characters by typing text into the console (see screenshots in Figure 2.10). Soares de Lima et al. [36] proposed another storytelling experience where the viewer watched the narrative via an augmented reality (AR) projection system and participated in it through drawing. The virtual characters were performing, and the viewer could physically draw objects on a tracked paper and then "transfer" them into the story scene as virtual objects to interact with the characters and the storyline.

FIGURE 2.10: Screenshots from the video game *Façade* [92] showing the two characters. The player interacts with them by typing text into the game window. The conversation is visible as text on the screen.

In another large-scale experience [26], the story unfolded as the viewer had natural conversations with another virtual character. The viewer was able to speak and move around in the scene freely. The system reacted to her behavior and progressed with the narrative. In these experiences, the viewers (or users) were given interactivity of a high fidelity (speech, gestures, locomotion, drawing). Because of the abundance of input and interaction techniques available to the viewer, these experiences are placed at the right end of the continuum, as shown in Figure 2.9.

2.3.5 Level of Interaction in CVR Storytelling

In the previous section, swivel-chair VR was mentioned as the preferred user scenario for CVR, where a passive viewer sits on a swivel chair (or, less likely, is standing). The viewer expects some agency in the storytelling but still wants to enjoy the narrative "leaning in" (involved in the storytelling, but not actively leading its progression) instead of "leaning forward" (actively participating in the storytelling activity and leading its progression via his/her own actions) [158]. Considering the placement of "CVR with interactivity" on the continuum (Figure 2.9), very limited interactivity is not considered as the desired option for implementing narrative VR. In many CVR experiences, the level of interactivity provided by the system does not match the viewer's expectation of participation when moving from a flat screen into this immersive medium. High-level interactivity is also not applicable because in CVR, especially the preferred swivel-chair VR, the viewer's capability of interaction, objectively, is still insufficient to perform full-body action or delicate operation on objects. Furthermore, subjectively, the viewer is still passive, unwilling to act with full interaction, but rather enjoying the story unfolding in front of him/her. Therefore, an initial inference can be made that a proper system design for CVR with interactivity will be placed somewhere in the "limited" or "medium" part of the continuum.

A storytelling framework presented by Reyes [121] fits well into this category. She presented possible diegetic interaction options on a pre-scripted story with different navigation alternatives. The aim was to produce an interactive narrative that is independent of the viewer's journey within the story; the plot is always created with a dramatic climax, "ensur[ing] the linear progression of the dramatic arc independently of the journey shaped by the viewer's choices" [121]. In her proposed system, Reyes put forward this idea of "limited interactivity", as the viewer has some level of free choice. However, the general narrative structure is based on the "hero's journey" and is controlled and made by the creator (similar to the double-diamond model in video game design [160]). The primary nodes are always defined (and will be a key element to drive the story forward), with secondary nodes of different paths for free choice. In their paths, two types of links were designed and implemented, where "external links" are jumps between pieces of the stories, moving alongside the general story arc, and "internal links" are extensions within a node, pointing to extra pieces of stories but not critical to the arc (providing more information to enrich the experience). A similar approach has been evaluated by Winters and Zhu [161]. They used the structural features of buildings and terrains in a 3D game world to implicitly guide the players along the storyteller's preferred path. They also placed prominent landmarks in the backdrop of the game world, representing key places where the main plot takes place, that the player, while freely roaming the world, would be attracted to and eventually choose to go to, thus driving the narrative forward.

It may in fact be inappropriate to represent all viewer interaction in CVR along a single-dimension continuum that groups them by "level" of interaction of a viewer during the experience. On the one hand, the level of viewer interaction varied from very limited to medium, high and abundant, determined by system design [128]. On the other, similar forms of interaction techniques also have a different impact on the story itself, such as simply controlling the playback progress on a one-dimentional timeline [108] or affecting how the story unfolds and ultimately changing the outcome, such as in *Bandersnatch* [140, 125]. Level of interactivity, that is, the complexity of the interaction technique, seems to not be strictly equivalent to depth of impact, as described in a previous section. It is possible that when an interactive storytelling system is implemented, its input methods and interaction techniques define its position on the continuum of interactivity. However, how the viewer will choose to participate and affect the narrative cannot be fully predicted because the actual use case varies from one viewer to another. This was discussed by researchers who looked at both technical and narrative aspects of immersive experiences separately [100]. Koenitz [73] also put forward a theory indicating that the viewer's participation in storytelling is also part of the narrative. Further exploration will be needed on the relationship between viewer interaction design choices and the actual viewer experience.

Furthermore, from the aforementioned interaction design case, it was acknowledged that there is no definite "better" interaction design for a satisfying viewer experience. One iteration of the digital marae experience from an ongoing research project can be a suitable example to demonstrate this point (see [109]). In one of the conditions in that experience, the viewer could spatially get "closer" with one of the ROIs and "further" from others. The system interpreted the viewer's choice by measuring which ROI s/he was closest to and presented the associated story. The viewer had a high level of narrative immersion because s/he felt the host knew where s/he was and what her focus was as s/he roamed the virtual environment. The viewer was participating in the progress of storytelling. The sequence of the content being delivered was determined by the viewer's activity. Instead of using pointing or clicking, the viewer participated via his/her *location at the moment*. However, this setup also significantly increased the system's complexity as the viewer needed input for teleportation, and the creator needed to cope with extra factors such as the viewer's distance and possible interruptions when the story about one specific ROI was being delivered. We can assume there is no "correct" choice between these two setups, and that it instead depends on the creator's choice of the purpose of the installation and what experience the viewers are expected to have.

Reviewing all types of interactions in storytelling experiences, at this stage, the following preliminary conclusions are those currently generated and acknowledged by CVR creators. These provide the common ground on which the research projects in this thesis are built.

1. A CVR viewer is a "lean-in" viewer. Creators are recommended to consider the viewer as a character in the scene. They should allow the viewers a certain

level of impact on the narrative, resulting in viewer agency, presence, and engagement. At the same time, the viewer's participation needs to be considered because his/her role is different from that in cinemas or video games;

2. The creator will need to consider a series of consequences when adding viewer interaction to CVR. Those include the tasks a viewer wants to carry out, the impact on the narrative each of these tasks will have, and the input method the viewer will use to perform these tasks. Especially in terms of input method, both explicit and implicit ones were employed in previous works. Based on conclusion one, since the viewer is only "leaning in" instead of fully willing to participate, creators are recommended to use implicit techniques in CVR;

3. Authorial control remains necessary to deliver a complete story. The creator still needs it to construct the story arc to evoke emotional engagement in the viewer. Internal links between the story elements need to remain intact despite the viewer's interactivity;

4. Whether a higher level of abundance and complexity of viewer interaction will necessarily lead to a higher level of narrative engagement and enjoyment remains unclear. Creators need to consider this in light of the general purpose of the storytelling activity itself.

2.4 CVR with Viewer Participation and Interactivity

From trials and tests with CVR storytelling involving viewer interaction, new questions arose for CVR content creators and researchers concerning viewer participation. While acknowledging the involvement of the viewer in the process of storytelling, they faced the following challenges:

- What part of the storytelling progress will the viewer be involved in and affect?
- To what extent will the viewer's input and interaction be incorporated into the storytelling itself?
- If the viewer's interaction will intervene with the storytelling process/narrative progression, what is her/his motivation and expectation?
- What changes are needed to the story structure itself to match the viewer's involvement?

- How can one create an acceptable combination of story structure and viewer motivation for interaction, and then get them to match?

Possible answers to these questions will be discussed in following sections.

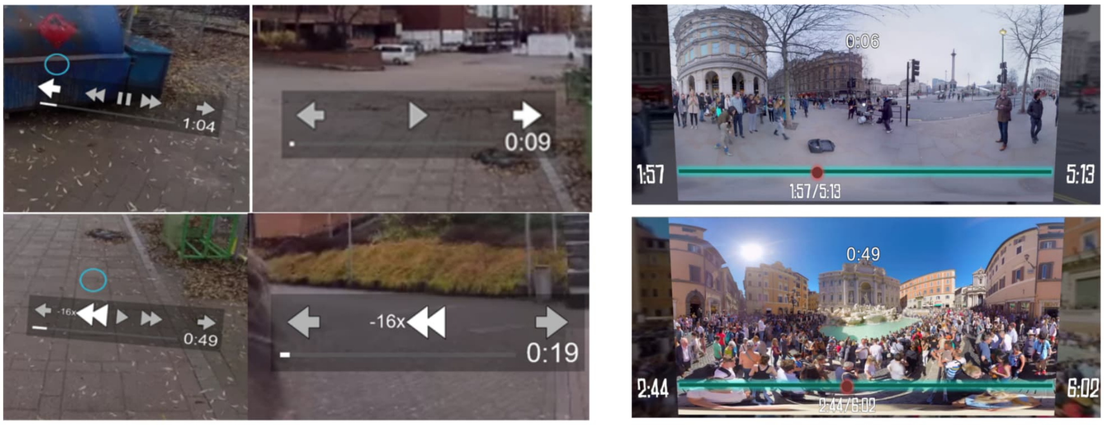
2.4.1 Viewer's Involvement the in Narrative

Acknowledging that viewers intend to interact, researchers have been exploring and evaluating various interactive techniques for CVR. Depending on the content, the genre, or the creator's specific intention, each approach registers its combination of choices in terms of several elements. Two of them are the main focus here: the interactive element's visibility and the depth of impact on the narrative.

Generally, when creating a CVR experience, the creator will need to go through three design decisions before eventually implementing the specific interaction technique: the *tasks* she wants the viewers to carry out, the *consequence/impact* each task will have, and the *input method* the viewers will use to perform them. Various interaction designs from example projects are reviewed in the following three sections, paying particular attention to their visibility to the viewer and their impact on the narrative.

Shallow Intrusion—Temporal Control. A typical viewer input in CVR is the temporal control, which means the viewer can only interact with the playback of a narrative (such as by changing the speed, pausing, or browsing), but is unable to break the linear progress. In a conventional 360-degree video player, the viewer is presented with a user interface (UI) in the form of a menu bar UI for interaction. Two examples, from Pakkanen et al. and Keijzer [108, 66], are shown in Figure 2.11. These researchers developed and tested various input techniques, but the control panel stayed similar. They were always derived from familiar desktop UIs. These UIs require eye and hand coordination to complete the "pointing-to-activate" two-step action, temporally taking away the viewer's capability of making spatial choices and looking at the scene itself. Petry and Huber [113] invented a new system that decoupled orientation control from temporal control. They kept free head rotation to control where to look and, simultaneously, enabled an extra pointing gesture for fast-forward/rewind. The viewers were therefore given parallel capabilities to look around and browse through events chronologically simultaneously. These two inputs also did not interfere with each other, unlike in the previous examples. However, they all kept the viewer at the level of controlling the temporal progress along a linear track; the storytelling itself was prerecorded and fixed.

Thus, the interaction was shallow and limited.



a) b)

FIGURE 2.11: Two examples of the menu bar UI used in CVR interaction: a) a UI with several buttons and a progress bar, from [108], where the viewer interacts with the buttons; b) a different UI from [66], where the viewer directly interacts with the progress bar to jump to a different point along the timeline.

Deeper Intrusion—Narrative Control. Creators are also aware that, in CVR, after acknowledging the viewer as a character in the scene, adding interaction enhances the viewer's feeling of presence because having an active role contributes to their enjoyment and engagement [43]. At the current stage, as most CVR content is prerecorded, the possibilities of interaction with the narrative itself are mainly limited to two: (1) choices over a bifurcated plot where every scene is a video clip, and (2) the overlapping of extra elements over each video clip, injected into the scene [121]. Interactive narrative content, also known as interactive fiction [96], is a form of narrative based on a bifurcated story and has been commonly available. Content can be found in both traditional flat-screen media, such as the sci-fi drama *Bandersnatch* from the series *Black Mirror* [140], and in immersive media, such as the virtual relic city tour *Bagan* [32] based on recorded 360-degree video and rendered 3D scenes. In interactive experiences, the viewers make choices at each "intersection" (story nodes in the design) and rearrange the linkage of fragments into the configuration of their choice [90]. One example is the previously mentioned drama *Bandersnatch*, in which the viewer is occasionally presented with two choices throughout the story, as shown in Figure 2.12. Each interaction within the experience is reactive, from a technological POV. Another type of narrative control strategy can be found in several on-site large-scale installations, where prerecorded stories are stored in scattered locations [15, 91, 101]. In these installations, each visitor goes through story segments in the order corresponding to how they arrived at each location or the route taken through the entire scene. However, this is challenging from a narratological/authorial POV, in terms of maintaining the flow of the story and viewer engagement with the narrative, because the story's creator cannot predict each viewer's actual chain of choices when the content is presented. Thus, authorial control over the narrative is disrupted. In these location-based narrative experiences, users did have control over the sequence of disclosure of the content but were still unable to change how stories precisely unfolded. In other words, they were not involved in the initiation and development of the narrative.

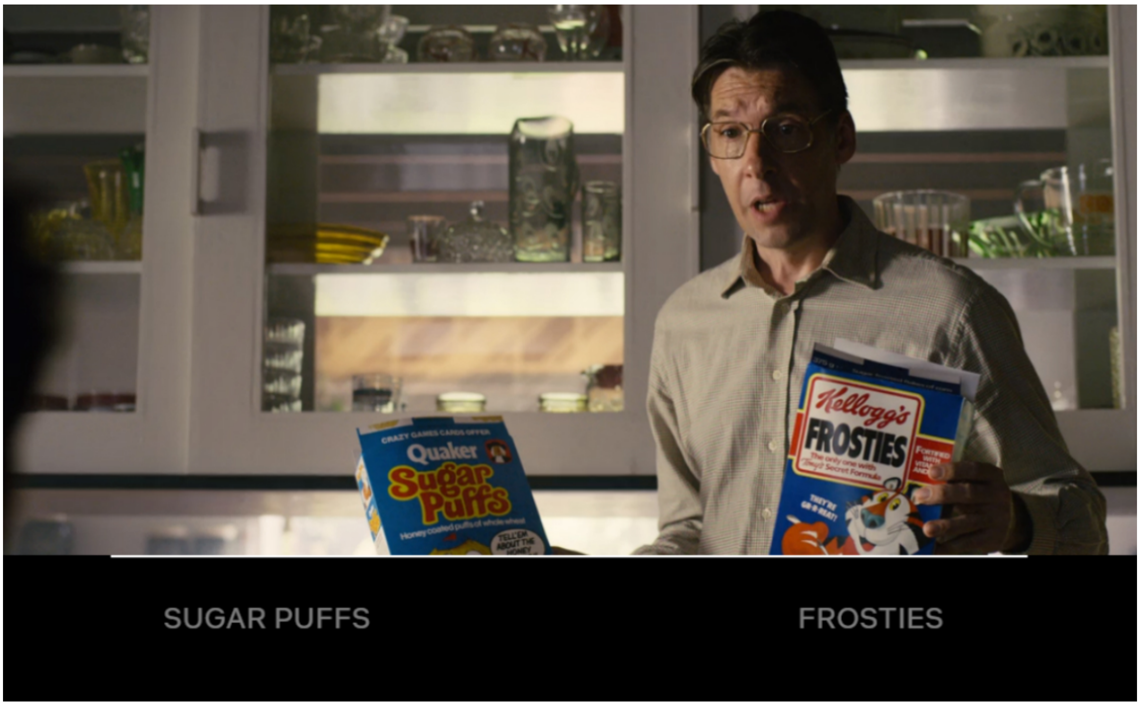


FIGURE 2.12: Screenshot from the interactive film *Bandersnatch* (source: Netflix). At several given moments in the film, the viewer is presented with two choices. These may affect the current character's decision of imminent action. A countdown timer is also presented (the thin line above the options). If the viewer does not make a choice before it runs out, a default option is automatically chosen.

Explicit and Implicit Interaction. In typical designs of a viewer's interaction with the UI of the narrative system, visible elements, such as a circle target, a countdown marker, or a translucent dot [63], assist the viewer in becoming aware of the location of an ROI, the effectiveness of a recent input, or a commencing event. When novel designs started to move away from a conventional UI, creators also explored the possibility of enabling viewer input without relying on any visual interface or reaction element. Rothe and Hussmann [127] list parameters around the activation targets. Targets can turn visible when triggered by the viewer or can remain invisible, depending on the requirement of the narrative itself. In most narrative VR, like with movies, even a small non-diegetic object can be disturbing and break the feeling of presence. Thus, the trigger or target will need to be either visible only when activated or invisible throughout the entire experience.

Explicit controls are also seen in museum tours, ancient site visits, etc., if choices for visitors are enabled. A typical interactive method in this sector, known as the perspective variation, was later proposed as a solution to this challenge [84]. A strategy described in that method recommends that creators include a cluster of different versions of the same story in their experiences. The viewer is then given the option to choose a perspective s/he would like to experience or the character to tell the story if there is an explicit storyteller. In the exhibitions *The Hague and the Atlantic Wall* [112] and *Narratives in the Trenches of WWI* [91], visitors can choose one of the characters from the scene by explicitly selecting one among several contextual objects at the start. Then the narratives the visitor will hear and see in the tour will all be from that character's perspective.

Other researchers chose the implicit design. Ibanez et al. [60] constructed a virtual tour system that generates stories based on the location designated by the viewer and the location where a virtual tour guide was standing. Therefore through the entire tour the viewer feels like s/he is guided by a real tour guide, with all the knowledge and responses to his/her choices of ROIs instead of linear prerecorded video footage. However, all the granular narratives are prefabricated. Therefore, compared to a conventional system where the viewer only gets to navigate along a one-dimensional timeline with explicit input, in this virtual tour system, the viewer naturally browses the scene, and the narrative structure changes accordingly on the fly. During the entire process, the viewer is providing implicit input to the system (naturally choosing where to look and focus, as one would do in a real-world tour), and is unaware that s/he is making choices.

2.4.2 Non-linear Story Structures and Their Interaction Designs

After acknowledging the necessity of viewer participation and exploring possible designs of viewer interactions in CVR, creators also realized that the traditional linear structure common in novels and films is no longer suitable for this type of new medium [98]. Since the viewer can intervene with the progress of the narrative (despite various levels and types) and the viewer's specific behaviors are not predictable, the specific progress routes of an interactive experience in each playback are also not fixed. New types of story structures are borrowed and developed from other interactive media, such as video games and improvisational theater, by creators of interactive CVR [138]. Carstensdottir et al. created a group of typical story structures for interactive narrative games whose aims are all storytelling. Other than the traditional linear story structure, they pointed out five types, *branching, foldback, broom, hidden*, and *opportunistic*, all applicable to some extent to immersive storytelling [24]. Some of the typical story structures are illustrated in Figure 2.13. Millard et al. came up with a simpler model, the CDP model, that classifies different non-linear story structures into three groups: *canyons, deltas*, and *plains* [95]. Creators and researchers later realized that the typical branching story easily leads to a tremendous amount of production and programming work on content to fulfill the needs of multiple lines and endings and will eventually cause the author to lose control of the progression of the narrative [122]. They upgraded the branching design with new forms that come with a controlled single ending but variable paths in the middle, such as the *string of pearls*, *detour* [157], and *gauntlet* [122].

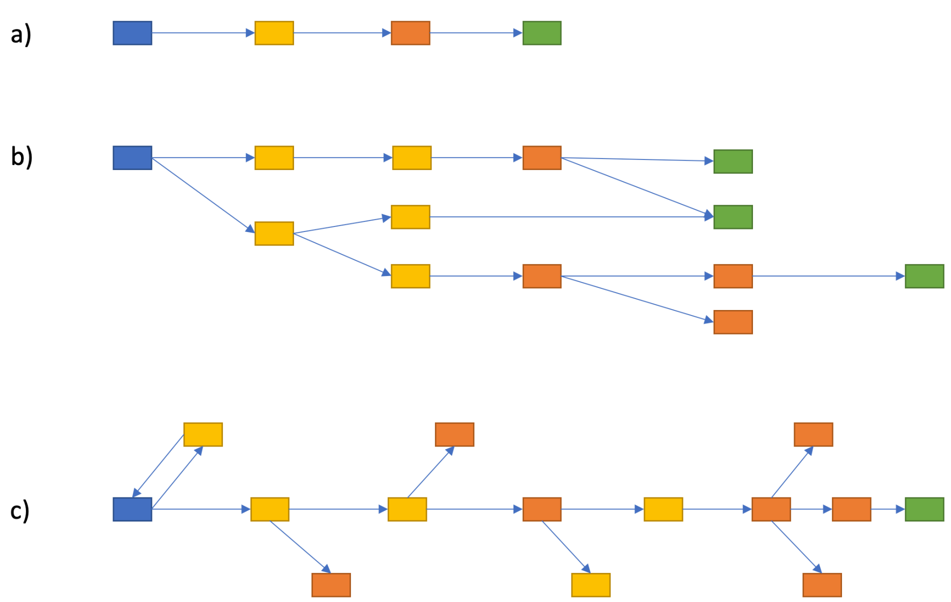


FIGURE 2.13: Three typical story structures. The color blocks are *events* in a story. The arrows connecting them are the flow of narrative. (a) The traditional linear structure: one event happens after another, the story only has one path, and everyone will get the same version; (b) the *branching* structure: at every event there are multiple paths to move forward, leading to several different endings; (c) the *gauntlet* [122] structure: although there are branches along the path, generally there is a main storyline and one ending. Other types of non-linear stories are not illustrated here due to space limitations (see [24] for more

.

The above exploration of interactive storytelling did not simply cover the static story structures but also examined how viewers dynamically interact with them. Carstensdottir et al. also listed four types of narrative progression mechanisms, in other words, how viewers will interact with the storytelling system, such as *progress through choices* and *progress through scripted scenarios*, and pointed out that the progression mechanisms and the story structures are paired, instead of a free choice of combination, when implemented [24]. Researchers who looked at other interactive storytelling experiences and creators who build storytelling projects have also justified that in immersive storytelling scenarios, the viewer needs to be presented with a form of interaction technique that matches their role in the experience, including their level of participation and capability for control in the narrative progress (see section 2.4.1) [126, 33]. The takeaway here is that creators have realized how the viewer will estimate the structure of the story when facing the type of progression mechanism they are provided with, such as *a viewer is presented with choices, s/he will expect a branch story structure*, and will expect the narrative to progress in the way s/he expect. Thus creators need to follow these patterns in the viewer's mind, otherwise the viewer's expectation and understanding will turn out to be incorrect and their experience will deteriorate.

2.5 CVR and Personalized Storytelling

After bringing viewer interaction to CVR, some researchers moved on to look at the possibility of acknowledging individual preference via the immersive storytelling system to create "personalized storytelling", since each person has her/his own characteristics when going through any experience. In a series of studies, creators and practitioners have explored methods for bringing the characteristics of flexibility and dynamic change of oral storytelling into other forms. In cultural heritage and the galleries, libraries, archives, and museums (GLAM) sectors, people have been looking at adding personalization to the visiting experience [35, 2]. Among those, Not and Petrelli [102] proposed that in order to enable personalization, three factors need to be implemented:

- Customization—the content is prepared to be able to be reshaped and morphed into different scenarios,
- Context awareness—the system can understand the current scenario and adjust accordingly, and
- Adaptivity—the system can change/alternate to deliver the correct form of experience to the current viewer.

Comparing these works, a common pattern with three necessary elements was found to construct a personalizable experience:

- Customizable content,
- Creation of a visitor's profile, and
- Creation of storytelling variation based on these profiles.

These will be described in detail in the following sections.

2.5.1 Customizable Content

In most types of media, such as books or films, stories are fixed once recorded or produced. Creators and researchers have long tried to reintroduce the characteristic of mutability to stories told other than orally. Zagalo et al. [163] and Lukin and Walker [84] discussed the fabula/syuzhet relationshp and proposed the idea of detaching the actual delivery of a story segment from how it was fixed in its core content. They also included detailed introductions about the use of this two-layer system in story composition and story presentation/disclosure. They explained that in storytelling, there are two levels of narrative representation: (1) the story, or fabula: the content of a narrative in terms of the sequence of events and relationships between them, the story characters and their traits and affects, and the properties and settings; and (2) discourse, or sujhet/suyzhet: the actual expressive telling of a story as a stream of words, gestures, images, or facial expressions in a storytelling medium [84]. Story and discourse are related but not identical or mirrored. Instead, in a narrative, events from the story are selected, ordered, and expressed in the discourse.

2.5.2 Creating a Visitor's Profile

In order to generate visitors' profiles, it needs to be worked out 1) what features are used for profile generation and 2) how the system is going to collect those features. The selection of features first depends on the type of profile the system will need, grouped as stable profiles, such as those based on demographic information, and dynamic profiles, those that are closely related to the current visit [102]. Normally for a stable profile, the system will use socioeconomic factors such as age, gender, or place of birth, or cognitive factors, such as personal interests in daily life [1, 111, 132]. For a profile that is dynamic and closely related to the current visit, the system will use features such as the visitor's motivation, knowledge related to the content, her expectation, and preferred itineraries for this visit [112].

Once the necessary features are selected, the next decision is around how the system will collect information in relation to those features from the visitors. Many previous projects used various methods to ask visitors questions directly and collect their answers as data. These include:

- Explicitly asking via a questionnaire before the experience [131, 118, 11]; this is similar to real-world rituals, such as the pōwhiri used by Māori to welcome new visitors before entering the marae;

- Implicitly collecting information via a short interactive interview rather than asking direct questions to avoid the cold-start problem [65, 1];
- Implementation of objective-based methods [61, 91, 102], whereby visitors were invited to pick an object among several choices all thematically related to the experience; the choice is used as a representation of visitor's preferences in terms of theme, language, character, etc.

There are also creators and practitioners who choose not to use visitors' answers as direct information as it requires the system to generalize every visitor's profile from the data. This requires enforcing a process that every visitor must go through and ensuring all required information is collected from every one of them, which can increase the time and effort expended.

Alternatively, some researchers chose a classification-based method [123, 131]. The content creator anticipates and prepares a list of expected visitor profiles (normally as personas) in advance. The system collects a few key features from the visitor and implements a matching process with the personas to put each visitor into a category. It then responds to labeled categories rather than to specific visitors. Although relatively coarse, the classification method reduces the number of profiles created and saves time spent on matching content to profiles [132].

2.5.3 Strategy of Creating Variations

After obtaining visitor profiles, the next step is to apply variations to the customizable content. However, selection of the specific variation strategy to use is a case-by-case decision and depends on the system's design and purpose of the experience. Variation was introduced towards two aspects in storytelling in the cases reviewed: the temporal succession of content segments and the disclosure of a given story segment. Variations based on rearranging the temporal succession of the content segments are completed either before the experience starts or at every junction of two adjacent segments. The choice of segments is usually given to the visitors and coupled with their physical input [15, 91, 101]. Variation based on changing the disclosure of a given story is more of an artistic design choice, with abundant implementation techniques. Several cultural heritage tours were presented in section 2.4.1, from the *Fabula Tales* framework proposed by Lukin et al. [84]. Although not explicitly set to one of the strategies, the disclosure of the same story varies in terms of POV and of style of speech as different characters delivered it. In these cases, visitors will sometimes need to revisit the same tour with different characters' voices to grasp the whole story.

2.6 Conclusion of Status Quo

In the previous sections, the background knowledge related to storytelling using VR was reviewed, including the basics around storytelling, the well-established storytelling forms (cinema, theater, and video games), early examples of immersive storytelling, and the more sophisticated and flexible storytelling device known as CVR. Viewer interaction and viewer personalization were also explained as new design considerations unique to immersive storytelling. Previous creators started by borrowing skills and grammar from filmmaking for the production of 360-degree videos, then began to consider the role of the viewer, their possible participation, and the balance of power between viewer and creator, specifically in terms of immersive storytelling experiences. Their work reveals that CVR creators do acknowledge the viewer's participation and the importance of viewer interaction. However, the design space is yet not fully explored, as:

- From the creator's side, there is still a lack of well-developed content preparation guidelines, design templates, and practical methods to assist the VR storytelling workflow compared to traditional 2D filmmaking;
- Also, creators continue to seek ways to effectively presentation content to the viewer addressing the issues of NP and FOMO;
- Creators and researchers continue to explore appropriate interaction techniques for CVR taking into account how viewers would like to interact. However, the means and extent of interaction viewers are willing to accept remain unknown. There is still a need to find a way of striking a balance between authorial control and the viewer's participatory involvement, one that satisfies the viewer's expectation on agency while also ensuring the storytelling is compelling and well delivered;
- In most current CVR productions, the viewer is treated by creators as having an abstract and featureless role in the storytelling process. Personalized CVR experience is yet not fully explored, nor has an engaging and enjoyable personalized CVR experience been published.

In the following chapters, studies conducted during this PhD research will be introduced. These are aimed at solving the aforementioned problems through further exploration and experimentation. The first study examines CVR content production techniques for creators. The second study was built upon the conclusions of the first, moving on to explore the viewer's role and viewer interaction in CVR experiences.

Chapter 3

Producing CVR Content with Attention Guidance

3.1 Introduction

In recent years, 360-degree videos have become a popular way for people to experience virtual reality (VR). Unlike conventional video, viewers watch 360-degree videos with VR headsets, using free head rotation to decide where to look. As described in section 2.2, this form factor leads to the issues of *narrative paradox (NP)* and *fear of missing out (FOMO)*, yielding weak narrative comprehension and low emotional engagement. To solve those issues, creators first tried to migrate filmmaking grammar from cinema to 360-degree videos. However, they discovered they are not as effective because the viewer's point of view (POV) and thus viewer's role has changed in the 360-degree videos. Creators and researchers then focused on using attention guidance cues to drive the viewer's attention to story-matter elements to maintain narrative awareness. However, those cues need post processing work [162] and will break the feeling of presence [55] as they are mostly non-diegetic (*Diegetic* is a Greek term meaning "belonging to the story scene, part of the story"). A few others have tried using diegetic cues but with limited exploration.

Based on previous work on 360-degree video production [39, 93, 18], the *Triangle System* was first created as a guideline system for 360-degree video creators to help them transition from traditional videos to 360-degree video productions. Derived from the well-known *Triangle Method* widely used by filmmakers to create conversation shots [3], but modified for 360-degree videos, the system focused on element arrangement by explicitly acknowledging the viewer as part of the scene and a character in the story. Figure 3.1 gives a visual presentation and explanation of the *Triangle System*, which covers the placement of the camera, the ROI and the storyteller her/himself in the scene, as well as the maintaining of the distance between each element. The Triangle System covers three scenarios, which are all depicted in Figure 3.1: A) A stationary camera with the ROI nearby, B) A stationary camera with the ROI far away, and C) A moving camera that moves with the storyteller. When the camera is stationary, and the ROI is close by, the storyteller will stand in a place that is beside the camera with a normal distance similar to those in daily conversation and within reachable distance to the ROI. When the ROI is far away, the storyteller then chooses a location where both s/he and the camera will have a similar field of view (FOV) to the ROI (B in the figure). In both scenarios A and B, the storyteller, the camera and the ROI form a triangle (as drawn in the figure, with blue shades). In scenario C, while moving, the storyteller will keep the camera moving on the same side all the time. Now the storyteller, the camera and the ROI form a triangle again, whether near or far. Thus, both the storyteller and the ROI will be in appropriate locations visible to the viewer, and the viewer will not have to turn her/his head back and forth.

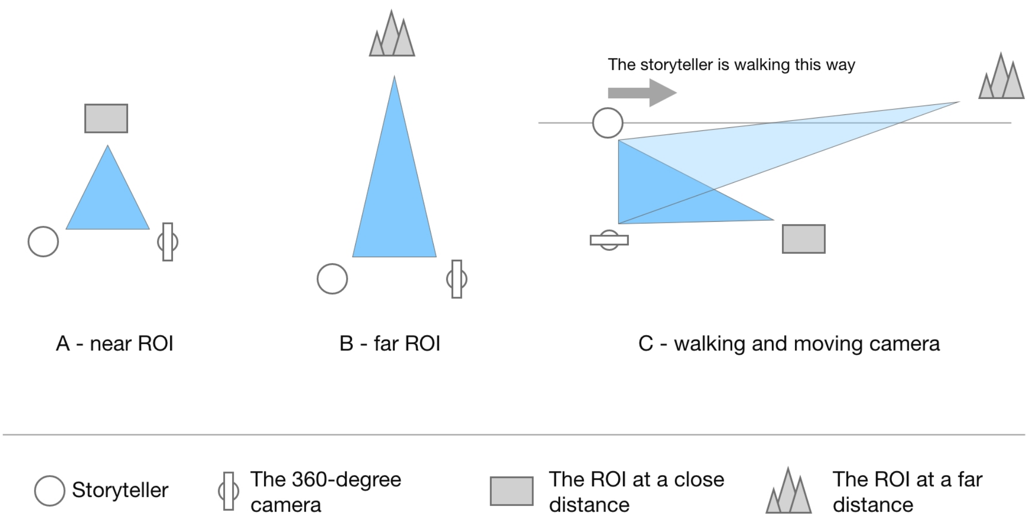


FIGURE 3.1: A visual presentation of the Triangle System. In various scenarios, the system reminds the storyteller to be aware of the distance to the ROI (close-by or distant) and the camera's mobility (moving or stationary), and then apply arrangements accordingly.

3.2 Our System - Action Units (AUs)

The Triangle System proved effective in helping compose content for a 360-degree video in some early trial runs [153]. But it had not yet touched the use of guiding techniques in the scene. The works done by Brown at el. and by Lindeman both indicated that applying social communication cues as attention guidance in 360-degree videos could be effective and at the same time, avoid post processing [19, 81]. Cues like those have also been intensively utilized on theater stages [82]. In this chapter, Action Units (AUs) are proposed as a directorial method and an attention guidance technique that a storyteller can adopt for onsite use. Compared to the other guidance techniques that require post-processing work, AUs can be created at the same time the creator captures the 360-degree videos.

The development of the AUs was started from the similarities between the storyteller-viewer relationship in 360-degree videos, and the face-to-face conversations in the real world. In a typical conversation, people use social cues to attract attention or convey intentions [72]. Tomasello et al. [152] stated that direct speech was necessary as the communicative context for mutual understanding. Accompanying speech and non-verbal cues, including facial features and gestures, were frequently utilized to convey extra information like intentions and changes of focus [30, 52]. By comparing films with real life conversations, Kappelhoff and Müller [64] pointed out that expressive movements, such as gestures, can trigger the same kind of felt experience in the spectator in a cinematic experience, as those in the real world. This happens as the viewer goes into a perceptual sensation of another ego, which is physically-sensually embedded in the delivered content. Similarly, when watching 360-degree videos and the viewer enters the temporary egocentric position of being addressed by the storyteller in a (virtually) face-to-face manner, social cues are believed to trigger the same effects as in real life communications and work as attention guidance cues. Ravenet et al. [119] also stated that people use recurring patterns, e.g. gestures, to map conceptual metaphors from one entity to another. This means that an effective attention guidance technique will not only need to be diegetic to the content or the scene of the video itself but also formed as prefabricated units. Thus their meanings can be widely acknowledged and used by storytellers and viewers, for both notability and comprehensibility.

A series of AUs were then created from three frequently used non-verbal units of social cues [48]. Among them, arm and head movements are used to emphasize objects, and eye contact is used for establishing social acknowledgment and enhancing the feeling of co-presence [67]. Therefore, three AUs were chosen for 360-degree video production based on these non-verbal cues (Examples are shown in Figure 3.2). They are:

- **AU01**: Conversation - the storyteller keeps eye contact with the viewer;

FIGURE 3.2: Examples of SMEs using AUs from virtual field trips recorded as 360-degree videos. The AUs are highlighted in the shots with bright yellow rectangles (the rectangles were for demonstration purposes only and were not visible in the actual 360-degree video). In the top row the expert used mainly AU02 and AU03. In the bottom row the expert used all three AUs under different scenarios, which are all shown in the pictures.

- **AU02**: Pointing - the storyteller points at an ROI when talking;
- **AU03**: Looking - the storyteller turns her head and looks at the ROI explicitly

In 360-degree video production, the storyteller can choose which of the AUs to use according to the content and the narrative she wants to deliver. AUs serve as call-to-action points, so when key elements appear, the storyteller will be assured that her viewers will follow and have an engaging viewing experience. When one is wearing a VR headset and watching a 360-degree video of a subject-matter expert (SME) giving a narration, a person will have the immersive experience of "being there", as if she is standing beside the expert, looking and listening. Thus, the expert in front of the 360-degree camera can use AUs when delivering the narrative to direct her to focus on the ROIs in the scene, as if she is having a face-to-face conversation with the expert.

3.3 User Evaluation of AUs

With AUs composed, the preferred user scenario for 360-degree videos, especially the ones where the storytellers give a presentation about a specific topic or object, was first explored.

3.3.1 Swivel-chair VR

In traditional filmmaking, the directors assume a passive viewer sitting in a chair looking straight ahead [17]. Similarly, the term *Swivel-chair VR* is put forward to specify the preferred way to consume 360-degree video content. *Swivel-chair VR* describes a scenario where the viewer watches a 360-degree video while sitting in a swivel chair, wearing a VR headset, as shown in Figure 3.3. By sitting on a swivel chair, the viewer can rotate around 360 degrees by turning the body together with the chair. It is more comfortable than only turning with one's neck since the viewer will have trouble looking directly backward, leading to muscle strain. The swivel chair also serves as a cue to imply the affordance of rotation, and an anchor point to physically ensure that the viewer is not moving around. The anchored position 1) matches the capture point of the camera while in production, providing a better immersive experience as the viewer feels like her body posture matches the visual perception within the virtual environment, and 2) gives the viewer more confidence in her safety in the real world, as the viewer is less likely to tip over or bump into obstacles. Cabling must be managed through, e.g. with the help of an operator.

3.3.2 The Pilot Study - Field Tests of AUs

A pilot study was first conducted to evaluate whether the general public (non-VR experts) can effectively use 360-degree videos as a novel technology in their routine work and apply AUs into the production workflow without intensive learning. Four SMEs (a volcanologist, a botanist, a glaciologist and a polar photographer) were invited to use 360-degree cameras to capture their scientific field works and implement AUs during the captures. The experts were first trained with the knowledge of how to use AUs and the scenarios suitable for each of the AUs. The SMEs then composed their shooting plans according to the locations and intended content with my assistance. As with traditional filmmaking, storytellers need to be able to convey their ideas about what they expect for a given shot or shot sequence. In those field trips, AUs were used to create anchor points for the experts to formulate shooting plans. Multiple snapshots from the 360-degree videos produced by the SMEs were extracted, then displayed in Figure 3.2. All AUs appeared in the snapshots are highlighted to demonstrate how they served the purpose of bringing effective storytelling to the 360-degree video.

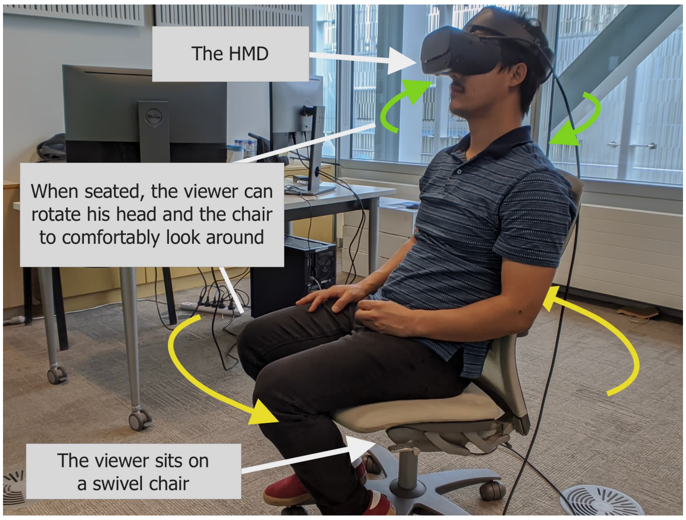


FIGURE 3.3: Swivel-chair VR: A viewer wearing a head-mounted display (HMD), watching a 360-degree video while sitting in a swivel chair. When seated, he can turn his head and the chair to look around comfortably.

After training, the glaciologist captured 360-degree videos when traveling with a group to Antarctica on a research mission. Sample video snapshots in the top row of Figure 3.2 show the camera was stationary among a group of people, mimicking the viewer just being one of them. The expert used AU02 and AU03 when introducing the mountains at a far distance and the crevices under the snow nearby. It helped the viewer locate and identify the ROIs that were being mentioned and keep up with the narrative. The volcanologist took a 360-degree camera on a solo field trip to capture presentation materials for his geology class. He extensively used all the AUs in his captures, as shown in the bottom row of Figure 3.2. In the first two snapshots, the expert applied both AU02 and AU03 for the lava features on the mountain far away and rocks nearby. The AUs worked as visual guides to aid the viewers in shifting focus from the expert to the ROIs, and then back to the expert himself. In the third snapshot, he used AU01 to deliver a speech, giving the viewers the feeling that they were standing by the stream with him and listening to the talk simultaneously.

The SMEs were also interviewed afterwards about their impressions of the use of 360-degree videos and the AUs. They indicated that AUs were easy to implement and required no extra burden of technical skills. The students from the SMEs' classes were also invited to watch those virtual field trip videos with the Swivel-chair VR setup. They anecdotally commented that they had strong feelings of presence and better memory of the narrative presented, compared to traditional 2D videos. The pilot study demonstrated the usefulness and practicality of using AUs in actual productions, from the perspective of SMEs as content creators, as well as their positive effects on attention guidance and reducing the FOMO issue.

3.3.3 Experimental Design

Based on the findings from the pilot study, further explorations were conducted to see if AUs could be as effective as the commonly used post-processing methods or even outperform them, from the perspective of viewer experience.

A within-subjects experiment was conducted to compare the AUs' performance with two other attention-guidance techniques, "Pointing Arrow" (PA) and "Angular Shift" (AS), as well as the baseline without any guidance (BK) [114, 55]. Based on the series of 360-degree videos of a campus tour as our experiment environment, different guidance styles were implemented on the videos as the independent variable (type of attention guidance) with three levels (AU, PA, AS). Each participant experienced all types of guidance. In the study, the effects of different levels on viewer behaviors (as the time taken to locate a ROI and the time of attention on that ROI), recall rate, and subjective feelings (as levels of engagement and enjoyment) were both measured. The AUs' performance compared to PA and AS was explored. The exploration was conducted on four aspects: 1) increasing levels of engagement and enjoyment, 2) affecting the recall rates, 3) reducing the search time and increase the attention time on an ROI, and 4) reducing cybersickness, when applied to 360-degree videos.

3.3.4 Hypotheses

Four hypotheses were proposed based on comparing AUs with guidance techniques used by other researchers in their experiments. In the study by Nielsen et al. [99], they pointed out that the synthesized non-diegetic cue reduced the viewer's perception of presence, as AUs were embedded within the content during the production stage. Since the AUs were performed by the storytellers themselves, it was considered diegetic, thus would impede less on the viewer's feeling of presence, thereby increasing the level of engagement and enjoyment. Regarding the recall rates, Li et al. [77] observed that the use of VR in learning cannot positively contribute to the learning effectiveness measured through a memory test when the students experienced a higher level of enjoyment compared to the use of conventional learning materials. Thus, the following two hypotheses were proposed:

- H_1: A viewer will feel more engagement and enjoyment when watching a 360-degree video shot using AU, compared to PA and AS.
- H_2: Compared to BK, a viewer's recall rate will be lower when AU, PA or AS is used in the video.

When testing various guidance techniques with 360-degree videos, both Speicher et al. [143] and Lin et al. [80] state that methods like the AS will introduce cybersickness and disorientation. They further point out that the accuracy of locating an expected item in the scene was higher when using a human actor as a guidance cue than when using an added object such as AS. Those indicated that AS would deteriorate the viewer's performance on searching and locating as it will introduce cybersickness. Also, the PA is an add-on cue, while AU uses human actors. Thus, another two hypotheses were also proposed:

- H_3: AU will reduce the Time-to-Search for ROIs and will extend the time a viewer focuses on ROIs, compared to PA and AS.
- H_4: AU will reduce a viewer's level of cybersickness when watching a 360-degree video, compared to PA and AS.

3.3.5 Measures

In this study, the levels of engagement and enjoyment were measured by the Questionnaire of Engagement Enjoyment and Immersion (E2IQ) [78]. The E2IQ was divided into two parts, E1 and E2, where E1 measured the level of engagement and E2 measured the level of enjoyment. The participants chose from a series of 5-point Likert scale options. The choices were summarized and converted into two numeric values, both between -1 and 1 (-1 = not engaging/enjoyable at all, 1 = very engaging/enjoyable). The memory effects were evaluated by recall tasks that asked participants to identify objects which were both visible and mentioned by the tour guide (but not described in detail) in the videos. The ease of search from the viewers was measured with Time-to-Search (TTS). TTS was counted from the point when the tour guide explicitly mentioned the ROI, to the point when the viewer located the ROI. The latter event was triggered when the gaze crossed the 29.2 degrees threshold (half of the single eye FOV of the HMD), which is regarded to be the peripheral limit [148], from the outside of her FOV. The level of attention was measured with the Time-on-Target (TOT) counted from when a participant located the ROI (ROI crossed the 29.2 degrees threshold) until it left her FOV (moved out of view). Both TTS and TOT were recorded in real-time by a script running in the background. The levels of cybersickness were measured by the Simulator Sickness Questionnaire (SSQ) [68].

3.3.6 Apparatus

In this study, the video playback and attention guidance techniques were all implemented in Unity3D 2018.3.11f1. A computer running 64-bit Windows 10 Professional with a 3.2GHz i7 processor and a GeForce RTX 2080 graphics card was used to implement the cues, record viewer behavior data, and ensure the smooth playback of the 5.7k videos. During the experiment, participants viewed the 360-degree videos wearing an Oculus Rift S[1] HMD without using its controllers, as shown in Figure 3.3.

3.3.7 Material

Before the experiment, eight 360-degree video clips of a campus tour were captured with an Insta360 ONE X 360-degree camera[2], at a resolution of 5.7k (5760 x 2880). The recordings were captured at four pre-selected locations (1 and 3 were outdoors, 2 and 4 were indoors). An actor played a tour guide. She introduced the places and described ROIs around, e.g., buildings, decorations and other unique objects. The ROIs in each clip were non-identical and none of them appeared more than once. Two takes were captured at each location. The actor used AUs with the narration in the first take (called “AU clips”), and she repeated the narration with neither head movements nor gestures in the second take (called “Blanks”). The Blanks were used as the Baseline (BK), and also augmented with PA and AS in a post-processing step. The camera was mounted on a tripod, thus, the viewpoint was fixed in each video clip. The actor’s distance to the camera was kept the same (2m) at each location so the viewer would feel a similar space between themselves and the tour guide when watching. The eight clips were then processed using Adobe Premiere Pro to adjust the volume levels and narration pace to minimize differences between the recordings. All the videos were trimmed to a length of approximately 90 seconds. After this adjustment, the AU clips were ready for the experiment since they already had the cues embedded in the content.

[1]https://www.oculus.com/rift-s
[2]https://www.insta360.com/product/insta360-onex

PA was implemented into the Blanks in Unity3D by adding an arrow fixed to the viewer's FOV. The arrow stayed in front of the viewer's FOV and constantly rotated to point itself towards the ROI being described by the tour guide at the moment, as shown in Figure 3.5b, d and f.

AS was also added to Blanks using Unity3D. The details of the AS mechanism are presented in Figure 3.4. Using the method from Tanaka et al. [148], the scenes were automatically rotated to shift the ROI to the front of the viewer when her head orientation was beyond a certain angular threshold from the ROI. In this experiment, the threshold was adjusted to 50 degrees (instead of 40 degrees in Tanaka et al.), which is half of the FOV of the HMD used in the experiment. Therefore the shift occurred when the ROI was outside of this threshold, and stopped once the ROI became visible to the viewer. The rotation speed was set to 15 degrees/s, the same as in Tanaka et al.

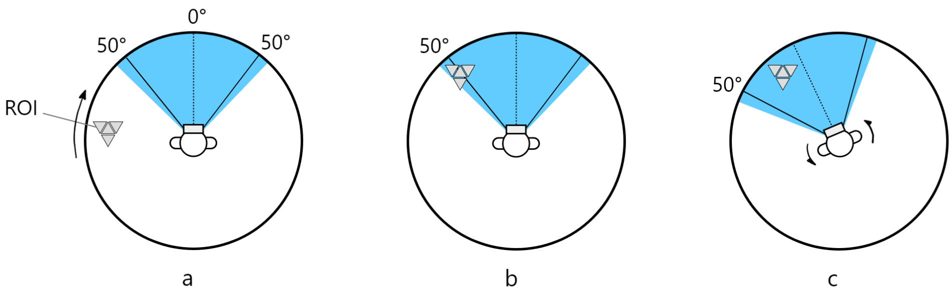


FIGURE 3.4: An example of the AS method, illustrated in a top-down view of the viewer wearing an HMD and the scene around her. The head orientation is marked as a dashed line, and the FOV is highlighted in blue. a) The viewer is looking towards the front of the scene. Now the ROI is located to her left, out of her FOV and outside the 50-degree threshold. As AS is taking place, it shifts the ROI towards the front of the viewer; b) The AS has reached the threshold, and AS stops. The ROI is now within the FOV, most likely triggering the viewer to turn her head towards it; c) When the ROI is within the FOV, although the viewer rotates her head, AS will not be re-activated unless the ROI again falls outside of the threshold.

3.3.8 Participants

Twenty four participants (14 females) were recruited from the university. All were between 18 and 39 years old ($M = 25.37, SD = 4.641$). Six of them had never used VR headsets, and eighteen of them reported they had experienced VR , but were limited to only a few times a year. Among those with limited experience, eleven had watched 360-degree videos (reported as "seldom, only a few times").

None of the participants had extensive use of VR headsets, nor any previous experience of 360-degree video production.

3.3.9 Procedures

In this study, four conditions (AU, PA, AS, and BK) were compared by applying them to the captured 360-degree videos and having the participants watch them with VR headsets. The experiment was approved by the Human Research Ethics Committee of the University of Canterbury.

Before the session started, each participant was presented with two examples of the recall tasks and received detailed instructions on how to perform those tasks. A debrief about how to use the Oculus Rift S VR headset was also given to each participant. Then they sat in a swivel chair, put on the HMD, and watched an introductory video to get familiar with the technology and this form of medium, and to reduce the effects of novelty and anxiety before starting to watch the actual content. The video had neither a tour guide nor any attention guidance technique attached.

After the introductory video, each participant watched four content video clips of the "campus tour." To keep the story's integrity, every participant watched content video clips 1 to 4 in the same order, but with four conditions (AU, PA, AS, and BK) randomly assigned and counterbalanced with a Latin Square. At the end of each clip, the participants were asked to remove the headset, and immediately move on to complete a viewing experience questionnaire to measure cybersickness, engagement, and enjoyment, followed by the recall tasks while the experience and memory were still fresh. Then they were given a one minute break before moving on to the next clip. Each participant watched five videos (1+4), answered four questionnaires, and completed four batches of recall tasks. After all video sessions, a short post-test interview was conducted to ask about the participants' general feelings, feedback, the preference of the guidance techniques, and the reasons for these. The entire session took approximately 40 minutes for each participant.

3.4 Results

The data from the participants' reported questionnaires and the logged time values from the Unity Engine were both analyzed using one-way ANOVA tests ($\alpha < 0.05$). Table 4.1 shows the mean values of SSQ scores, E2IQ scores for Engagement (E1) and Enjoyment (E2), and the recall rates, TTS, and TOT, for each condition. The

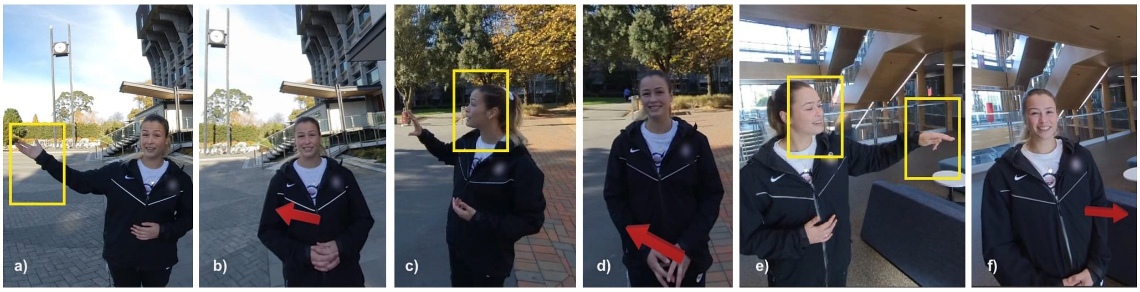


FIGURE 3.5: Snapshots from the campus tour video with attention guidance methods applied. The AUs used (highlighted in yellow boxes) were a) pointing only, c) pointing then looking back, and e) looking and pointing at the same time. "Pointing Arrow" (red arrows pointing to the targets) is shown in b), d) and f), with the scenes matching to a), c), and e), respectively.

most outstanding values of each measure (if present) are highlighted in the table. The overall results are also summarized and plotted in Figure 3.6.

TABLE 3.1: The mean values of the results of Engagement, Enjoyment, Recall rate, TTS, TOT, and Cybersickness for each condition.

Measure	BK	AU	PA	AS
Engagement	0.32	**0.41**	0.32	0.16
Enjoyment	0.13	0.23	0.17	-0.04
Recall Rate (%)	88.25	74.33	77.32	73.96
TTS (s)	4.39	**2.74**	3.38	8.30
TOT (s)	4.44	4.12	**6.52**	2.28
Cybersickness	26.33	22.51	15.38	**80.77**

3.4.1 Engagement and Enjoyment

The levels of engagement were higher in the videos with AU or PA (AU: $M = 0.41, SD = 0.222$, PA: $M = 0.32, SD = 0.260$), compared to BK ($M = 0.32, SD = 0.190$). For videos with AS, the level was lower ($M = 0.16, SD = 0.455$). The ANOVA test indicated significant differences among the conditions ($F = 2.871; p = 0.041$). A Tukey's HSD Post-hoc test indicated that AU performed significantly better than AS on level of engagement ($p = 0.025$). But no other statistical differences were found between AU and PA, PA and AS, or AU and BK. The results indicated that participants felt higher engagement with the narrative of the videos with AU, compared to those with AS.

Participants also reported higher levels of enjoyment when watching the videos with AU and PA (AU: $M = 0.2283, SD = 0.35$, PA: $M = 0.174, SD = 0.38$), compared to BK ($M = 0.12, SD = 0.39$). Nevertheless, participants reported lower levels of enjoyment with AS ($M = -0.039, SD = 0.53$). However, the ANOVA test did not reveal

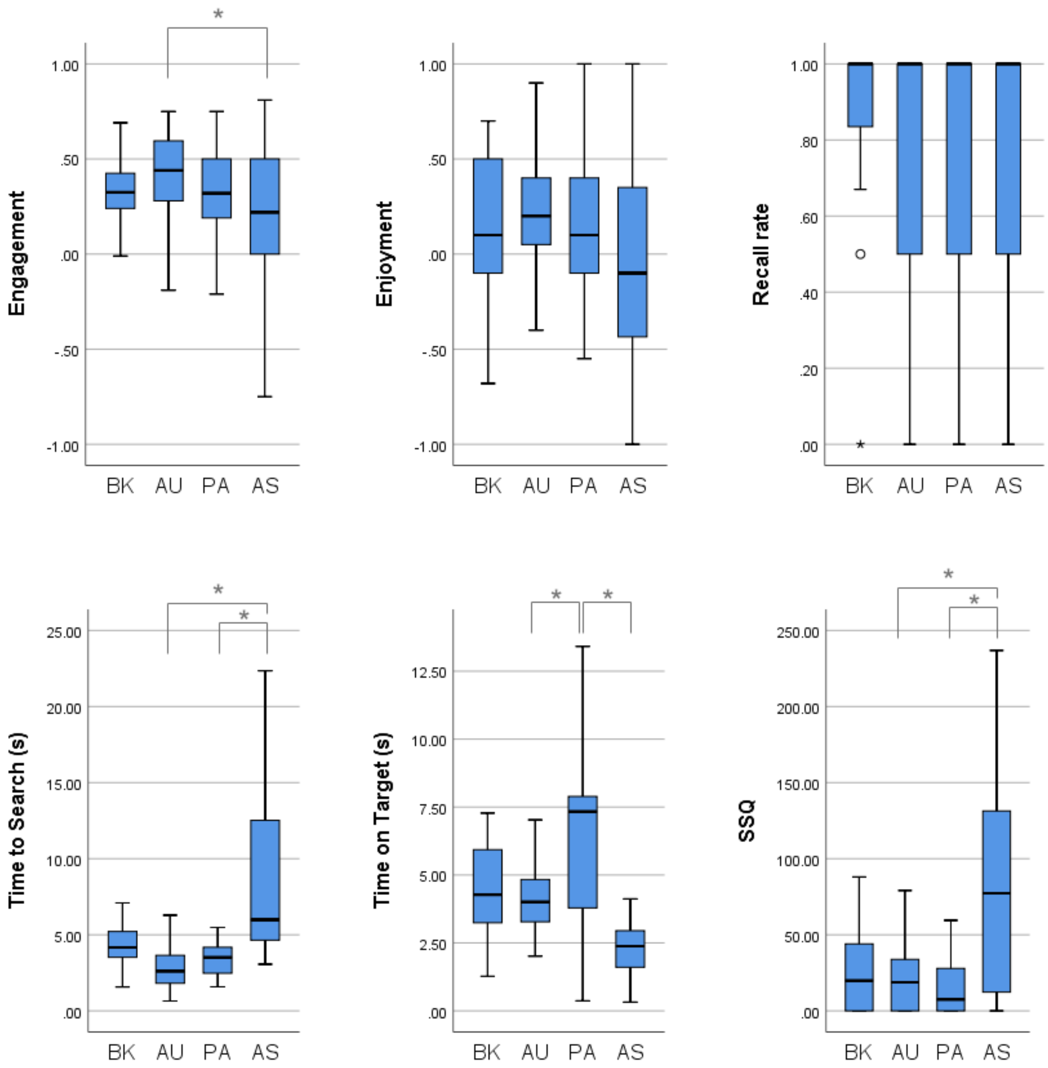


FIGURE 3.6: Boxplots summarizing the results of levels of Engagement and Enjoyment, Recall Rates, viewer performance of search (TTS) and attention (TOT), and the level of discomfort of each condition (SSQ), in terms of medians, interquartile ranges, minimum and maximum ratings. Top row, from left to right: Engagement, Enjoyment, and Recall rate. Bottom row: TTS, TOT, and SSQ.

any statistically significant differences in the level of enjoyment among the four conditions.

3.4.2 Memory

The recall rates were all lower than the BK ($M = 0.8825, SD = 0.24$), when an attention guidance was applied, whether it was AU ($M = 0.74, SD = 0.380$), PA ($M = 0.77, SD = 0.308$), or AS ($M = 0.74, SD = 0.354$). However, there was no significant difference among the recall task performance of the four conditions ($F = 0.968; p = 0.411$). A Post-hoc comparison was also conducted using the Tukey's HSD test. The results also indicated there were no significant difference pairwise in AU and BK ($p = 0.455$), AU and PA ($p = 0.996$), or AU and AS ($p = 0.999$).

3.4.3 Search and Attention

TTS results indicated that participants took less time to find the targets (a viewer turns her head to search for the ROI when mentioned by the storyteller) with AU and PA (AU: $M = 2.74, SD = 1.260$, PA: $M = 3.38, SD = 1.115$), compared to BK ($M = 4.38, SD = 1.540$). However, with AS, participants took longer to locate the targets ($M = 8.30, SD = 5.310$). An ANOVA test ($F = 17.992, p < 0.001$) also indicated significant differences among the conditions. A post-hoc test found participants performed significantly faster when searching for the targets with AU and PA compared to AS ($p < 0.001$ for both AU-AS, and PA-AS). Nevertheless, no significant difference was discovered between AU and PA.

TOT results showed that participants stayed on the ROI longer with PA (PA: $M = 6.515, SD = 3.18$, BK: $M = 4.43, SD = 1.77$). However, both AU and AS shortened the average time participants stayed on a given ROI (AU: $M = 4.117, SD = 1.268$, AS: $M = 2.277, SD = 1.00$), compared to BK. ANOVA also indicated there were significant differences among the conditions ($F = 17.754, p < 0.001$). The Post-hoc test also indicated PA performed significantly better than the other three conditions (PA-BK: $p = 0.013$, PA-AU: $p < 0.001$, PA-AS: $p < 0.001$). Also, there were significant differences between AU and AS ($p = 0.013$), with participants spending significantly longer on target with AU.

3.4.4 Discomfort

The levels of cybersickness of those videos with AU or PA were reduced (AU: $M = 22.5129, SD = 22.11$, PA: $M = 15.3808, SD = 18.403$) compared to BK ($M = 26.3325, SD = 27.533$), according to the experimental results. For the AS videos, the cybersickness level increased ($M = 80.773, SD = 74.013$). The ANOVA test showed significant differences in their performance ($F = 12.344, p < 0.001$). The Post-hoc test indicated AS was out-performed significantly by the other three conditions (BK-AS: $p < 0.001$, AU-AS: $p < 0.001$, PA-AS: $p < 0.001$). However, significant differences between the AU, PA, and BK, were not detected. The results indicate that participants felt less discomfort when watching the videos with AU or PA, compared to those with AS.

3.4.5 User Preference

In post-test interviews, each participant was asked Q1: Which attention guidance method did they think was the easiest to use?, Q2: Which method was the most uncomfortable?, and Q3: Which in general did they prefer? The totals of each participant choice are shown in Figure 3.7. The results show that 13 out of 24 participants felt AU was the easiest for searching for the ROIs, another 11 chose PA and none chose AS. As for the most uncomfortable method, 21 participants chose AS. When asked about their general preference of the methods, 13 chose AU, 11 chose PA, and none chose AS. It was also noticed that the choice for most straightforward to use was always identical to the choice for general preference. When asked for reasons, several participants who preferred AU stated they felt high levels of presence and less distraction from the narration. Their impressions of the AU were mainly on its being a diegetic feature and easy-to-understand, such as "[The AU was] not an artificial add-on but performed natively by the person in the scene"(P15) and "I can easily understand what he [the host] means and what I should do when I saw him doing a pointing gesture"(P19). Those who preferred PA stated they liked how the arrow stood out from the video and was immediately recognizable. However, several also pointed out that the arrow felt strange because it was not part of the scene but was an add-on.

3.5 Interpreting the Data

In the following sections, the results of the experiments will be discussed together to highlight the main findings.

3.5.1 Implications of Engagement, Enjoyment, and Discomfort

The study results showed that both AU and PA increased engagement and enjoyment compared to the baseline (BK). They also both reduced the level of cybersickness. These results support hypothesis H_1, that the viewer did feel more engagement and enjoyment when watching a 360-degree video shot using AU, compared to PA and AS. Since AU did reduce the level of cybersickness compared to

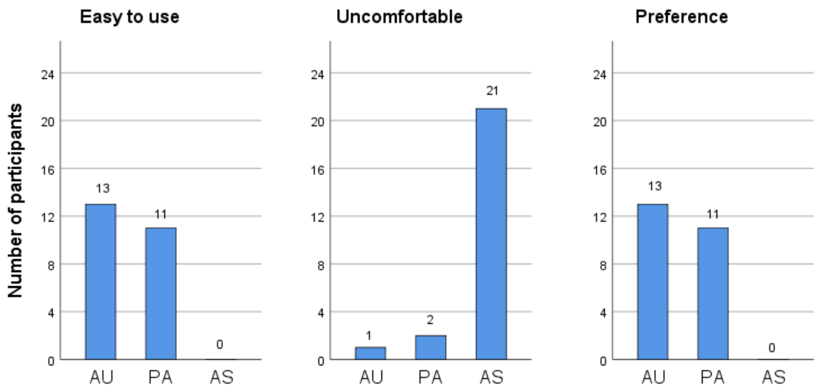


FIGURE 3.7: Bar charts showing the total count of participants on each of the preference choices. Left: Which method was the easiest to use? Middle: Which method was the most uncomfortable? Right: Which did you prefer?

AS, but not PA, hypothesis H_4, which assumed AU will reduce cybersickness when compared to both PA and AS, is only partially supported. For AU, the storyteller implemented them during the production phase, using social cues similar to those in daily conversations. Thus, the cues are naturally embedded in the narrative itself. "Looking" (AU02) and "Pointing" (AU03) also introduced eye contact between the storyteller and the viewer, acknowledging the social existence of the viewer. Viewers felt like the storyteller was addressing them face-to-face, and thus felt more engaged in the narrative. They were also less aware that they were wearing a VR headset and looking at simulated images.

3.5.2 Implication of Memory Effects

The study results indicated that the recall rates of the three conditions with cues showed no significant difference from the baseline. The hypothesis H_2, which predicted that attention guidance would lower the recall rates since acting partially as a distraction in the recall task, was not supported. Nevertheless, distraction was reported by some participants, per their feedback after the exposure to the materials. In the experiment, the participants were informed about the content-related recall tasks and practiced on sample questions before starting to watch the videos, so they were expected to pay attention to the narrative, actively searching for the ROIs mentioned, and remembering their details. Since the details were not included in the narrative by the storyteller, the introduction of cues was believed to have distracted some of the participants from carefully inspecting the ROIs visually. The guiding cues, paired with the acting from the storyteller, both dragged the viewer's attention more towards the narrative. Although the results were not significant and we cannot conclude that guiding cues reduce recall rates, we do recommend that creators take into account a cue's impact on memory effects if they want their audience to retain key information from the material.

3.5.3 Implications of Viewer Behavior

The TTS and TOT were measured to indicate viewer behavior (search and attention, respectively) while watching the videos. Results showed that participants took less time to find the ROIs in the AU and PA conditions compared to BK, and took longer finding the ROIs when AS was used. We can draw a conclusion that the visible cues, such as the AU "Pointing" and the arrow of PA, helped the viewer to locate ROIs faster. This result supports hypothesis H_3, that AU will reduce TTS when applied as an attention guidance technique. The AS was also believed to have elongated TTS because when the AS was applied, the video content was still a blank scene with no visible cue. The scene rotation during the narrative confused viewers, and they thus took more time to understand what was happening before resuming the search task. In the post-test interviews, several participants stated that they thought the AS was a software error and had difficulty understanding what to do before they realized it was part of the system design.

The results of TOT showed a different trend than TTS. Viewers stayed significantly longer on the ROIs after the storyteller mentioned them, when PA was applied as the attention guidance. A conclusion was drawn that since the arrow was constantly visible in the viewer's viewport, it became such an explicit cue that the viewer subconsciously followed it and drifted away from the ROIs much less frequently. In contrast, the AU was only visible when the viewer was looking at the storyteller. This result does not support hypothesis H_3, indicating the need for a further look into the effects of cue exposure time. TOT results also showed that participants dwelt less when AS was applied than in other conditions. One apparent reason is that since AS introduced a higher level of cybersickness, participants were less able to stay focused on the ROIs. In the post-test interviews, several participants also mentioned trying to "fight the system" and subjectively looked away to resist. They felt the system was forcing them to look at something and taking away their free agency to look around, so they looked at the ROIs for only a relatively short period.

3.6 Design Reference for Content Creators

During the material preparation and running of the experiments, we learned how the AUs' properties differed from other attention guidance methods. AUs were applied directly during the capture of content, without the requirement of any post-processing work. Storytellers could compose AUs into their scripts as a part of the content, then naturally perform them when delivering the narrative. AUs have considerable potential and this is a suitable method for creators who need to share and publish their work soon after it is captured, such as documentary videos for science-outreach purposes. In other storytelling scenarios other than using 360-degree videos, AUs are still applicable. AUs can be applied to the actors, humanoid avatars and virtual characters in computer-synthesized 3D scenes, and other immersive environments. They will remain effective as long as there are social communications in the narrative. The usage of AUs in the production stage is expected to help presenters ensure the players follow the storyline and have a higher level of presence in the immersive scene.

However, a trade-off between the AUs' effects on attention (i.e., the recall rate and TOT) and the effects on the levels of both engagement and enjoyment is also noteworthy. As seen in previous analyses, AUs increased the participants' levels of engagement and enjoyment, and directed them faster to the ROIs when applied to videos. However, they also distracted them from paying attention to details of the ROIs. The participants recalled the details less than expected. This was also reinforced by the feedback from participants during the post-test interview, as some preferred AUs for their diegetic characteristics and others preferred PA for its outstanding contrast and substantial implications. This could be a vital design reference for content creators, as one needs to make decisions by considering the effects of AUs, the desired content, and the expected response and results from viewers. AUs can be useful for scenarios like scientific presentations or other face-to-face conversations where a natural social atmosphere is essential. But in other scenarios, such as training and educational applications, where the recall rate and transfer of knowledge are more emphasized and other aspects such as comfort and enjoyment are less dominant, using AUs as the only attention guidance may not be sufficient, and adding extra cues will be recommended. In sum, AUs (based on human acting) can be effective for attention guidance in CVR, based on our study results and the analysis. They also help reduce the risk of breaking presence and mitigating sickness because of their diegetic characteristics. Hence, RQ_1 from Section 1.2, whether the social cues from human characters (actor, storyteller, etc.) in a CVR scene can be used as attention guiding techniques to solve NP and FOMO issues in CVR, is answered.

3.7 Conclusion

In this research, Action Units (AUs) were proposed as a set of directorial cues, aimed at solving the fundamental challenge CVR content creators are facing, the *narrative paradox*. The results from both the pilot study and the formal user study indicated that AUs could effectively guide the viewer's attention when applied, while maintaining the advantage of being diegetic and easy to use. Creators can adopt the AUs onsite to direct viewer attention in an immersive storytelling experience. In the following study, I will discuss another challenge CVR creators face in production: the viewer's intention to interact, and expectation of agency since they are now immersed in the story world.

Chapter 4

Viewer Participation and Interaction in CVR

4.1 Introduction

In Chapter 3, previous research was introduced [153, 19, 81, 82] aimed at solving *narrative paradox (NP)* in CVR. Their focus was on how, over time, CVR creators noticed a change of how they thought of the viewer's perspective from cinema to CVR. They then used various attention-guidance cues to support the viewer's narrative engagement. Action Units (AUs) were also presented as a diegetic guidance cue for easy application during the production stage. These solutions, however, have mainly focused on the creator's role, as opposed to viewer agency. Viewer's interaction, and viewer intention to participate in the narrative progression, was not addressed. Since the viewer is now seen as also being a part of the story world, creators must also consider viewer interaction when they design their CVR experiences. One of the essential design decisions for creators, as stated in detail in Section 2.3, is to define a proper level of interaction (including the complexity of input, and the depth of consequence of viewer input). A *Continuum of Interactivity* (Figure 2.9) was proposed to gauge and locate a balanced point where CVR with viewer interaction needs to sit among other types of interactive experience, if creators acknowledge they will move on from the simple passive-watching type of 360-degree videos.

In a summary, interactive CVR storytelling has inherited from immersive storytelling, thus still fits the preferred user scenario for CVR as well as "Swivel-chair VR". CVR viewers, although given the capability to interact with the system, are still inclined to enjoy the narrative with a "lean in" mindset (the viewer feels attracted to the narrative and temporary forgets the physical world around him/her, such as when watching a high-pace thrilling action movie) instead of "lean forward" (as the viewer feels actively involved in the narrative and subjectively intends to control how the narrative will progress, such as in a role-playing video game) [158]. Those factors position CVR experiences near "limited" or "medium" on the Continuum of Interactivity. What is more important than simply choosing an interaction level, is that creators should give interaction a "contextual meaning." This means they need to consider the tasks a viewer wants to carry out by interacting, the impact on the narrative each of those tasks will have, and the input method the viewer will use to perform those tasks.

4.2 Two Interaction Elements - Control and Visibility

Involving viewers in the control of narrative progression is a popular topic in interactive storytelling research [157]. 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 [154]. On the other hand, researchers have also discovered that for cultural- or education-related content, a clear structure helps in audience understanding and retention [83]. Researchers who focus on games have also found that, in particular, explicit narrative progression designed by the creators has a positive effect on declarative knowledge acquisition [56].

Viewer awareness of interaction also impacts the viewer experience of storytelling [122]. 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 [154]. A system that is responsive to viewer input will also increase viewer involvement and immersion [126, 133].

Concern has also been raised by some researchers, pointing toward the consequences of choices. Rezk et al. [122] cautioned that if viewers are given explicit choices during storytelling, they are likely to hesitate when faced with too many options. They pointed out that the viewer 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 "implicit control," where the viewer still participates in the progression of the narrative, but is not *explicitly aware* of making choices. To achieve this, one group implemented a system that monitored viewer behavior during virtual museum tours at predefined spots, and responded to this behavior by making unannounced narrative choices over branches [60]. Another researcher recorded the player's actions in a game, used the data to determine his/her overall contextual intention, and then presented a matching ending from several parallels [147]. Thus, the focus of the study described in this chapter is narrowed to a two-element combination, control of narrative progression (creator vs. viewer), and the visibility of interaction affordances (implicit vs. explicit interaction).

4.3 The Experiment

The main focus of this study is how viewer interaction can be enabled in interactive CVR, such as those that use 360-degree videos. However, there are various methods for different user scenarios, defined by factors including the content, the emotions that the creator wants to invoke in the viewers, and the type of interaction viewers will use [154]. Due to resource limitations, the first constraint on this study is the *applicable content*, and we limit the content to those that:

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.

The structure of the employed story 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, a between-subjects experiment was conducted to compare the viewer experiences between design variants. The content was an in-house guided tour through a series of 360-degree videos, with more details about the production process presented in Section 4.3.3.

In the experiment, four conditions were set up for the interactive and immersive storytelling experience:

1. Creator control (CC);
2. Creator control, with randomly rearranged segments (CR);
3. Viewer control, with implicit input (VI); and
4. Viewer control, with explicit input (VE).

The structure of the content remained unchanged between conditions and will be illustrated in detail in Section 4.3.3. 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 4.3.4. The CR condition was included to observe if the viewer's experience of this specific story would change when the order of the segments was different from what the creator initially intended. Each condition's effect on viewers was measured, including the level of engagement with the content, enjoyment from the experience, and the general user experience. A series of content-related questions were also designed, 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 in the transfer of knowledge.

4.3.1 Research Questions and Hypotheses

In this study, the research questions were:

- RQ_1: For the hub-and-spoke structure, which type of control pattern will bring a higher level of engagement and enjoyment: creator control or viewer control?
- RQ_2: 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 or explicit control?
- RQ_3: Between implicit and explicit control, which one will lead to a better memorization of the content?

Hypotheses were formulated corresponding to the research questions based on previous research on the viewer's role and behavior in narrative CVRs. Firstly, the viewer-controlled modes (VI & VE) were expected to bring higher levels of engagement and enjoyment. As in an immersive environment, agency increased the level of presence and brought a deeper feeling of being directly involved in the story scene, as well as a stronger feeling of fun [47]. As has already been verified by previous research, the viewer's role in immersive media differs from that in a regular movie. The viewer felt they were a character in the scene and formed the expectation that they had some influence over how the guided tour will progress

[126, 133]. Secondly, as mentioned in the previous section, explicit choices for the viewer made them think about the potential consequence of each choice, thus possibly breaking immersion and deteriorating the general experience [122]. The implicit control method (VI) was also assumed to have less negative impact on the viewer's general experience of 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, the condition VI was expected to also lead to a better result in the viewer's memory test of the content, because in the viewer's perception, they were *passively* watching and more focused on the content. In summary, the following hypotheses were formulated:

- H_1: Compared to creator control, viewer control will lead to higher levels of engagement and enjoyment with the experience;
- H_2: Compared to the explicit control-based method, implicit ones will lead to a more positive viewer experience;
- H_3: Implicit control will lead to better performance on memory tests.

4.3.2 Apparatus

The interactive system was implemented on a computer running 64-bit Windows 10 Professional with a 3.2GHz i7 processor and a GeForce RTX 2080 graphics card. The computer also recorded viewer behavior data and ensured the smooth playback of the 5.7k 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 4.1A and B.

4.3.3 Materials

The study was approved by the Human Research Ethics Committee of the University of Canterbury. For this experiment, a series of 360-degree videos were captured in a laboratory room, then assembled into a virtual guided tour. In the lab space, several large experiment installations were placed against each wall, as shown in the top-down view of the space in Figure 4.2. 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 installation. The real world visit was duplicated into this study by capturing 360-degree videos of it, as the spatial layout of the *space*

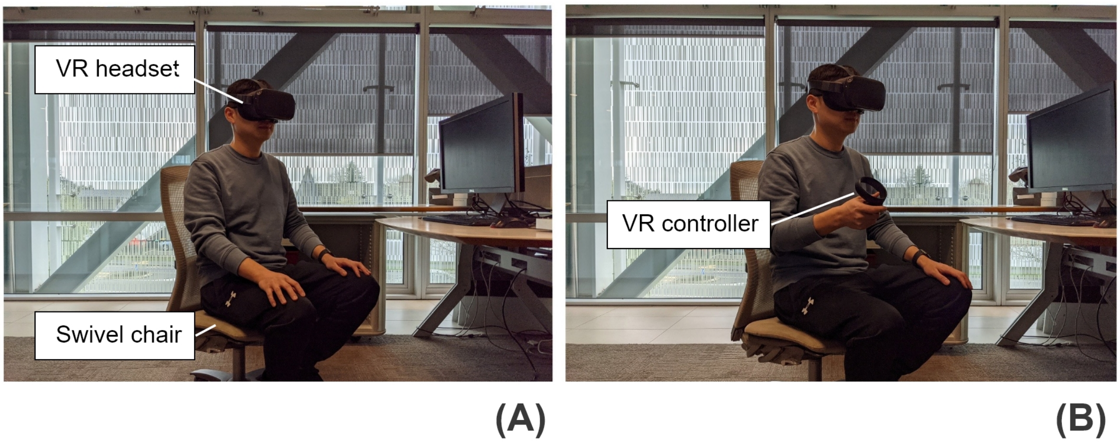


FIGURE 4.1: Two types of participant setups. (A) For conditions CC, CR, and VI, the participant does not use a controller but wears a headset and sits on a swivel chair, and has the freedom to look around in the scene. (B) For condition VE, the participant also holds a VR controller in the right hand and uses it to make explicit choices.

and the *installations* matched 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.

The 360-degree video clips were then captured when the host was introducing the installations at each designated spot, using an Insta360 ONE R 360-degree camera at a resolution of 5.7k (5760 x 2880). The camera was mounted at eye height on a tripod so the viewpoint was fixed in each segment. The positions of the camera and storyteller (host) for each segment were carefully chosen so they are not blocking views, as shown in Figure 4.2. The distance from the host to the camera was kept the same (2m) at each location. Thus, 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, the "Introduction" clip was captured first 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 it. The narrative structures of each segment at the spokes were also scripted to be similar in terms of running time. A total of five clips were captured, 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,

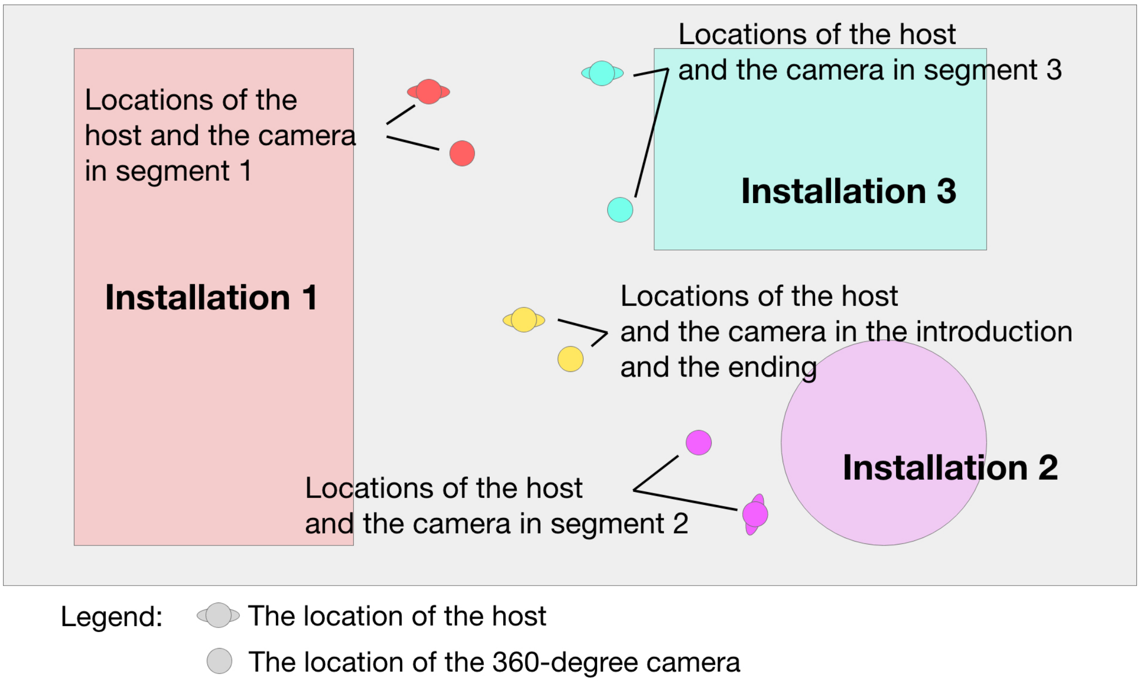


FIGURE 4.2: The layout illustration with a top-down view of the lab room in which the footage was captured. The camera was initially placed in the middle of the room (the "hub") when capturing the introduction and ending clips, and placed near each installation when major segments (the "spokes") were captured. Each spoke is shown in a different color, and the placement of the camera and the host when capturing each spoke segment are also marked. At all times during the tour, all spokes were visible to each other.

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.

4.3.4 Implementation of the Interactive System

In this study, the interactive storytelling system consisted of two components: (1) a video player to present the 360-degree video clips to the participants via the VR headset, and (2) a mediator component to deliver the designated order from the creator (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.

CC: in this condition, the five clips were loaded into the library of the system following a specific order (the Introduction, three main segments, then the Ending).

The system played those 360-degree video clips one after another. There was no viewer input in this condition.

CR: similar to CC, the system played those 360-degree videos without viewer input. The only difference was that the system randomly rearranged the three main 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 the participant was watching the 360-degree 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 4.3. 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 dwelt on a gate longer than a threshold (set to 3 seconds in the system in this study), the moderator would decide that the viewer might be interested in this object and pull the corresponding segment to the top of the playlist as the next one to play after the current clip. The gate threshold could be triggered multiple times along the playback experience until the currently-viewed clip played 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 moderator executed the entire process in the background, and the viewers were not aware of it at any moment.

VE: A hand-held laser-pointer method was implemented in this condition to enable the viewer's explicit interaction. It used a simple point-and-activate mechanism [130] to help reduce the unnecessary learning workload of the viewers. Based on the gate design from VI, those invisible gates were replaced with visible cards, as shown in the screenshot in Figure 4.4. Each card had text showing the segment it represented, plus the time length of that video segment as an extra aid for the viewer's decision-making process. Viewers holding the controller could then point at the cards with the laser pointer and pull the trigger to make a selection, as shown in Figure 4.5B. A popup message would temporarily appear in the viewer's field of view (FOV) to confirm the choice. The cards were set to appear when the progress of the current video clip reached 70%, so they were not always shown to cause distractions, nor appearing too late and leaving the viewers with too little time to make selections. Since the interactable cards were overlaid on top of topic-related

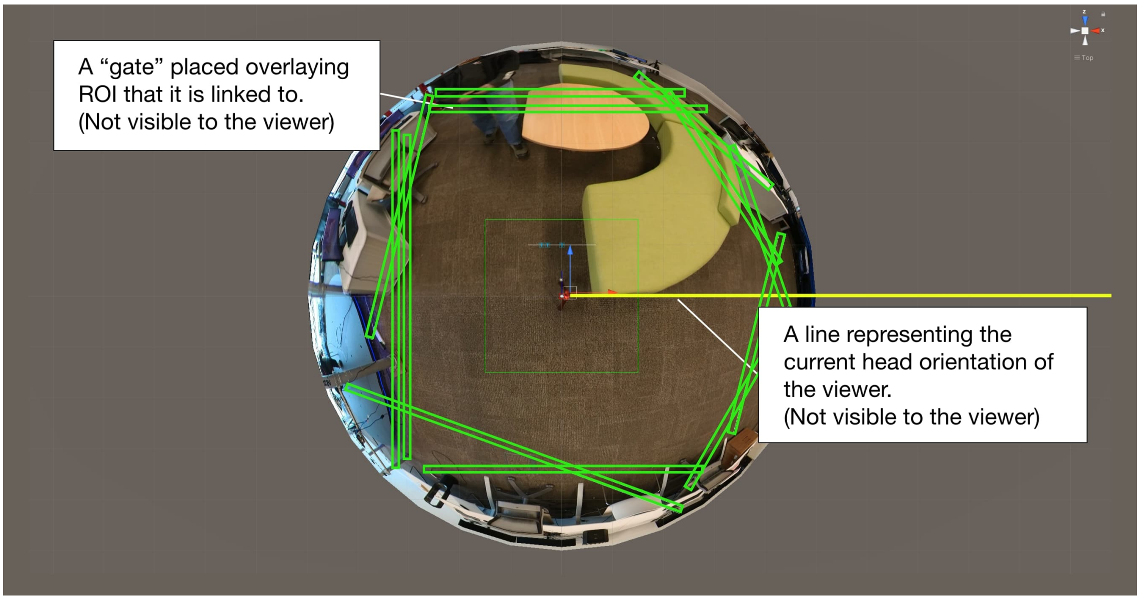


FIGURE 4.3: A top-down view of the 360-degree spherical screen showing the gates scattered around covering the ROI behind them, waiting to detect the viewer's head orientation vector (which is represented as a thin yellow line in this screenshot) to collide with them, registering gaze dwelling over the ROIs. The image is a top-down view, so the gates are the green rectangles. Technically, in the 3D scene, they are thin boards standing in front of the ROIs, and serve only as detection mechanisms and are not visible to the viewers.

objects and were scattered around the scene, popup messages were programmed to appear in the center of the viewer's FOV as a reminder for when cards were available for selection, reducing the FOMO. The default order from CC would be used if the viewer did not make any choices before the video segment ended.

In all four conditions, a Helper heads-up display (HUD) was set up and constantly visible at the lower-left corner of the viewer's FOV, as shown in Figure 4.5. This aided the viewer in being aware of the current segment's progress and the entire tour's progress. Thus, the viewers would not get lost in the tour and experience difficulty recalling the content of a segment. Instead, they maintained an awareness of the progress and pace.

4.3.5 Measures

When designing the experiment, two essential facts emerged during piloting. First, in the VI condition, the viewer was unaware that the system was monitoring her head orientation and making alternations to the playback progress. Second, since a participant only experienced one of the four conditions across CC, CR and

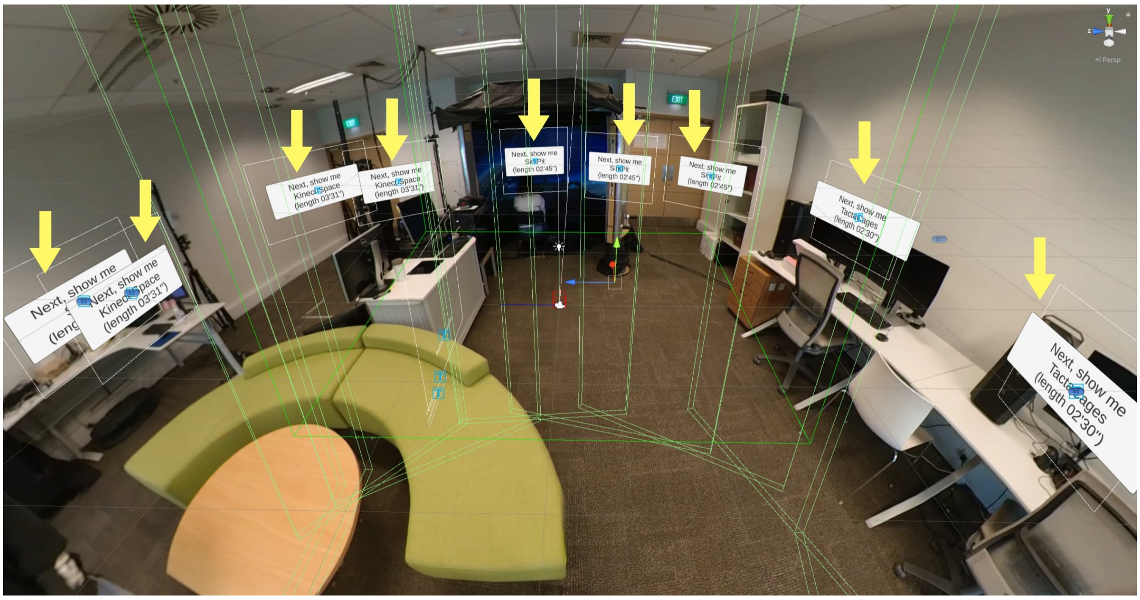
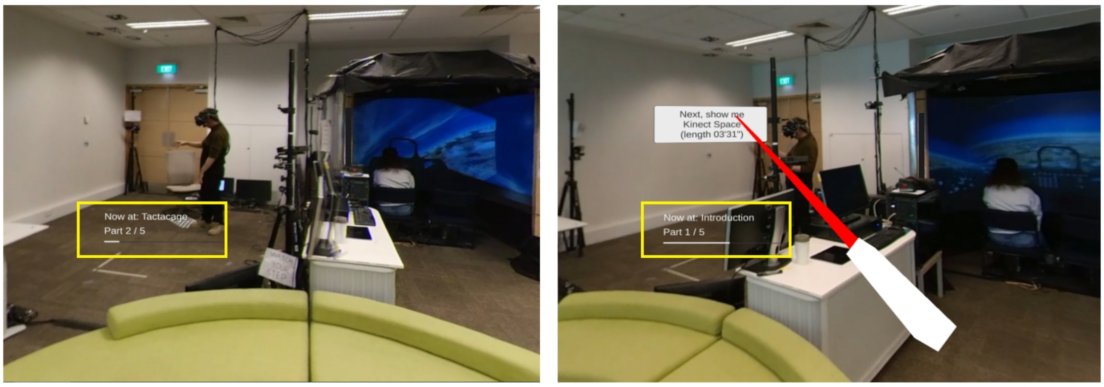


FIGURE 4.4: A screenshot from the Unity editor showing the cards distributed around the scene. They were interactable with the virtual controller in the hand of the viewer. In this figure, all of the cards are visible and pointed to by yellow arrows for illustration purposes only; in the actual scene the yellow arrows were not visible. Also, the cards would not show simultaneously as in this figure. They were programmed to show at certain locations and times according to the progress of the current segment.



(A) (B)

FIGURE 4.5: Screenshots of what participants see in four conditions. (A) For conditions CC, CR, and VI, the participants sat and watch the 360-degree videos. The Helper HUD was 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 was not visible in the experiment. (B) For condition VE, the participant also saw a virtual controller with a red laser pointer in her right hand. There were also interactive cards with texts visible in the scene at certain times (like in the screenshot). The participant then used the virtual controller to point and make explicit choices.

VI, participants were also unaware that some conditions might impact the narrative progression while others would not. Taken together, the viewer's subjective experience across the conditions would not be equivalent unless they were told afterward which one they had experienced. Thus, assessing and comparisons of the viewer experience by explicitly asking about their preferences was not a viable option in this study. Instead, their subjective feelings were measured by regarding all conditions as generic immersive storytelling experiences. Design preference was also derived by comparing the participants' overall experience across the conditions.

For subjective measures, each participant's level of engagement with the content, enjoyment of the experience, and general usability experience were assessed. The engagement was measured by the widely-applied User Engagement Survey Short Form (UES-SF) [107]. The operational guidelines provided by Schmitz et al. [134] was also used when applying this measurement. Enjoyment was measured using the 12-item scale provided by Ip et al. [62]. They used this scale to measure learner enjoyment after watching a series of Massive Open Online Courses (MOOCs) in 360-degree videos, similar to our study setup. A simple three-part evaluation form, provided by Shah et al. [137] in their study evaluating a 360-degree video playback system, measured the general user experience towards the system. After removing one non-relevant item, it was applied to see if the participants found the system easy to use, both with the implicit interaction and the traditional point-and-activate methods.

Content-related questions were used to assess participants' memory performance from each condition. Three multiple-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 (such as Q4-1: *In this immersive video, the guide described those things in the Vision Space: (choose all that you think were in the video), 1 TactaCages; 2 SimPit; 3 Aviator Trainer; 4 Kinect Space; 5 Voxel Portal.*). Three single-choice questions then asked the participant to verify some facts visible in the scene (such as Q4-6: *In the Kinect Space, the person using the system was holding VR controllers in his hand: 1 True; 2 False.*). 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 insight into how much the participant remembered from the tour and understood the introductions. The questions did not require any reasoning or deduction, and so did not require any background knowledge.

A plugin was also added to the 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. The heat maps were then used to identify whether the viewer showed a relatively active exploratory behavior or mostly passive, static watching.

4.3.6 Experiment Procedure

Four conditions (CC, CR, VI, and VE) were compared by applying them to the same five-segment virtual guided tour. Since the experiment had a between-subject configuration, and a one-way ANOVA for fixed effects was chosen for data analysis, G*Power analysis was conducted to determine the target group size. Results indicated that there was a 95.8% chance of correctly rejecting the null hypothesis of no difference between the test groups and control group, with a total of 40 test subjects. Later in the actual experiment, 44 participants were recruited (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 daily. 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, consent was obtained from each participant. They were then introduced to the "virtual guided tour" and explained how they would experience it with the HMD and the Swivel-Chair VR setup [153]. 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 (CR) and each viewer's interaction (VE) or behavior (VI) during the experiment. When the tour ended, the participant was asked to remove the HMD and move on to complete the questionnaires.

After the questionnaires, a semi-structured interview was conducted 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.
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, an extra question was added asking about their motivations for making choices using the controller via the interactive cards. The entire session took approximately 35 minutes for each participant.

4.4 Results

The data from the questionnaires were analyzed 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 4.1. The overall results are summarized and plotted in Figure 4.6. 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 were both multiple-choice and single-choice questions, the scores were calculated using the following rules: for a single-choice question, the participant scored one point for the correct answer; otherwise, zero points. For a multiple-choice question, the participant scored a full mark (five points) if s/he selected all and only the correct items, losing 0.5 points for missing one of the correct items and a full point if a wrong item was selected.

TABLE 4.1: The mean values of the results for Engagement, Enjoyment, General User Experience and Memory Performance, for each condition.

Measure	CC	CR	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

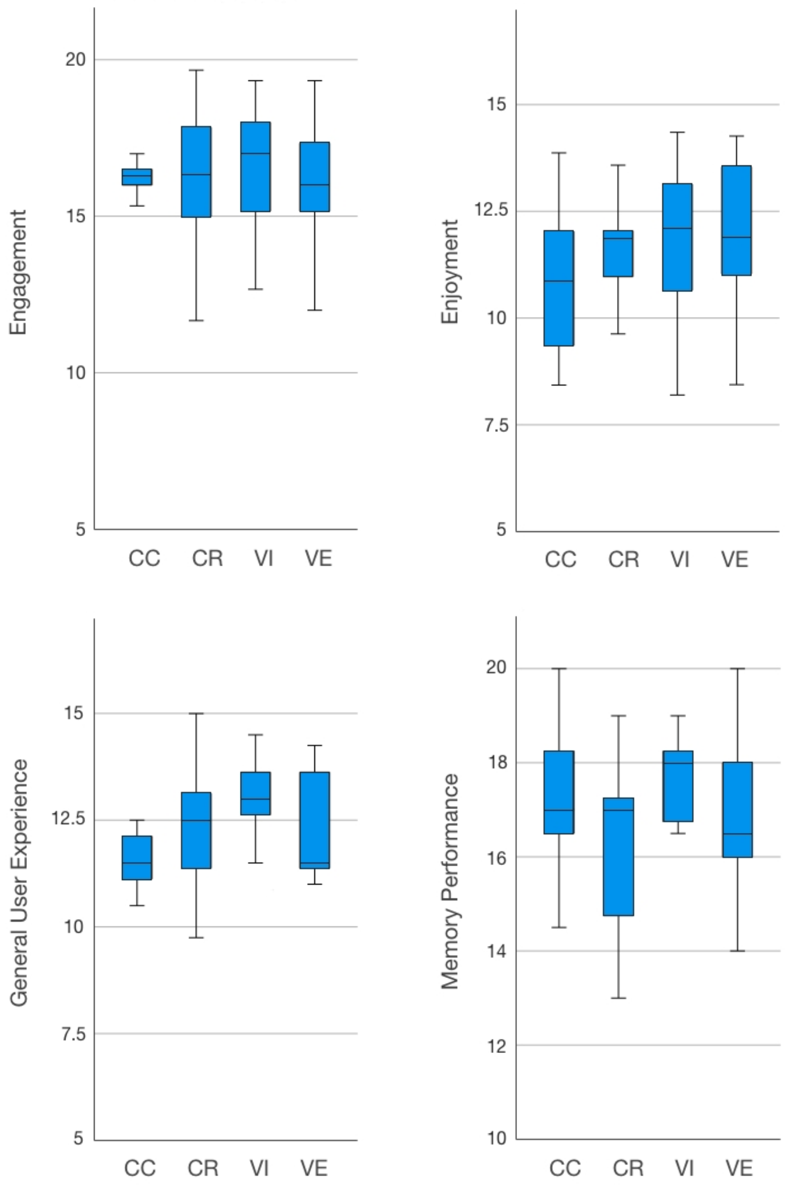


FIGURE 4.6: Boxplots summarizing the results of levels of engagement for each condition (converted from the UES-SF items), levels of enjoyment, general user experience scores towards the system, and the memory performance scores from the content-related questions in terms of medians, interquartile ranges, minimum and maximum ratings. Top row, from left to right: Engagement, Enjoyment. Bottom row: General User Experience and Memory Performance.

4.4.1 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 (CC: $M = 15.97, SD = 0.327$, CR: $M = 16, 30, SD = 0.608$, VE: $M =$

$15.58, SD = 0.777$). However, the ANOVA test indicated no significant differences among the conditions ($F = 0.344, p = 0.794$).

4.4.2 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 (CC: $M = 10.78, SD = 1.813$, CR: $M = 11.59, SD = 1.211$, VI: $M = 11.79, SD = 1.774$). However, the ANOVA test indicated no significant differences among the conditions ($F = 0.973, p = 0.415$).

4.4.3 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 (CC: $M = 11.52, SD = 1.207$, CR: $M = 12.38, SD = 1.514$, VE: $M = 12.31, SD = 1.260$). However, the ANOVA test indicated no significant differences among the conditions ($F = 2.269; p = 0.092$).

4.4.4 Memory Performance

Looking at the memory performance scores from participants of the four conditions, one can notice that those who experienced the CC condition scored higher in the memory test (CC: $M = 17.27, SD = 1.664$), compared to the other three conditions (CR: $M = 16.22, SD = 2.029$, 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 ($F = 0.662; p = 0.580$).

4.4.5 Behaviors of Attention and Focus

Each participant's attention over the entire scene was also recorded 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 4.7. The area covered by the colored clouds are areas in the scene that had been looked by the viewer over the entire process. The color represents how long the viewer had dwelt over a certain spot. The cloud started from blue, and turned to green, then red as the viewer stayed on the spot longer.

Comparing across the conditions, the heat maps from the VE group, showed larger scanned areas recorded on the canvas than those of the other groups, indicating that participants who experienced the VE condition showed a higher tendency toward exploration (actively looking around). Four examples were extracted from each condition and grouped together to clearly show such difference. In Figure 4.8, four heat maps were all taken from segment 3 but from different conditions. We can clearly see the colored cloud spread more across the scene, compared to other three conditions, indicating a more actively-exploring behavior from the viewer on condition VE. The dense red colored areas are also larger in condition VE, indicating the viewer was actually looking carefully at wider areas, instead of unconsciously scanning them. Comparing the segments under the same condition, participants showed more exploratory behavior in the first segment (Introduction) than in the other three main segments where the host introduced details of each installation. The participant behaviors are further discussed in Section 4.5.3 by looking deeply into the heat maps and participant answers during the interviews.

4.4.6 Subjective Feedback from the Interviews

The interview results were transcribed and analyzed using a thematic approach. Topics were identified and generalized into two high-level themes, with five sub-level topics. Below the themes from the interviews alongside the interpretations are presented. A further analysis and discussion by connecting the themes with the conditions participants experienced are presented in Section 4.5.2.

Theme 1: Impressions of the Virtual Tour Experience

The first theme concerns the participants' impressions of the virtual tour as a general experience. The viewers mainly paid attention to the experience itself, the content presented, or the system during the virtual tour. Generally speaking, being immersive helped participant engagement with the experience.

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 like a 360-degree*

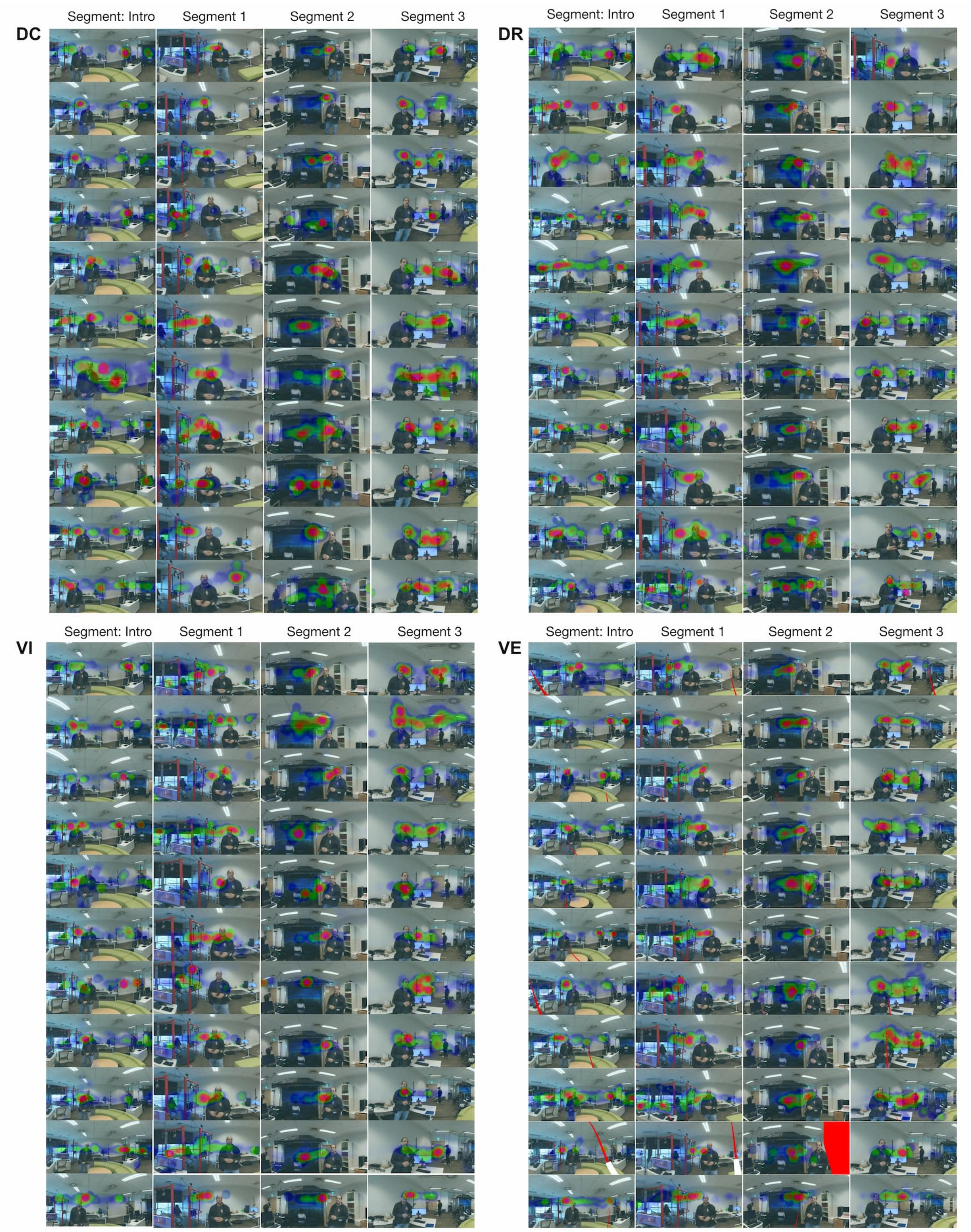


FIGURE 4.7: A grid summarizing the generated heat maps, grouping by the conditions. The images are zoomed into the central part to clearly show the gaze dwelling, from the original equirectangular projections. Each sector has 4×11 (or 44) heat maps collected from one condition, four segments. In each row, four heat maps represent one participant's behaviors. Each column represents the data from all 11 participants from this condition, on this specific segment. In CR, VI, and VE, although participants watched segments 1, 2, and 3 in random orders, the images here are arranged in CC order for easy comparison.

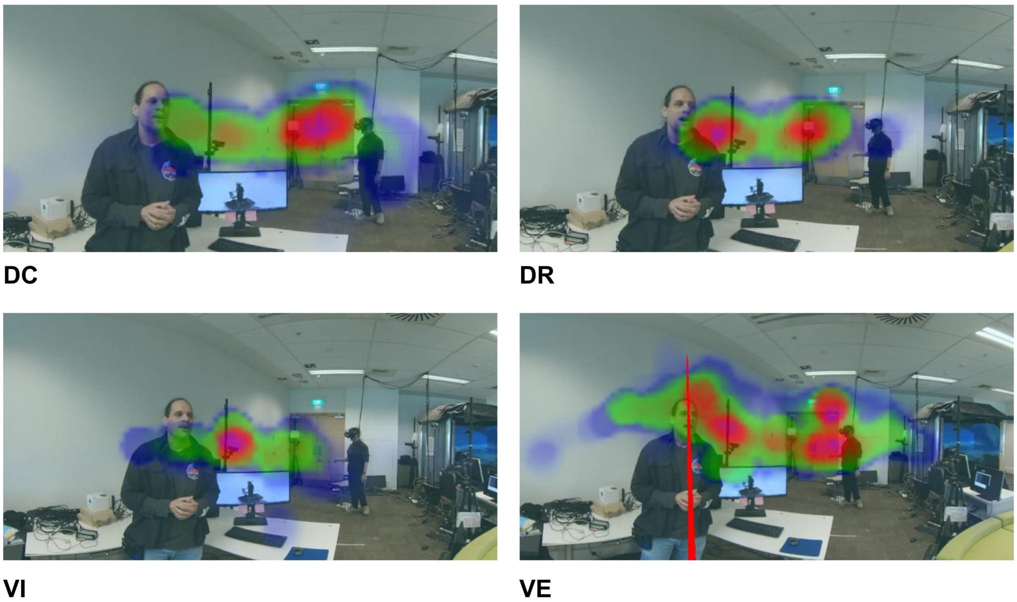


FIGURE 4.8: Four enlarged examples of the heat maps. They were all taken from segment 3 but from different conditions to show how the viewer's behavior differentiate from one condition to another.

experience, an immersive tour...I can look around". The characteristic of 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."*

Feeling like 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]..."* I 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.

Impressions 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. These are also related to another theme discussed later, but a few also pointed out their expectation of design improvement: *"I do not want to use the VR headset frequently...not for long* *time use because it is uncomfortable...I was unable to walk around in it".*

Theme 2: As a Medium for Transfer of Knowledge

The previous theme explained how the participants saw the virtual tour and their impressions. With this theme, the virtual tour's performance on serving as an immersive tool to assist in learning about the lab and the installation presented in the materials, was inspected.

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 in, such as *"The [name of the installation] is quite impressive because I think it is interesting/cool..."*. Since the participants were 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.

Immersion 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."*

4.5 Interpreting the Data and User's Feedback

The main findings and implications of the results of the experiment will be discussed in this section. Insight into viewer behaviors from the interviews will also be presented, together with observations of the participants while they were watching the 360-degree videos, and the heat maps across each segment of the tour.

4.5.1 Objective Measures

Since there were no significant differences detected between the objective results from the four conditions, the hypotheses (H_1 & H_2) were not supported; higher levels of engagement and enjoyment were not found in those who experienced the conditions with viewer control (VI & VE). Data analysis detected no significant differences among the memory performance scores from the four conditions, so H_3 is also not supported. These indicate that participants did not react to the four conditions differently on the objective responses. Alternatively, they did not objectively become aware or care about all four conditions applied to the virtual guided tour experience. There are at least three possible reasons for this:

1. The participants were unaware of any alternative other than creator 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). At the same time, 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 in the next section.

2. Constraints were set on content in advance so they 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. Participants are expected to have higher levels of enjoyment and engagement if the content was well-prepared and creatively made by professionals.

3. The amount and granularity of viewer interaction on content also contributed to the observation. In conditions VI and VE, viewer interaction was only used to drive the narrative progression. More specifically, among various elements within a storytelling experience, the viewers were allowed to control *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, viewers could have been allowed to interact much more than simply picking segments on a playlist in other interactive experiences. One might interact with an object within the scene, or 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 can interact with more specific and visible elements in the scene (and maybe also linked to narrative progression/branching).

4.5.2 Viewer's Main Impression and the Role of Control

In the previous section, the comments from the interviews, common themes, and sub-level topics were summarized. The transcription of each participant's responses were matched to particular topics and we counted the times each topic was mentioned. The conditions each participant experienced were then inspected and the numbers separated into two groups, creator control (CC & CR) and viewer control (VI & VE). The final results are shown in Table 4.2. The total count of Sub-level Topic 3 of Theme 1 is not further divided, as it is not condition-dependent.

TABLE 4.2: 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	CC & CR	VI & VE
Theme 1: Impressions of the Virtual Tour Experience			
Being Immersed in the Virtual Environment	31	13	18
Feeling like a Real Guided Tour	18	13	6
Impressions of the Technology	37	-	-
Theme 2: As a Medium for Transfer of Knowledge			
Actual Knowledge Gained	24	14	10
Immersion as an Advantage for Learning	20	8	12

For Theme 1, the viewer's primary impression of the immersive experience was believed to have changed according to whether or not the viewer had been given control. When a viewer is given control over the narrative progression and notices her agency within the experience, her primary impression/recall will mainly be the feeling of "interactivity" or "being immersed," instead of focusing on the content. This conclusion is supported by the number distribution between CC & CR and VI & 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. "Viewer control" leads users to remember more about the *experience*, rather than the *content*.

4.5.3 Exploratory Behaviors

On the heat maps in Figure 4.7, clustered areas (the area a viewer focused on for the longest time) and the size of the cluster (a small cluster if the viewer rarely moved her/his head when looking at that area) were both carefully inspected. The focus point of most viewers fell on the host during the virtual tour because they were reminded to pay attention to the content as there would be questions about it at the end. Most viewers took it as a "test" and tended to listen carefully 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. Several participants also chose to look at the ROIs. The reasons suggested from the interviews were mainly (1) they noticed some topic that they were interested in and would like to know more about, and (2) they were attracted by something in the video and wanted to have a closer look (zooming 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 (a VR lab), thus were driven by curiosity to look around actively. This was also verified by their answers in the background questionnaires. Most participants who showed "curiosity" behaviors had answered "never" or "only a few times a year" when asked about previous VR use or watching 360-degree videos.

This observation motivates a conclusion that participants' tendencies to watch passively or actively look around could have been affected by their preconceived impression of the experience. During the experiment, before starting the viewing sessions, the experience was described as "a virtual guided tour made from a series of 360-degree videos." The actual mechanism behind the scene was not mentioned because the participants were not supposed to be aware of the differences between conditions. They were also notified that there were "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." They showed minor exploratory behaviors in the main segments. But first impression is not the only case where this occurs. Personal experience and interests also contribute to exploratory behaviors [65].

4.6 Discussions

There are some less-perceptible viewer behaviors and intentions, which are not as explicit as losing track of key story elements, and are all based on the data and interpretations presented above. The first two sections focus on viewer needs for pace control and their motivations and hesitation when faced with choices. The third section transforms this discussion into implications for CVR creators.

4.6.1 The Non-stop Flow of Time and the Control over Pacing

In most cases, the viewers' actual perception of the experience was different compared to the conceptual scenario described above [154]. During the experiments on interactive CVR with various control types and methods, viewers showed their needs for pace control. Several participants of the VE condition reported in the interview that they felt there was not enough time to make selections. Some saw the cards appearing in the scene, but did not have enough time to consider the consequences and make a selection. Such observation 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 creator 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 in the viewer's hands could be crucial. In other words, when the storytelling process encompasses viewer interaction with the narrative, it also enables the viewer to perceive him/herself in a role involved in the story. Therefore, the control of story progression should also be appropriately handed over to the viewer, at least partially, to match the agency a viewer now has.

4.6.2 Making Choices, or Not Making Choices

During the second user study (see Chapter 4), an extra question was added during the interview for participants in the VE condition to ask about their motivations for making their selections. This was based on my observation that in the VE condition, there were two types of users. Some actively made selections when the cards appeared, and others did not make any choices along the tour even though the controller was in their hands. The following are discussions on viewer's choices after dividing them into separate groups.

For those who made selections, the responses indicate 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 a vital 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 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 either differed from the original design, or they went beyond the instructions for 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 not foreseen 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 the selection mechanism's consequences, they would have chosen according to whichever segment interested them most.

The confusion was assumed to be related to the wording on the cards, as the labels read "Next up: [the name of the ROI]," which may have confused some participants. However, this raises the possibility of a mismatch between a viewer's perception of the consequence of control, and the actual consequences. As mentioned by Carstensdottir et al. [24], the creator of an interactive experience must consider how the viewers establish a specific mental model in their minds when a control or input method is put in their hands. Design choices for interactive elements needed further investigation in this case.

4.6.3 Implications for CVR Creators Considering Viewer Interaction

Another set of CVR production guidelines was derived by comparing participants' interview responses in the different conditions. We have learned that a 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 sole creator control, implicit viewer control, or explicit viewer control. This guideline is also a part of the Adaptive Playback Control (APC) framework, assisting CVR creators in choosing proper viewer interaction designs for projects with different purposes. The framework will be introduced with details in next chapter.

The necessity of having a balance between *the viewer's natural tendency to explore* and *incentives to invoke exploratory behaviors in viewers* was also noticed. The former was observed due to 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 observations from the participant behaviors during our experiments. The latter was derive from the combination of both quantitative data and qualitative observations. The viewer's tendency to explore in an immersive environment actively 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. During the experiment, while the participants were carrying out the task of "I need to pay careful attention to the speech from the tour guide," they 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. Since no sufficient data was collected in the second user study (see Section 4.4) to fully understand this mechanism, only preliminary insights are presented 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*. Such contrast made viewers unclear of the consequences behind their actions, making them step back into a passive-viewing mode. To encourage viewer interaction, CVR content creators need to be aware of viewers' intention to explore actively, and treat that as a firm grounding in immersive storytelling. They also must help the viewers strike a balance between being willing to act and being afraid to act, via story structure design and system design. On story structure, one example would be the *virtual tour* content used in the second user study, which was designed with a flexible structure that the actual sequence a viewer choose to visit the three spots does not affect the main narrative of the entire tour (see Section 4.3.3). On system design, a good example would be the *pace-control* discussed back in Section 4.6.1.

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

1. Viewer control is recommended for immersive storytelling projects that focus on providing an experience of participation and feeling of agency. For other projects, where learning and knowledge transfer is the priority, the creator might want to consider leaning more heavily on creator control. While this might 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 utilize best the viewer's natural tendency to explore actively. Interaction design with implicit input 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.

To sum up, although viewer interaction comes as a natural tendency in CVR, both the level of control given to the viewer and the visibility of the interactors should depend on the content and its purpose. It is recommended as a design rule for all CVR creators, as well as a response to RQ_2, about the design considerations CVR creators need after acknowledging that the viewer as a participatory role in the experience, stated in Section 1.2.

4.7 Conclusion

In the study, the introduction of viewer interaction into CVR storytelling was explored, by focusing on a two-element combination: narrative progression, and interaction visibility. This formal user study evaluated four combinations of 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. Semi-structured interviews were also conducted to assess their subjective feedback. The objective data did not reveal any significant differences. However, by combining both objective and subjective data, insights into viewer's role in CVR and their behaviors were generated. CVR creators are suggested to design control modes and interaction visibility based on the original purpose of the storytelling experience.

Compared to the previous study, this one was built upon acknowledging the viewer's new role in CVR storytelling (they can actively look around in the story world). In this study, the viewers were given more extensive capabilities to interact with the narrative (enabling them to participate in the narrative progress). However, for the matter of "giving the ability to interact with the story" and "actively participate" in CVR, the viewers could not accurately tell if interaction was enabled or explicitly given, or not, during the experiment. This observation is supported by the fact that the objective measures from participants all failed to show statistically significant differences for the different setups in the experiences. It is subjectively inferred that the design differences in the system's interactivity may not have been well described and perceived by users. To further evaluate whether *involving viewers in the progress of storytelling* enhances the experience and delivers the intended reactions expected by creators, more care must be taken to balance the control provided to both the creators and viewers. This is explored next in the thesis.

Chapter 5

Adaptive Playback Control - A Comprehensive Framework for CVR

In Chapters 3 and 4, the results of two studies exploring effective attention guidance for CVR viewers (stated as RQ_1 in 1.2) were presented. In this chapter, learnings from the studies, including the use of AUs in CVR production and the addition of viewer interaction to CVR experiences, are discussed in more depth. These techniques are then combined into an integrated framework of CVR production guidelines, which covers a complete workflow for producing a CVR experience with viewer interaction. Responses to the research questions posed earlier are also probed from a more holistic perspective in this chapter.

5.1 Knowledge Gained from User Studies

In the first study, practical methods to guide viewer attention in an immersive storytelling environment were investigated. The results of that study provided insight into the effects of attention guidance cues on viewer behaviors and the watching experience (see sections 3.4 and 3.5). The results from the first user study threw light on how viewer experiences were affected by AUs in CVR. In the group where AUs were applied, the viewer's feelings of engagement and enjoyment with CVR content increased, while the time spent searching for the ROI was reduced. These results indicate that the application of AUs helped mitigate the FOMO issue, proving that AUs, as a set of cues, are effective tools to assist creators in CVR production. In the second study, the viewer's involvement and agency in terms of controlling the narrative progression in CVR were explored.Unlike the first study, which emphasized the creator's role in production, the second study focused on a combination of two elements on the viewer's side: narrative progression and interaction visibility. Via a series of measures of the viewer's subjective feelings along with semi-structured interviews, insights were gained despite the data not having revealed any significant differences. Another guideline derived from this user study is that the CVR creators are suggested to design control modes and interaction visibility based on the original purpose of the storytelling experience, as well as providing incentives and guidance to support the viewers to participate. Especially in a CVR experience with implicit interaction, the viewers were given more extensive capabilities to interact with the narrative (enabling them to participate in the narrative progression), without being aware that they were empowered to actually affect the narrative progress. Therefore such a design is recommended for creators working toward a more "lean back" and passive experience, such as a documentary.

It is important to note that these resulting guidelines remain unorganized and disconnected. Some deal with content creation, others with system design, and none approach CVR as a co-creative experience involving active engagement on the part of both the creator and the viewer. In addition, the CVR content used in these studies is only available in the "guided tour" format and place a human narrator in each scene at all times. It has not been determined whether other types of CVR should follow the same rules. Therefore, the following sections (5.2 and 5.3) present follow-on work that unifies the rules into a comprehensive framework. The applicability of such a framework to other types of CVR content will be investigated in a later chapter (section 7.3).

5.2 Merging CVR Guidelines into a Coherent Framework

In the two user studies, knowledge about using CVR for effective storytelling was gained. How viewers, as a necessary part of the CVR experience, can be placed in a balanced role in the narrative process was also explored. However, the guidelines produced remain scattered through different realms, and each covers only one of the many challenges creators are facing. CVR creators currently need to use a trial-and-error model to identify which guidelines to use in resolving problems encountered in production. Therefore extra steps were taken following the user studies to formulate a coherent *adaptive playback control* (APC) framework for CVR creators. The framework was designed with the goal of matching the guidelines with the workflow of CVR production, from content preparation, through plot design, to viewer interaction design, so creators can naturally choose among the guidelines and look for those referencing their current stage of their work. The framework is also intended to provide a *holistic* view of how the parts are linked together in a CVR experience. The vision is that through this framework, CVR creators can produce engaging and enjoyable work, knowing they will maintain authorial control over the plot and the theme of the story while being assured that the viewers will not experience the FOMO issue. At the same time, the viewers will have the freedom to interact alongside the narrative progression, thus living up to their expectation of agency. The framework will also assist creators in designing interactive CVR experiences without overemphasizing the interactivity, such that it becomes more of a gaming experience. The following section discusses the motivation for the APC framework, how it was developed, and how it is inspired by other frameworks for different interactive media, and then describes the APC in detail.

5.2.1 Previous Guidelines for CVR

The first step in the exploration of the CVR framework was to review the existing guidelines or references already established by other creators and researchers in their work. Over decades of the development of cinema, filmmakers established practices and guidelines for effective storytelling through this medium [17]. As 360-degree video became more widely available and easily accessible for the general public, creators integrated it into their toolkit as they tried to tell long and complete stories [59, 13]. The exploration of new sets of storytelling grammar coincided with the start of production of experimental works [146].

At first, creators were producing CVR works in trial-and-error mode, experimenting with prototypes and gathering design references from filmmaking projects while retrofitting them to suitable storytelling grammars and design principles for CVR [18, 60]. Examples include setup instructions around the 360-degree camera, and instructions around placing characters on stage for the new medium and new POV of viewers (such as [18, 27, 79] previously discussed in section 2.2.1). Later, some creators and researchers reached into the realm of game design to borrowed theories such as *co-construction of the story* and *ludo-narrative* [157, 73]. These theories were imported to CVR to highlight the viewer's necessary contribution toward the progress of delivering a complete story and providing a circle of experience. They served as a type of plot design reference moving beyond *how 360-degree video should be filmed so viewers can watch comfortably* to consider *immersive story* *construction whereby the viewer should also be regarded as an author* and suggesting creators place special consideration on the intrinsic characteristics of the immersive experience at the early stage of design process [126]. Their discussions, however, remained at the theoretical analytical level and did not offer specific, useful references. Direct application of these theoretical concepts to practical creative and shooting tasks is challenging for CVR creators. A well-constructed set of guidelines available for direct use is still missing for creators moving from filmmaking or regular 360-degree video to creating an immersive storytelling experience with viewer interaction.

5.2.2 Frameworks from Other Interactive Media

In the second step of exploration, a survey of previous works that provided frameworks for other types of storytelling was conducted. The survey covered various types of *storytelling media*, such as tabletop games, video games, and interactive TV programs, with the aim of finding common patterns to describe the necessary elements for a framework to support content creation. Among the cases investigated, Carstensdottir et al. [24] examined interaction design in interactive narrative games, specifically the structure and progression mechanisms to establish common ground between the designer and the player (see previous discussion in section 2.4). They listed six types of story structures commonly found in narrative games and summarized four types of progression mechanism. Most importantly, they stated that the choice of story structure and of progression mechanism for a game is not a random combination. However, a reasonable decision needs to be taken by designers in order to match the mental model players will form when they see the progression mechanism towards the structure of the plot, even before they have fully experienced it. Ursu et al. [155] also listed a series of existing programs and summarized a system structure for an effective software to support both authoring and delivery of TV programs. Although not directly in the realm of immersive storytelling, both still served as useful examples of how reference frameworks are constructed for narrative experience creators. Moreover, when viewer participation is regarded as a key component for story comprehension and becomes the aim of the entire experience, such as for [85, 58, 97], viewer interaction is more carefully designed and generally considered at the very beginning when the story is composed. In these works, the researchers thought to use frameworks to guide the process by merging interaction design and content preparation organically such that viewer/player interaction became an essential part of the narrative.

Nevertheless, extra adjustments are still necessary before these frameworks can be applied to CVR, since CVR is more of a *lean-back* medium that involves less active and intensive input from viewers compared to the media types discussed above.

5.3 Adaptive Playback Control: A Framework for Interactive CVR

The review of previous work on interactive and immersive media yielded an important conclusion: that a coherent framework for CVR production should align with the workflow and how creators actually build the experience. More specifically, the framework needs to be compatible in terms of helping creators understand the intrinsic characteristics of CVR (such as 360-degree video), how it differs from a traditional flat medium, the unique challenges it presents, and the solutions available. The framework also needs to encourage creators to see a CVR experience from the viewer's perspective. It is expected to help the creator recognize a viewer's expectations as well as their concerns when entering a CVR experience. When these two aspects are both achieved, the framework is *coherent* and the creator can be confident in creating an engaging and effective CVR storytelling experience from both a *bottom-up* and a *top-down* perspective (cf. 4.6.3). Therefore, to support the creation of effective storytelling experiences for CVR, the APC framework is designed to fulfill the aforementioned requirements, providing support for both *content creation with viewer participation in mind* and *viewer interaction design by involving the viewer as a participating character in the storytelling scene*. Further, because the frameworks used in the works mentioned in the last section have been proven effective by others who used them as references for design [47], a similar framework following this workflow for interactive CVR production is constructed here, with two parts, one focusing on the creator's role and the other on the viewers:

- **Immersive Content Production Guidelines**: Being different from traditional flat-screen media, immersive content requires creators to adopt new mindsets to take into consideration, for example, the role of the viewer, the new mise-en-scène, the FOMO issue, and the necessity to resolve it;
- **Viewer Interaction Design Guidelines for CVR**: Since viewers are also involved in the immersive storytelling experience, their capability to interact and the agency the system enables them shall meet their expectations when

they step into the immersive scene. Creators will need to design the interactive elements carefully and consider viewer interaction as part of the narrative process from the beginning.

Content creators are expected to use this 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, culminating in an engaging and satisfying immersive storytelling experience with interactivity. As per discussions in earlier chapters, via this framework the creators retain essential *authorial control* of the story but at the same time allow a certain freedom of viewers so they have agency and find the storytelling experience moulded and shaped according to their expectation (which makes it *adaptive*). Thus the name of the framework: ***Adaptive Playback Control (APC)***.

5.3.1 The Immersive Content Production Guidelines

As stated in previous chapters, immersive content production differs from traditional filmmaking or video production. Creating immersive content requires directors to consider the viewer's role, the new mise-en-scène, and the FOMO issue, which are all non-existent with the traditional 2D screen. It becomes necessary to resolve these issues with proper techniques to ensure the desired user experience of the final product. The addition of attention guidance cues into the immersive scene, as learned from the research, effectively kept the viewer's attention on critical story elements to avoid losing track of the plot and thus confusion and failing to comprehend the story. The following points build on knowledge from previous studies and added into the content production guidelines:

- CVR creators need to consider the choice of suitable cue designs. They need to consider factors including whether the cues are diegetic or temporarily or consistently visible, for different contexts and purposes. Generally, diegetic and temporary/dynamic cues (such as the AUs used in previous studies (see Chapter 3)) are recommended to achieve CVR that aims for dramatic narrative.

- The design of content structure should also match the purpose of the storytelling, as well as its context and background. In some cases the content is organized in a more complex structure to meet the need for viewer participation and interaction that goes beyond just actively looking around (discussed in detail in the case study in Chapter 6).

- References on structural content design and assembly are also provided, assisting CVR creators in designing and preparing scripts and blueprints ahead of actual capture (further discussed in Chapter 6, especially section 6.2.1). This aligns with the fact that for interactive CVR, the viewer will likely expect interaction at a medium level with input and some involvement in the narrative progression, as stated in Chapter 2.

We should also be aware that this guideline is suitable not only for 360-degree video for compelling storytelling but also for other immersive media. In sum, it includes basic areas such as scene setup, camera manipulation, and the use of attention guidance cues like AUs for production work for specific CVR content, then also a more abstracted structural reference to guide creators in integrating this content into a digital storytelling system. The content produced also becomes the "content library" (a concept mentioned in Section 4.6.1) to support the type of structure chosen for that storytelling experience.

5.3.2 The Viewer Interaction Design Guidelines for CVR

According to the workflow for APC, after the content creator lays down the story structure and prepares the content following the guidelines, the story is ready for application of the APC guidelines. The next step is to set up how viewers will interact with each piece of the content and progress along the storyline. In previous research, parameters from multiple axes of the viewer's interaction in immersive media were identified to define a method specifically for the CVR experience. The core concept is to remember that a perfect and universal solution does not exist. To the contrary, content creators will need to design interactive elements for each project, aiming to balance the viewer's capability to interact and their agency of the experience with the expectations they form as they step into the immersive scene.

The complexity and capability of viewer interactivity was discussed in Chapter 4 and represented as a continuum of interaction. The depth of viewer agency, or the extent to which a viewer can have an impact on the narrative, was also described. The observations of subjective viewer impression, their understanding of their roles, and their control capabilities will also be included here as a reference for what viewer expectations will be like for certain types of immersive and interactive content, as well as a part of the APC framework and guideline. Finally, all of the above will be consolidated and provided as a reference for viewer interaction design for CVR, covering the choice of parameters related to viewer interaction techniques such as the type of control, the main role of controller of the narrative progress, and whether this control is perceptible to the viewers. These factors were already extensively explored and discussed in the second user study, in Chapter 4. The following items will be included in the guideline:

- The viewer needs to have control over the pace of progression of the narrative. The control of story progression might also rightly be handed over to the viewer, at least partially, to take into account the fact that the viewer is also involved in the storytelling activity, being also a character in the scene and expecting some agency over the story;
- The option to disable control should be available when viewers tend to withdraw from the controller role. Since the possibility that a viewer may shift from medium-level participation back to passive observation during the storytelling cannot be ruled out, the creator of an interactive experience needs to consider how viewers establish a certain mental model in their minds when a control or input method is put in their hands, and be flexible in following that mental model;
- CVR creators could exploit viewers' natural tendency to explore in an immersive environment and apply appropriate incentives to invoke exploratory behaviors in them. Meanwhile, the system's design needs to support viewers' intentions and expectations, for example if their exploratory behaviors are being sparked, and avert frustration when viewers hit an "interaction obstacle" and veer to "being too afraid to act".

5.3.3 A Conceptual Example of an APC-Powered Interactive CVR Experience

To demonstrate how the APC guidelines can be applied to a CVR production, a conceptual example is proposed here, enabling implicit viewer interaction while maintaining authorial control. The example includes these preconditions:

1. The viewer is seated in a swivel chair and wearing an HMD for the VR experience;

2. The viewer is passively watching, but the creator's screenwriting will define the viewer as (at least) a participatory-level character in the story;
3. The viewer's experience will not be interrupted by non-diegetic elements, and the time flow will not be explicitly affected by viewer's interaction.

Th system emphasizes both the viewer's expectation of agency and the creator's expectation of authorial control, following the continuum of interactivity outlined in Chapter 2. To strike this equilibrium, the following three fundamental elements are created (Figure 5.1):

- A *story builder* component for the creator to fill in raw material and references to the internal relationships between key story nodes and their weights of priority;
- A *moderator* component running in the background to capture the viewer's real-time behavior and mediating the playback process as a response to the viewer's actions;
- A *projector* component presenting the content to the viewer and receiving commands from the moderator updating the presentation when necessary.

In this system, the creator, as story builder, provides the content in the form of nodes, plus map-like references for the moderator. This map is the results of assigning the nodes and paths with "weights of necessity" defining, for example, which nodes are mandatory and which parts of the path are predefined and cannot be waived. The projector captures implicit input from the viewer (such as head rotation and gaze change), for the moderator to be read as signals, altering the path of how the story progresses from one key node to another in real time. Before playback, the creator feeds the content (as nodes) and the map to the story builder as raw material. The moderator uses the map as a reference to the possible paths a viewer can later take connecting the nodes in playback, the mandatory paths, and their priorities (all visible in Figure 5.1). The viewer watches the content when playback starts, wearing an HMD. The moderator monitors and analyzes the viewer's behavior as they watch to work out which node the viewer is interested in at a given moment, cross-references it with the map, and determines which node will be presented next, and accordingly sends the directives to the projector. In response the projector presents the chosen node (as content) to the viewer. Since this is all running in the background, the viewer is unaware of the fact that s/he is making choices what will affect the narrative progression, albeit implicitly.

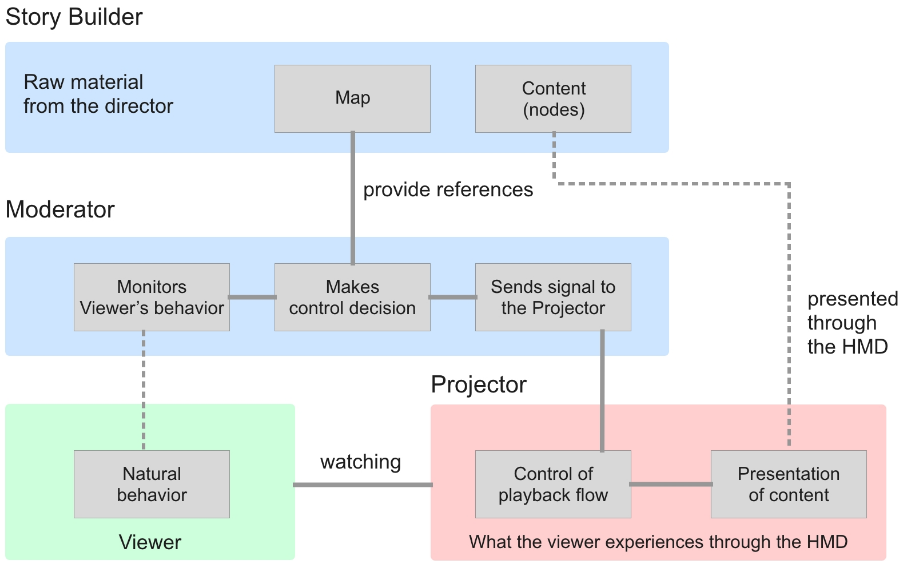


FIGURE 5.1: Structural diagram of the interactive system using the APC guidelines, showing the system's three components and a viewer who is using the system, and demonstrating how the components work together to deliver the final result to the viewer.

In this conceptual example, the APC guidelines ensure the experience will meet the expectations of both the creator and the viewers. A viewer in such an experience can find him/herself with the agency to take a role in the story through a certain level of interaction with the system, because the path the story follows changes (within certain boundaries) according to the viewer's behavior. Meanwhile, the primary purpose of the storytelling and the emotional response the creator wants to invoke in the viewer can still be ensured, because the essential narrative arc and the main plot are still laid by the creator and not affected by the viewer's behavior while watching. In the next chapter a real-world case will be presented further exploring the topic storyteller control and acknowledgment of viewer involvement. It may also be possible to observe whether the viewer's sense of narrative immersion and engagement and enjoyment will be affected by the fact that his/her interactions within the story world are at a *limited level*, compared to a sic degrees of freedom (6DOF) version or full-scale VR game.

5.4 Conclusion

In this chapter, insights from the two user studies were summarized in regards to the creator's toolkit and viewer interaction in the CVR storytelling experience.

The need for a broader framework for creators of CVR was demonstrated and the Adaptive Playback Control (APC) framework introduced, including a description of its motivation and the process behind its creation. An overview was then provided of the two components of the APC framework: the content production guidelines focusing on the creator's side and the viewer interaction design guidelines highlighting viewer's role, and considerations around these. In the next chapter, a case study featuring storytelling co-design workshops focusing on indigenous storytelling will be described, serving as a real-life example of the application of the APC framework.

Chapter 6

Case Study - Use APC Framework to create Māori Storytelling Experiences

6.1 Introduction

In the previous chapters, two studies were described to cover a series of guidelines for CVR creators to produce interactive storytelling experiences. Those guidelines were then merged into the APC framework that covers the workflow of CVR content production, providing references to creators when designing the experiences and involving viewer interaction in the progress of the narrative. In this chapter, a case study, where the APC framework was put into practical use, is presented. We aim to create a digital experience using CVR, allowing users to experience stories told by virtual storytellers in a way very similar to the real world face-to-face experience.

The case study started with a series of co-design sessions with *kaikōrero* (Māori storyteller), collaborators from the Te Rau Aroha Marae in Bluff, Aotearoa New Zealand. Such whanaungatanga (connection between people) was established with those collaborators in the first place because of the geographical uniqueness of the marae and the Bluff township. Te Rau Aroha Marae is the world's southern-most marae, located in Bluff, which is the southern most town on the South Island of New Zealand (Figure 6.1)[1]. The rūnanga (tribal council) of the marae today connects people who live in both Bluff and on the islands off the south coast such as Ruapuke Island and Rakiura (Stewart Island). Stats data[2] shows more than 40 percent of the

[1]See more on: https://ngaitahu.iwi.nz/te-runanga-o-ngai-tahu/papatipu-runanga/awarua/
[2]See more on: https://www.stats.govt.nz/tools/2018-census-place-summaries/bluff

population there self-identify as Māori and maintain the traditional lifestyle, including the seasonal collection of muttonbirds and oysters. This uniqueness of the Bluff community sets a strong cultural background for this case study, and emphasizes the necessity for being culturally appropriate and following Tikanga Māori (Māori customary practices, behaviors and protocols) when applying digital technologies to traditional cultural activities.

FIGURE 6.1: A map view showing the location of Bluff in the South Island of New Zealand. An enlarged view shows the Bluff town and islands off its coast. A photo of the Te Rau Aroha Marae is also presented.

The context of this study also supports a second motivation. It aims to find out if a digital storytelling can help indigenous people (Māori) connect with the storyteller and the scene in the story, when the digital experience is built following their understanding of maintaining connections via storytelling, or matching their expectations of such experiences. The APC framework was also introduced to all the relevant parties in the project and used as a reference for the group to design and produce CVR content.

The third motivation to conduct this co-design project stemmed from the previous study (see Chapter 4). In that study, since viewers had control over the narrative progress, they experienced different versions of the tour because of their choices. However, the results indicated that regular viewers were not aware of the personalized component of the storytelling experience, nor could tell the difference between a standard linear story and the interactive version where their behavior and input had been considered. Rather than asking for an evaluation from general end-users, the storytellers are assumed to be in a position to validate if the digital storytelling experience was appropriate and feasible when used to bring people feelings similar to those in the real world, because in the end, what this digital storytelling has been trying to mimic is the storytelling experience those kaikōrero offer in the physical world; they are the experts in this topic and are the ones who have the most potential to tell the difference and give practical feedback.

With these observations and considerations in mind, the following components were drafted:

- Planning the overall structure of the CVR experience with a focus on the "personalization" feature;
- Conducting a series of co-design workshops with Subject-matter Experts (SMEs; in this case, *kaikōrero* of Te Rūnanga o Awarua from Te Rau Aroha Marae in Bluff, New Zealand), around what elements are needed to enable personalization to bring it close to the activity of oral storytelling;
- Capturing a series of customizable CVR content using the APC framework as its reference.
- Developing a system that delivers personalized storytelling experience using CVR.
- Evaluation of the system and its experience by the SMEs to see if 1) it matches the expectation of storytellers, like the oral storytelling experience they would like the viewers to have; 2) it is culturally appropriate in their perspective.

In the following sections, the process of narrowing the exploration area of this project will be first elaborated. It starts from a design space about viewer interactive storytelling in CVR, then moves onto the co-design workshops conducted with kaikōrero Māori, including the results and reflections, followed by an evaluation and discussion of the CVR system.

6.2 Enabling Further Viewer Participation in Immersive Storytelling

In deriving and generalizing the APC framework, bringing the viewer's participation into CVR storytelling was the main focus. However, its effects are limited to only the temporal succession of content segments. The other possible effects were not fully explored, such as how it affects the content disclosure. From the viewers' perspective, in the past studies, they were allowed to control the narrative progression (responding to what they do), but not yet to change the narrative based on their profiles (responding to who they are). In real world storytelling, the teller embraces changes; for example, the same Tama Rereti story (a traditional Māori story about their ancestors) has variations across different *iwi* (tribes), when told by different *kaikōrero*, and also towards different audiences [104]. Thus alongside using APC in an actual project, it is also extended to create a more personalized CVR experience, resembling real world oral storytelling. A more "personalized viewer participation" means the viewer will not only find s/he is experiencing the story as if being there with the teller, but will also find the narrative is tailored towards him/herself.

In Section 2.5, the typical pattern used in viewer-centered (personalized) digital storytelling experience was introduced, which involves a customizable content pool, a system to collect and create the viewer's profile, and a strategy to generate variations in the stories disclosed. This pattern is not exclusive to such digital experience, but has been widely used in our daily life, even before digital technology. In oral storytelling activities, Māori people do not tell a fixed story, but draw necessary elements from a pool and assemble and arrange them into a story that fits the current context and need, plus adding the teller's personal touches [124]. When telling a story, the teller draws from a pool of conventional imagery and dramatic devices, but takes liberties with how they used and arranged them, depending on who the listeners are, the context and the purpose of the telling. Different tales would often highlight different characters or events, and each story would have its own individual touches of the storyteller [117]. In some oral traditions, oral storytelling is also participatory; it is not only a matter of the teller, but also related to the listeners. Both parties see the storytelling activity as a communal sharing that binds communities spiritually and relationally. Thus both the stories and the ways of telling are as varied as the people who tell them and the attending listeners [139].

A fact to admit is that the design space around storytelling personalization and traditional oral storytelling is large and complex, thus this study is not going to explore all the details and effectiveness of each combination. This would require considerable resources, some outside the scope of the research's focus. Some design decisions were made in advance and set for this co-design project, helping to narrow down design space.

6.2.1 The Segment and Take Structure

In order to frame customizable content preparation, a flexible structure for a story was first explored, enabling storytellers to generate narratological variations when presenting it to their listeners. In the scope of our research, the content is always in the form of CVR, and more specifically, recorded 360-degree videos. Unlike museum exhibitions where the content are distributed in space, CVR content are linked together by time; mainly presented in chronological order rather than spatial arrangement. Thus in this study the idea of **Matrix of Segments and Takes (MS&T)** for CVR was introduced to demonstrate that the topology, content blocks, repetitive variations and alterations on time sequences can all be consolidated together to be explored holistically. In the MS&T model the structure of a story in its finest granular elements were also implemented, namely the *Segments* and *Takes*. Their relationship in the formation of a story is shown in Figure 6.2.

Segments are small content units within a story following its temporal succession. A segment is a unit containing a piece of the story that can no longer be divided into smaller parts to maintain its integrity and be comprehensible to the viewers. In other words, a segment is like a mini-story that if consumed by the viewers, should at least be understandable within this same segment. An example is a scene of a conversation among several actors on a set with several lines of dialogue, where the viewers can get the general picture of what those actors are discussing. Such a segment cannot be further separated or the viewers will have difficulty understanding it. One story typically contains several segments, for example the five-segment structure used in many stories, i.e. beginning, rising, climax, falling, and end. However, the order can be different in different disclosures, such as the end presented first followed by a flashback to the beginning of the story.

Takes are the cluster of content clips of the same **segment**. Like the *takes* cinematographers capture onset in filmmaking, the **Takes** here comprise multiple captured clips of the same story segment. Content-wise, they are the same story, but

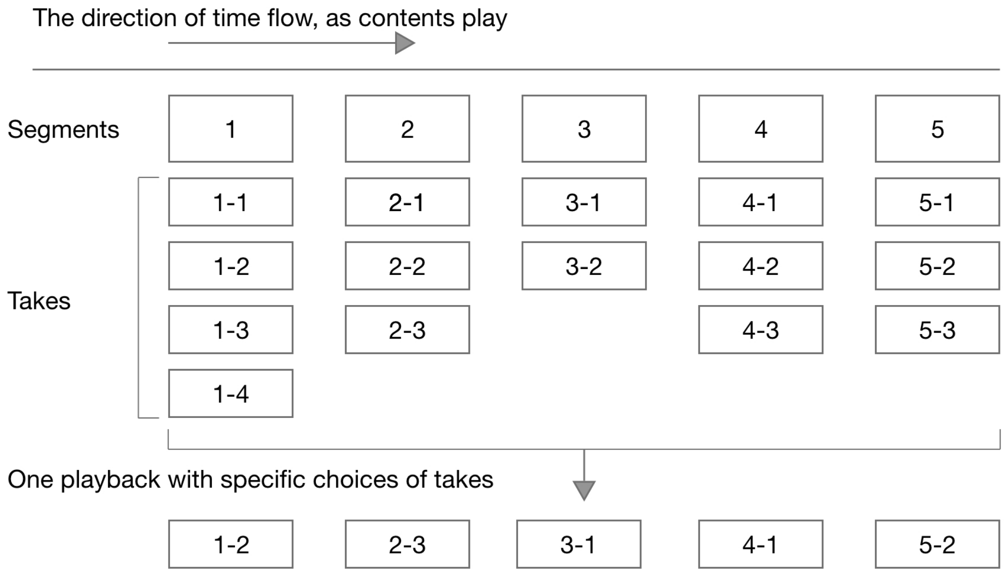


FIGURE 6.2: A figure showing the MS&T, consisting of segments (rows) and takes (columns). Each segment along the timeline may have different numbers of takes available in parallel. On each playback, the viewer sees only one version of the story with specific choices of takes assembled. As an example, there are five connected segments chosen for one playback, shown at the bottom of the figure. Notice each of them contains a specific take

each is unique in detail. For example, the beginning of a story can be told in different ways, with different styles and patterns, by different storytellers, and facing different groups of listeners (e.g., visiting groups, and personas, which will be introduced in the next section). In an interactive or immersive storytelling context, using takes can lead to the viewers experiencing variations between repetitions. In storytelling personalization, a creator can tweak one segment into several variations, i.e. *takes*, to match different viewers. Since viewers will have different knowledge bases, personal connections, and expectations towards the story, if the content is adjusted to match the viewer's profile, everyone will find the content they experience fitting their expectations and yielding higher engagement and enjoyment.

In general, the MS&T represents a two-dimensional mutable story structure (as shown in Figure 6.2). On one dimension is the segments' order, structure, and dependency; the other, within a segment, is a parallel array of sibling takes with the same content but varying presentation. Eventually, at play-back, a viewer or a group of viewers, at one time, will only get one version of the story, which is a combination of 1) a series of segments in one temporal order and 2) a series specific choice of takes under each segment block. Such combination is shown in the bottom row in Figure 6.2. The shown structure also support a different story variation in which the policy for selecting *takes* is random.

This case is set to mainly looks at narrative variation and personalization when a kaikōrero Māori (storyteller) tells a story about his/her marae, whānau, or essential things related to that. Since it is preferred in Tikanga Māori that the story and its disclosure are separated and treated differently, the MS&T model was imported as a base structure to prepare the content. The 360-degree videos were used to built a CVR experience, with storytellers in the scene presenting content related to the objects visible in the scene. Action Units were used during capturing as the primary method to attract the viewer's attention to the ROI.

6.2.2 The Strategy of Creating Variations

In this study only the *Takes* of the MS&T model were used, leaving segments fixed as the storyteller presented. The reason for this is that in the Māori storytelling context, it was unsuitable for these to be disassembled for a segment-based strategy. The Māori *kōrero* (the activity of storytelling) is a principal means to give identity and *mana* (honour) to the individual, and expresses their connection with the whānau (family), hapuūand gives identity to the individual, and tīpuna, the ancestors) [14]. Because of this centrality of purpose in conferring identity, and following Tikanga Māori, it is not preferred to disassemble a story told by one teller as it may compromise the mana contained within.

Once the story is captured, the temporal succession of the segments is fixed (complete creator control). Variation happens on the takes, as one segment will have several takes in parallel. Some early discussion and consultation with Māori experts also led to a preliminary understanding that it is inappropriate to use a system to mimics the kaikōrero mechanically. Instead, the variation should be introduced into the stories completely by the kaikōrero, with the personas as references. In this project, the following strategies were applied to create narrative variations:

1. **Same story, same storyteller but towards different visitors**. This is the primary variation initiative, aiming to introduce a variation on the perspective of presenting the story, the character and the information included in the story, and also the language used to present the story (English vs Te Reo (the

Māori language)). But the final decision and presentation would remain in the hands of kaikōrero, not us.

2. **Same story told by different tellers**. Variations will be introduced on different tellers' perspectives, the use of different characters and different presentation styles.

As decisions were made on content structure, visitor's profiling and variation strategies, the design space was now narrowed down to a reasonable size for one study. The research questions, content preparation, system development and evaluation, are all introduced in the following sections.

6.3 Co-design Sessions - Inviting Listeners into Storytelling Process

As described at the beginning of this chapter, viewers were not aware of the variations introduced into the storytelling experience. The quantitative methods conducted on general viewers did not reflect the differences on designs of system variants (see Chapter 4). In such a scenario, many cultural heritage researchers have used qualitative methods, such as co-design sessions and working directly with SMEs [103, 149, 159]. In this project, kaikōrero Māori collaborators from Te Rau Aroha Marae of Bluff, New Zealand were invited to join a series of co-design sessions. They were introduced to the project and the idea of *building a personalized storytelling experience* using digital technology. Tools to represent viewer variations were initiated, discussed and iterated. Content capture plans were also formalized during these sessions.

The sessions with experts of Tikanga Māori and Mātauranga Māori were to discuss and explore the following session questions (SQs), aim to ensure the cultural appropriateness through out the process:

- **SQ1:** What appropriate features from visitors a kaikōrero (storyteller) will consider?
- **SQ2:** What design tools can be chosen to appropriately represent the visitors that fit real storytelling scenarios?

The first decision of the sessions was the use of "personas", as a set of fictional people [89], to represent the typical visitors in the Māori storytelling context. Because the content preparation and the actual visiting are separated stages in digital storytelling, the personas were used in the former one. A list of preliminary personas were prepared in advance as a starting point, with two virtual events as their backgrounds. The first one is a virtual Waitangi Day (New Zealand national holiday) event held at the Bluff Marae is also set as the background for storytelling, as is usual. On that vitural event, the kaikōrero will welcome visitors and tell them the stories of the Marae, the iwi, and about the carvings in the *wharenui* (meeting house). The second one is future visits of younger generations to Te Rau Aroha Marae, viewing stored stories told by the kaumātua (elders who are also the storytellers). People could experience the stories digitally, and the system would be able to tell who the visitors are so it could give them a story as if they were there with the kaumātua telling it. After reviewing some previous works, key features grouping into two categories emerged, fitting to differentiate visitors in a Māori storytelling context:

1. Demographic features, including: age, gender, ethnicity (Māori or Pākehā [non-Māori]), and genealogy (whether s/he has whakapapa to the marae);
2. Context-related features, including: viewer's motivation, expectation, related knowledge and experience, any previous visits, personal relationship with the storyteller.

As a first step, six virtual visitors (as preliminary personas) were created from those features, as shown in Table 6.1.

The purpose of those personas, or "virtual visitors", is to help storytellers deliver different versions of the story as they would in real world scenarios. For example, a kaikōrero from the marae was asked to look at one of the personas and assume s/he was the one visiting the marae today and was going to tell the visitor about the *whānau* (family) wall in the *wharenui* (main meeting house). Those stories would be then captured digitally. This set of "prototype personas" was then bought to the co-design sessions to be reviewed for any missing critical factors and whether they are appropriate for a Māori storytelling context. Because of the pandemic, all the co-design sessions were not held face-to-face but as online video conferences.

6.3.1 Session 1 - Introducing Digital Storytelling and Personalization

In Session 1, the Te Rūnanga o Awarua kaikōrero invited were first introduced with the project's background and motivation. The prototype personas were then introduced and shared with them. After that the group discussed Māori storytelling

TABLE 6.1: The original preliminary personas created.

Name	Age	Gender	Ethnicity	Whakapapa	Motivation	New	Kwn	Previous experience
Rawiti	25	tāne	Māori	Yes	Meet family friends	Yes	No	Basic understandings
Claire	30	wāhine	Pākehā	No	Learn about the carvings	Yes	No	Good general knowledge, basically from internet
Kaia	70	wāhine	Māori	Yes	Reconnect to whānau and iwi	No	Yes	Closely connected to the whānau
Simon	35	tāne	Pākehā, Asian	No	Experience stories and learn knowledge	No	Yes	Basic to intermediate understandings
Atarangi	55	wāhine	Māori	No	Social and meet people	Yes	Yes	well-versed in mātauranga Māori (Māori knowledge)
Jin	65	tāne	Pākehā, Asian	No	See the buildings	No	No	Minimum knowledge

In the head row, several abbreviations were used to save space: New = new visitor, Kwn = known to the host.

and what they thought were essential factors in a storytelling activity. After the co-design session the recorded audio files were transcribed to texts using an online tool. The scripts were then manually read through for possible errors, organized and analyzed using Thomas' General Inductive Analysis [151] method as a guideline. However, due to the limited time and the number of experts we could reach out to, a formal thematic analysis as described in the guide was unable to be conducted in full scale. The procedure of the actual analysis carried out are listed as follow:

1. The transcription was carefully read through and text sections closely related to the research topic were collected;
2. A series of topics are generalized, representing the main information from the collected sections from step 1;
3. Relationships and potential links between the topics are explored. Important insight gained from the co-design session were summarized;
4. The personas were updated following the insight and results.

The discussion first highlighted the fact that Storytelling is about the transfer of knowledge, the sharing of knowledge, and reciprocal knowledge sharing. It mainly states that the activity of storytelling is access control to knowledge. Storytelling is also regarded as sharing of knowledge by Māori to set up a conversation's common ground. Other items include:

1. **Since knowledge transfer, sharing, and exchanging all have "access control" and different people will get access to different knowledge, the mechanism of *access control* is a process of accessing the viewer's knowledge base.** The kaikōrero will consider the following factor of visitors as essential. Those factors include their knowledge base level, age bands, and previous visits (whether s/he is a waewae tapu, which means newcomer or rare visitor, or a regular visitor). Meanwhile, gender is less considered a factor. Especially, kaikōrero suggested the use of age bands instead of addressing specific numeric ages individually. Those bands are grouped as follow:
   - Tamariki 0~12 years old;
   - Rangatahi 13~25 years old, or extending to 30, typical college student, no child (in some contexts the tamariki and rangatahi can be merged);
   - Pakeke, 25~mid 50s;

- Kaumātua 55+ and older

2. **The visitor's intention also reflects what knowledge they are after**. The kaikōrero also pointed out that storytelling is mainly towards groups, not individuals. During the storytelling activity following knowledge level access, kaikōrero adjust the length and details of the stories they will tell;

3. **Another important purpose of storytelling is for reconnection and establishing whakapapa, or showing hospitality.** The kaikōrero mentioned that they would like to see the digital storytelling experience as a teaser for returning to physical reconnecting. They would also like to follow the principle of Manaakitanga (sharing, mutual respect, and hospitality) for the digital version of storytelling. It was also brought up that among the visitor groups, those looking for whakapapa constitute a significant part;

4. **Digital tech brings new extra possibilities for storytelling.** The kaikōrero stated their awareness of the necessity to consider digital tech and the young generation while protecting and passing on their knowledge. They also pointed out the necessity of considering language and subtitle options when turning storytelling into digital experiences.

Learnings from the co-design session regarding the storytelling activity are distilled from the main insights and their supporting statements as follows:

1. In the view of Te Rūnanga o Awarua kaikōrero from Te Rau Aroha marae, storytelling is a process of knowledge sharing, and a way of knowledge access control;

2. The kaikōrero use viewer personal traits to determine knowledge access level. They consider age groups, whakapapa and *waewae tapu* with high priorities;

3. For kaikōrero, the storytelling variation was sharing different knowledge based on the listeners' knowledge access privileges and motivation.

In the preliminary version, a fictional Waitangi Day event commemorated at the Te Rau Aroha Marae was assumed as the motivation for those visitors in the persona to be invited to the marae. During the co-design session, new insight into the motivation and behaviors of the visitors were gained, including:

- Kaikōrero normally face a group of people, instead of individuals;
- Storytelling is a process of knowledge sharing. The personalized storytelling experience is actually a process of determining knowledge access;

- Knowledge access rights are not directly observed, but rather determined by the kaikōrero. A normal visitor to the marae could been given access to only the generic kōrero (story). However a visitor who has whakapapa to the marae and had previously been here could be given access to more specific kōrero such as around tīpuna (ancestors);
- Demographic features are used to determine the visitor's proper level of knowledge access;
- The visitor's age band, ethnicity, genealogy, and waewae tapu are important;
- The visitor's intention is also considered by the kaikōrero for knowledge access rights are.

Following these insights, the personas were updated to a new version to represent a group with mixed gender, but homogenous in other characteristics. Several items and instances were also changed, including the *Age bands, Ethnicity, Whakapapa, First-time visitor, and Intention of visiting,* and resulted in 96 combinations in total. A tree diagram presents all of them, starting from the first feature to the fifth, as in Figure 6.3.

However, it was impractical to use all the personas. Further observation to the tree map revealed that some nearby combinations would result in similar intentions. Those similar intentions then led to one single story variation, indicating the potential of treating them as one persona group. This observation was also further validated by the kaikōrero, as stated in Section 3.3. After merging, six group personas are screened out from the tree as typical representations. The six representative groups and how they would cover a series of nearby similar instances, are presented in Figure 6.4. As group personas, each of them will contain approximately 5-10 people sharing the same traits. The details of those six new group personas are listed in Table 6.2.

6.3.2 Session 2 - Updating the Personas and More on Oral Storytelling

After the group personas upgrades, the second co-design session was conducted with the same group of Te Rūnanga o Awarua kaikōrero. The purposes of the second session were reviewing the updated personas, validating the motivations of the proposed digital storytelling experience, and learning more about story variation strategies kaikōrero use in real life scenarios.

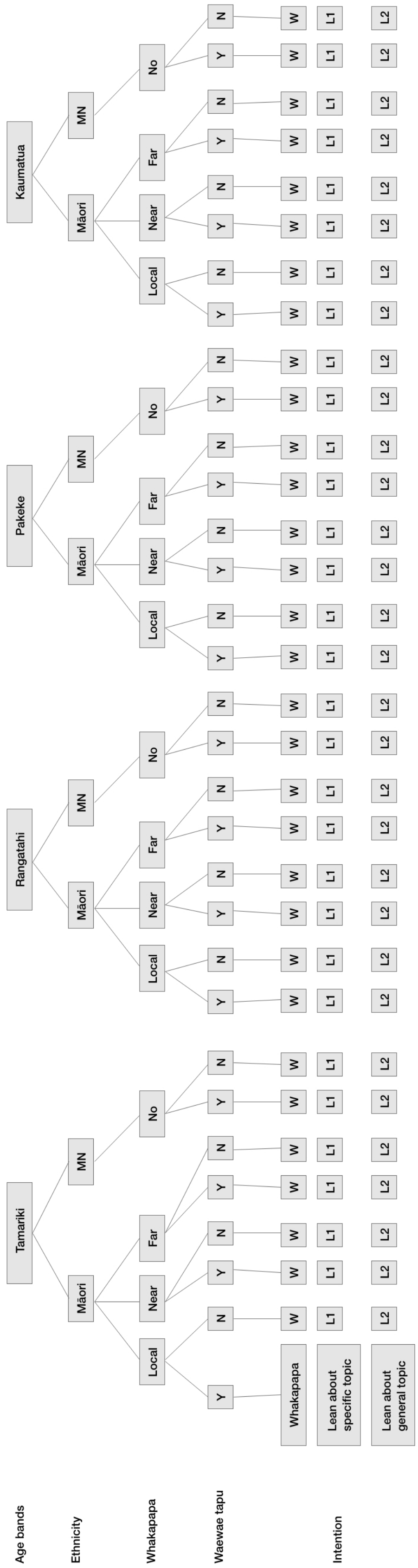


FIGURE 6.3: The full combination of persona instances from the 5 features, result in 96 instances in total.
In the row of **Intentions**, **W** stands for "with the intention of looking for whakapapa", **L1** stands for "to learn about specific topics" and **L2** stands for "to learn about general topics".

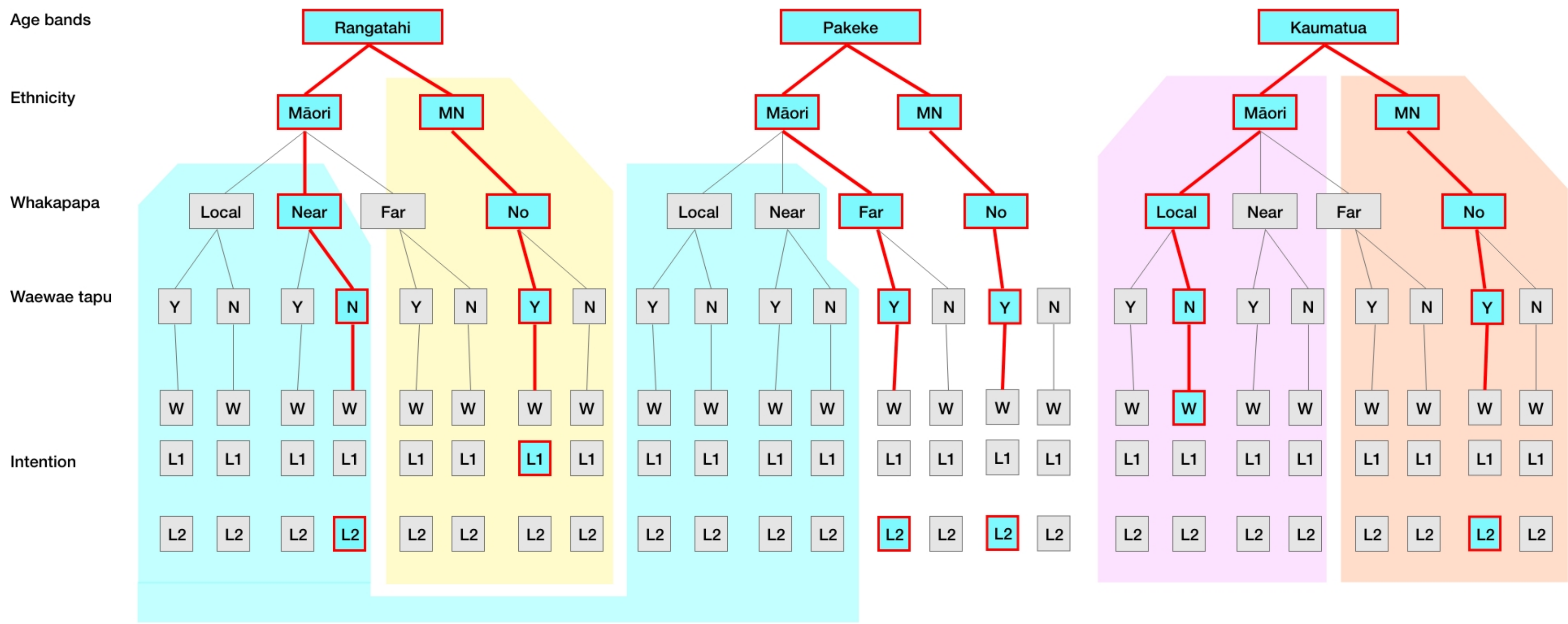


FIGURE 6.4: The six representative choices from the full combination, and their area of coverage to represent a series of combinations that may result in similar story variation outcome. The chosen groups are highlighted with red boxes and lines, with a colored backdrop showing the nearby groups they represent.
In the row of **Intentions**, **W** stands for “with the intention of looking for whakapapa”, **L1** stands for “to learn about specific topics” and **L2** stands for “to learn about general topics”.

TABLE 6.2: Six group personas, updated version.

Group	Age	Ethnicity	Whakapapa	New	Intention of the visit
1	Rangatahi	Māori	Near	No	To hear the stories from the Bluff Marae, their history and the carvings (feature), and share stories from their own marae (story sharing and exchange, mostly generic)
2	Rangatahi	MN	No	Yes	Visit Bluff Marae as a field trip from their school's indigenous art class. They especially would like to learn about the carvings and decorations in the wharenui (Learn about the carvings - history and story oriented).
3	Pakeke	MN	No	Yes	Visiting because it's part of their ocean resource summit in Invercargill and then invited to the marae(generic).
4	Pakeke	Māori	Far	Yes	They are attending the Waitangi Day event, as representatives of the Rūnanga.
5	Kaumātua	MN	No	Yes	Honorable officials and VIPs, come to give a speech at the ceremony (about the importance of keeping people learning and understanding the Māori culture), and visiting the marae after the event.
6	Kaumātua	Māori	Local	No	Looking for family reconnection.

In the head row, abbreviations are used to save space. New = new visitor(waewae tapu).
In the content, MN = Multi-national

During Session 2, data analysis results from the first session were introduced, followed by the updated group personas. The session was also recorded, reviewed and summarized for key points:

1. **Ethnicity** In the updated personas, kaikōrero pointed out how the term "multicultural" refers to a wide range of cultural origins, which is not necessary information in this context. For the purposes of this project the feature *ethnicity* was updated to a binary option, either Māori or Others;

2. **Age bands** Although kaikōrero suggested that they do not recommend merging the groups of Tamariki and Rangatahi, they supported the decision of ruling out Tamariki from the personas design because they would not be able to use the VR headsets and experience the digital version, and to reduce the number of combinations;

3. **Previous visits** In Section 5.3.2, on the updated personas, whether the viewers were first time visitors was listed as a key feature, with options of yes or no. It was clarified by kaikōrero that the idea of *never been to the marae* (represented by the option *No*) would not work out, because if one was listening to the story (in the real world scenario), s/he must have already been on the site of the marae and have been welcomed through the pōwhiri process (a ceremony that formally welcomes a visitor onto the marae.) Thus s/he would not waewae tapu anymore. After discussion we decided to change this variable as *Visited before?*, then two values, 1) waewae tapu, 2) regular visitor;

4. **Whakapapa** Kaikōrero also pointed out that the value *no* on the variable whakapapa, in the updated personas, is also not an appropriate design. They explained their world view that everyone, eventually, is connected and can be included in a large family looking deeper into everyone's genealogy. Everyone eventually has whakapapa to the same spot, as they said. It is also a part of manaatikanga, respecting everyone's origin and background. In a further discussion, according to the kaikorero, the local people who hold *manawhenua* (authority over the land or territory, and people who live nearby or from a nearby marae, may share some family connection, to distinguish them from other visitors. They are more likely to get stories that are very similar and close. Therefore, "local" and "near" were combined into one group *Manawhenua*. Now the variables under whakapapa are simplified to two: the Manawhenua includes local people and people from nearby, and *Ngā Hau e Whā*, which literally means "four winds"), includes everyone else from

afar; both Māori and others. Kaikōrero also emphasized that this view about whakapapa specifically applies to Te Rau Aroha Marae of Bluff; it can be very different in other marae.

The personas were further iterated and updated, following the feedback and suggestions about their features and values, emphasizing ethnicity, age bands, previous visits and whakapapa. The final version of the group personas is presented in Table 6.3.

Another discussion carried out with Te Rūnanga o Awarua kaikōrero was on how the content were prepared in their real world storytelling activities (e.g. the strategy of assembling story segments, as option style A and B, in Figure 6.5). The learnings from that discussion revealed that the kaikōrero would consider and do:

1. In 90 percent (their estimation) of the scenarios, they would use the cascading 3-block variation model (style B in the Figure 6.5) and give everyone the generic version at the start, then a detailed add-on block, and then an in-depth add-on block if the visitors are those who will have access to those content plus time allows. In most cases, only the generic and detailed block were used;

2. For the remainder, where the kaikōrero knows who will be the audience, for example a wānanga (a seminar or forum held and led by Māori), they will prepare a *special* version of the story with in-detail content. Those cases are relatively uncommon. However they did point out that the Group 4 in the final version of the updated personas (Table 6.3), would be one of those cases. They would prepare an in-detail extended version of the story in advance. For group 6, a group of kaumātua visiting for family reconnection, they would also prepare content in advance.

The kaikōrero also talked about language options. They pointed out that the language to use in storytelling is actually decided at the very beginning of the activity. They will know which language to use within a few moments of meeting the visitors. They also stated that, in most real cases, having a group of Te Reo speakers visiting the marae is an extensively formal matter. The kaikōrero will no doubt be aware of the event and prepare it carefully, including deciding to use Te Reo for storytelling. This is described in one of the *special* cases mentioned in the previous paragraph.

TABLE 6.3: Six group personas, final version.

Group	**Age**	**Ethnicity**	**Whakapapa**	**Previous visit**	**Intention of the visit**
1	Rangatahi	Māori	Manawhenua	Regular visitors	They are here to hear the stories from the Bluff marae, including the history and the carvings (feature). They also will share stories from their own marae. (story sharing and exchange, mostly generic)
2	Rangatahi	Others	Ngā Hau e Whā	Waewae tapu	They come to visit on a field trip from school's indigenous art class. They would like to learn the carvings and decorations in the wharenui. (learn about the carvings - history and story oriented)
3	Pakeke	Others	Ngā Hau e Whā	Waewae tapu	Their visit is arranged as a part of the Ocean Resource Summit held nearby in Invercargill. (generic)
4	Kaumātua	Others	Ngā Hau e Whā	Waewae tapu	They are honorable veterans invited to a local ceremony about remembering the people's past and look into the future, on the ANZAC day. They will visit the marae after the event.
5*	Pakeke	Māori	Ngā Hau e Whā	Waewae tapu	They are representatives of the Rūnanga, coming to attend the Waitangi Day event at the marae.
6*	Kaumātua	Māori	Manawhenua	Regular visitors	They come to visit the marae as they want to look for reconnection with the whānau.

*Groups 5 and 6 are the ones that the kaikōrero commented they would treat as special cases. To these groups the kaikōrero would prepare content suitable for them in advance. For example, the story for Group 5 would very likely be presented in full Te Reo.

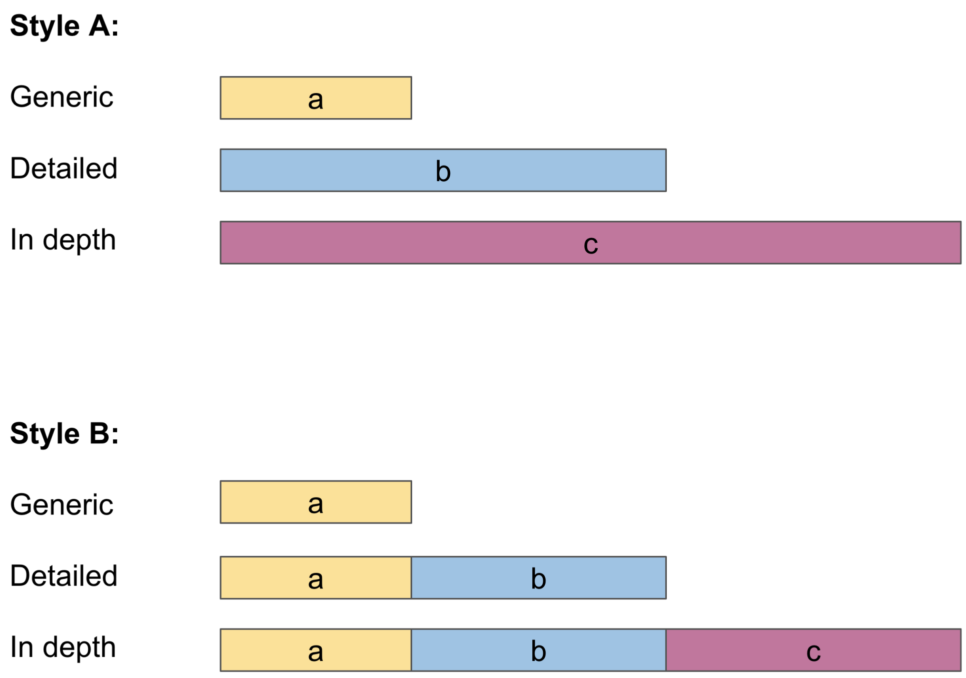


FIGURE 6.5: Two variation models for storytelling. In style A, different groups will get completely different versions of the story. In style B, every group will start from getting the generic version a, and some of the groups will get further content depending on who they are and the context of storytelling.

To sum up, for Te Rau Aroha Marae of Bluff, in most scenarios the kaikōrero will not decide exactly which version of the story to give in advance, but rather when facing the visitor groups. However, in some scenarios, maybe more formal ones, the kaikōrero will prepare the story in advance, as they would already know who the visitors are, their intentions, and which story will fit for that case. Those variation strategies were then reconnected to the persona groups in Table 6.3:

1. For groups 1 to 4, the visitors will start by getting a generic story, and maybe some additional details if time or visitor's interest allow that. An interactive way can also be used, such as setting up prompts for visitors from the kaikōrero (as it is in a digital version), and the visitors themselves can decide if they want to continue into a detailed version story;

2. For group 5, it is more like a specific case, because as their visiting is important, the kaikōrero will be aware of that in advance, and also prepare a specific story in advance;

3. For group 6, where the group of kaumātuas' intention is looking for family reconnection, the kaikōrero know that they will give a more detailed and in depth version story for this case.

Up to this point, part 1 of RQ_3 in Section 1.2, asking the key features kaikōrero will consider in personalized storytelling and the variation strategies they will apply to the content, has been discussed, clarified, and hence is answered.

6.3.3 Extra Reflections after the Co-design Sessions

It is learned from the previous analysis that in the Māori world view, storytelling is a process of reciprocal knowledge sharing. The process of personalization of storytelling then determines the viewer's knowledge base level and access level to knowledge. In a kanohi ki te kanohi (face-to-face) scenario, the kaikōrero also gets feedback from the visitors in the interactions, and will gain an understanding of the visitor's characteristics during the storytelling. In contrast, since the telling and viewing/listening of the stories are two separate stages in the digital experience, such real time reciprocal exchange is not available. However the importance of having information flowing from the visitors to the kaikōrero was still acknowledged. Thus the personas are used as instruments to bridge the telling and viewing. For kaikōrero they are references when telling the story. For the system they are the channels of collecting information from the visitors during an adaptation stage in the viewing experience, before the actual story starts.

A difference in where the visitors are placed in such activity was also noticed, between the traditional context (or say western context) and the Māori context. As stated in Chapter 2, the story content in those traditional contexts, such as in the GLAM (galleries, libraries, archives, and museums) sector, is customized and altered based on the visitor's behavior, intention, and preferences. The system reconstructs itself to deliver the content that fits the visitor's needs. However, in a Māori context, **storytelling is a bilateral process** because it is a process of knowledge sharing and exchanging. The storyteller (kaikōrero) has the authorial control. S/he accesses the knowledge level of the visitor, determines which knowledge is suitable for him/her and matches the "access level", then that knowledge is shared to the visitor via storytelling. Therefore, it is important to point out that, with Māori storytelling, the personas are references for the storyteller to *presume* such knowledge access level, during the recording stage of the digital storytelling, since in this form the content preparation and the actual consumption of it are separated stages.

This is a departure from the assumption of typical user representatives and design anchors familiar in modern design workflows.

Moreover, the motivation of using the digital experience for storytelling with kaikōrero Māori was also revisited after the sessions and persona modifications. The digital experience is proposed to be delivered based on a virtual stage where, because of the distance or travel safety, those groups of visitors are unable to visit the Bluff Marae now, and would like to first experience the storytelling in the digital form (such as CVR). Such experiences are expected to help future virtual visitors to have some initial experience, and will be more likely to one day come to Bluff, visit the marae and listen to the stories face to face with the kaikōrero.

6.4 Content Preparation and System Implementation

For this case study, stories about the whānau wall were captured inside the wharenui of the Te Rau Aroha Marae in Bluff (Figure 6.6 shows a behind-the-scene view of the capture work, including the whānau wall, the kaikōrero, and the 360-degree camera). To create a personalized storytelling experience with viewer interaction included, a series of footage was captured comprising the storytelling guided by personas. Transitional content such as the introduction, welcoming message, and transitions between blocks to make the experience more fluid were also captured and included in the footage. The following sections will describe them in detail.

A kaikōrero did all narrations for the content from the marae. During capture, the choice of language for storytelling was determined by the kaikōrero. They were free to use Te Reo (the Māori language) whenever they thought it necessary. Te Reo could also be used in stories whose major language is English, when certain expressions were needed. The kaikōrero could also choose to tell some of the stories entirely in Te Reo, such as those intended for kaumātua, Māori (*special* cases 5 and 6).

FIGURE 6.6: A behind-the-stage photo showing the capture work at Te Rau Aroha Marae in Bluff. It shows the kaikōrero conducting the storytelling in front of the whānau wall, in side the wharenui. On the wall and other panels around it also filled with carvings and weavings that tell different stories. The 360-degree camera is placed near him, recording everything in the scene.

During early co-design sessions, the kaikōrero stated that they would (most of the time) use a 3-segment model to tell stories (see Section 6.3.2), but using only the *generic* and the *detailed* segments in this project. The decision was made because 1) kaikōrero expected the digital experience would not completely replace the physical visit thus it should not cover all story content, 2) time and resources were limited in this project. The kaikōrero also stated the *Special* cases 5 and 6 would be outside the 3-segment model. Those two stories were recorded separately.

Figure 6.7 summarizes the content segments captured. In general, corresponding to the MS&T model, a story will have either one generic segment and a detailed segment or only one segment (for special cases 5 and 6). Personas 1 to 4 will lead to four takes of each segment. In total there were four generic stories, four detailed stories, and two special stories. The generic segments were designed to be 1-2 minutes, and detailed segments to be 2-5 minutes long. For personas 5 and 6, their generic segments were 8-10 minutes long. Between different visiting groups (personas), different disclosure strategies were expected to be used by the kaikōrero. This could be either the same story told in a different style, or a slightly different story but under the same topic. Since on the whānau wall there are sub-threads and sub-segments, where different stories can be extracted and assembled from it, the choice of which sub-stories to tell was given to the kaikōrero themselves.

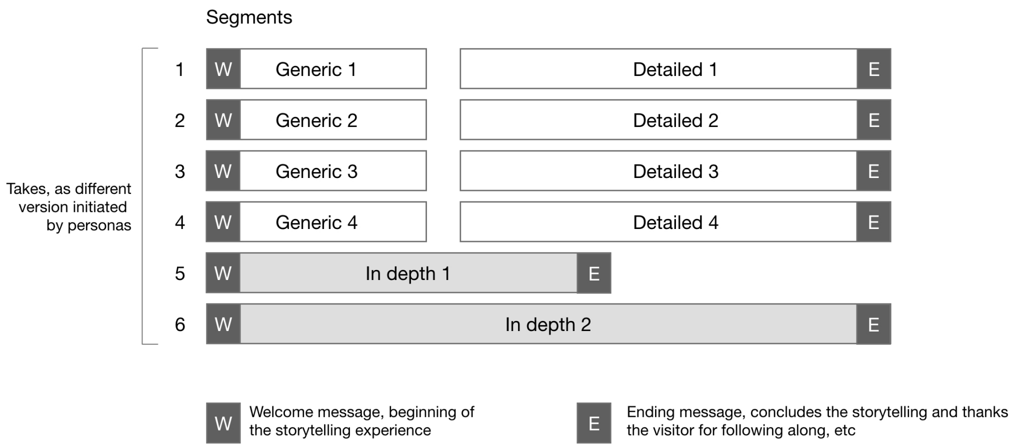


FIGURE 6.7: The detailed capture plan, showing how the generic segments, detailed segments and two in-depth segments are arranged around the six visiting groups. There are also welcome messages and ending messages captured and attached to the main content segments.

At the scene, all content was captured with an Insta360 Pro 2 360-degree camera, at its full resolution of 8K (7680 x 3840) at 30 fps, with HDR on. The camera was stationary in all scenes. Those footages were later downgraded to 5K (5120 x 2560) during the stitching phase, using the Insta360 stitcher. The camera was mounted on a tripod so the viewpoint was fixed in each segment. As shown in Figure 6.8, the positions of the camera and the storyteller were carefully chosen following the *Triangle-system* (see Section 3.1), from the content creation guideline of the APC framework. A separate remote lapel mic was used to ensure the audio captured from the kaikōrero was clear at anytime, while allowing him to move around and make gestures. Before the capture started, one of the six personas was presented to the kaikōrero. The personas were printed on paper, listing all the key features from Table 6.3, plus a group photo collected online matching the persona features, as shown in Figure 6.9. The kaikōrero was instructed to take a few minutes to become familiar with the persona, looking at the key features, imagining the group described on the paper were actually coming to visit so he could prepare a story for them. The process was repeated for each group persona, until all was completed. Each story for every persona group was recorded as a single video clip when captured (generic and detailed segments combined as one clip). However, the kaikōrero was instructed to stand still and leave blanks between the generic and detailed story intentionally. The clips were then post-processed by cutting into two clips at the still part.

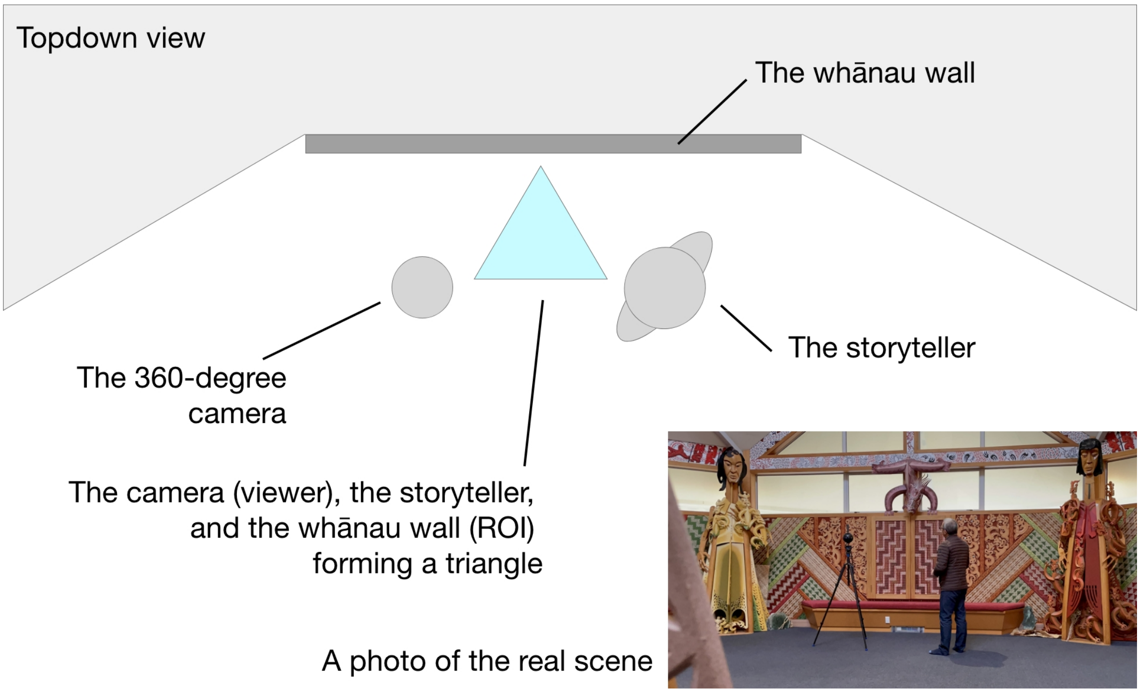


FIGURE 6.8: The spatial arrangement (topdown view) of the 360-degree camera and the storyteller in front of the whānau wall. Notice it follows the *Triangle-system* presented in Section 3.1 so the viewer can later see both the storyteller and the wall and follow storyteller's gestures easily.

6.5 System Development

The system to control and present the recorded 360-degree videos was developed in Unity 2020.3.36f1 and runs on a laptop with i9 CPU and RTX3080 graphic cards to ensure the smooth playback of the relatively high resolution 360-degree footage. The captured footage were first numbered and loaded into the content library of the system, being the *customized content*. A user interface was then added for the viewers to choose which visiting group s/he belongs to, as shown in the screenshot in Figure 6.10.

The case study is designed to investigate whether the implementation of personalization will make a digital experience more appropriate in the Māori cultural context. The accurate collection and matching of viewers to their persona groups was not in the scope of the study (see Section 6.2). Therefore the system was built without a module for persona matching. Instead, the kaikōrero who would join

FIGURE 6.9: The printouts of the personas used both in the capture stage and the evaluation stage. Each printout shows a group photo of virtual visitors on top, representing that group persona. Their key features from Table 6.3 are listed at the bottom. There are six in total.

the evaluation session were presented with a simple prototype system with six personas printed on paper. The prototype system was deployed on a Oculus Quest 2 headset with only the right-hand controller. A ray cast pointing method was implemented for the viewers to choose a visiting group from the six and proceeded to watch the corresponding 360-degree video (Figure 6.10). Noteworthy that in the UI, the texts under the photos cover only their age group, whakapapa, and whether first time visiting, due to limited space on the canvas. However the full details of each group persona were printed and described on paper (Figure 6.9, notice the photos and their arrangements are identical between the screenshot and the printouts) and were provided to the kaikōrero at the beginning of the evaluation stage (see Section 6.6).

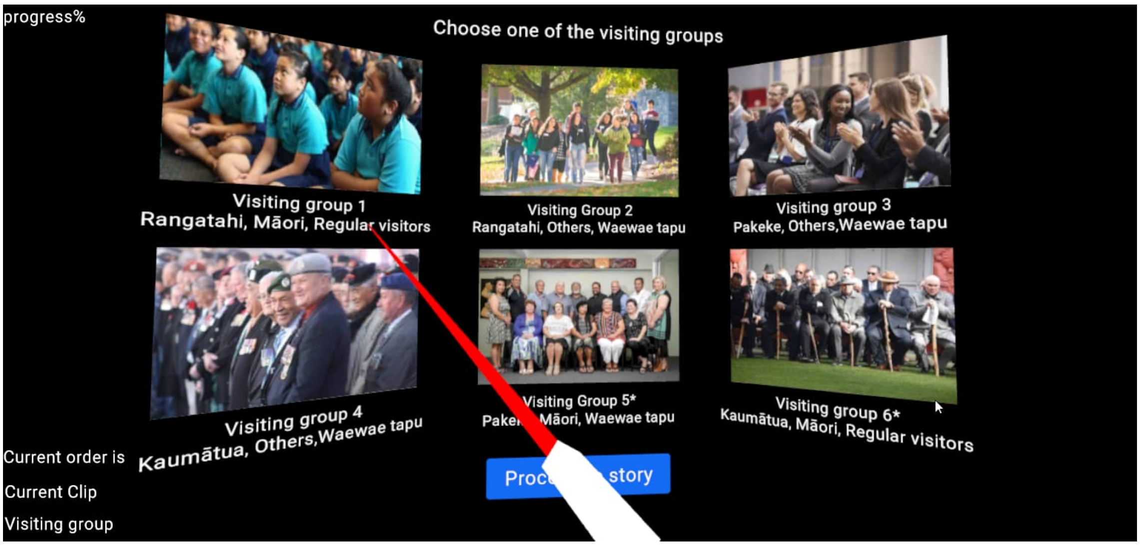


FIGURE 6.10: A screenshot showing the personas selection panel. The virtual representation of the controller and the ray cast are also visible in this screenshot, as the white stick and the red ray. Although the screenshot was taken from the PC running the VR system, it represented the first-person-view of the VR viewer. The texts and digits on the left upper and lower corners are visible only on the laptop for reference purpose only. They are not visible to the viewer wearing the HMD.

For the first four persona groups (1 to 4), a two-part story structure was used, with a *branch* at the end of generic segment to determine how the experience will flow, as explained in Figure 6.11. While watching, the viewer will be first presented the generic segment. In a viewer's eye, the system pauses the video at the end of the first segment and turns on another panel asking if the viewer would like to continue to hear more story (the detailed segment), or end the experience there (as shown in Figure 6.12). If the viewer chooses to continue, the detailed segment will be presented. Otherwise, the system will reset the video player, remove the existing panel, and show the persona panel again (as shown in Figure 6.10).

6.6 Evaluation

One noteworthy observation from the previous study is that the complete structure of the contents was not visible in the eyes of the viewers. Each of them could only see just one disclosure of the story. The viewers did not know that the sequence of how segments are presented was adjustable and each of their behaviors during the experience was involved in determining which disclosure was chosen for a specific case. It was assumed that end users may be unable to perceive,

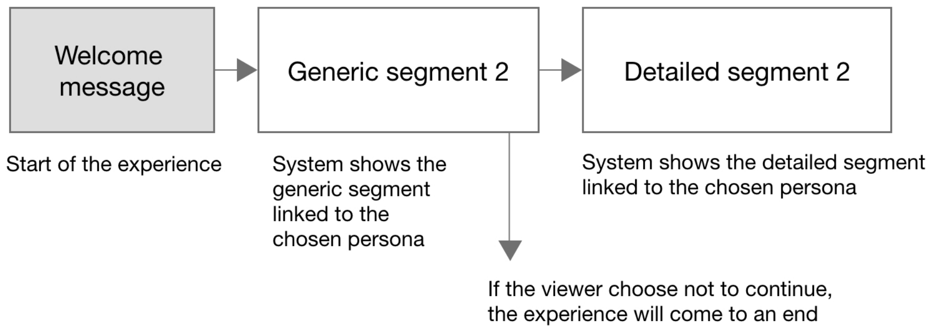


FIGURE 6.11: A flow chart showing how a viewer can go through the experience with a branching point in the middle, between the generic segment and the detailed segment, for group personas 1 to 4. In the case shown, the generic segment and detailed segment connect to the persona 2 were chosen.

compare and then evaluate the variation between multiple story disclosures unless they are explicitly told and exposed of the backstage mechanisms. However, that would require the viewer to revisit the same story several times and try on a series of versions, which introduces significant uncertainty to the impression as the novelty factor comes into play with the repetition. Thus the focus of this case study shifted to the storytellers instead of the viewers. In this evaluation, the novelty factor was removed as kaikōrero were all well versed in storytelling and familiar with most Māori story content, especially those originating from the marae they all have whakapapa to. To explore the appropriateness of the digital storytelling system on replacing kanohi ki te kanohi storytelling, the evaluation compared the characteristics of the digital system to the oral storytelling activities, from Tikanga Māori that they all hold.

6.6.1 Method

The evaluation was conducted as a group session, kanohi ki te kanohi, on-site in the wharenui (meeting house) at the Te Rau Aroha Marae in Bluff, New Zealand. That session had three phases; the briefing, the experience, and the debriefing (group discussion). The phases of the process and the questions for the group discussion were both created in a way that began with open and informal talks before moving on to funnel the discussion towards the important factors anticipated to be

FIGURE 6.12: A screenshot of the panel between the generic and detailed segments, for group personas 1 to 4. Although the screenshot was taken from the PC running the VR system, it represented the first-person-view of the VR viewer. The texts and digits on the left upper and lower corners are visible only on the laptop for reference purpose only. They are not visible to the viewer wearing the HMD.

pushed through evaluation, both of which were inspired by the approach from Kriz et al. [74]. During a briefing phase at the beginning, the digital storytelling system and its feature of personalization were introduced, followed by the use of personas and how the content were captured. The motivation for developing such a digital system and the purpose of running the evaluation session were also explained. Finally the use of the VR headset, the controller, and the digital system itself were all demonstrated.

During the experience phase, four kaikōrero rotated to use the digital storytelling system wearing an Oculus Quest 2 HMD. After putting on the HMD, each viewer got to choose two group personas from the menu and watched the content prepared for those two personas. It took 15 minutes for each viewer to go through the experience. While one kaikōrero was using the HMD, the laptop connected to the HMD mirrored the VR view to a monitor nearby, so the others were able to follow the viewpoint and discuss freely. This non-immersive exposure was specially designed for this evaluation to give users pre-exposure to the experience when they were not actively participating, so conversations and thoughts would start to formulate during the experience phase. It was also easier for the kaikōrero to talk about their thoughts and feelings when moved onto the debriefing phase, and cultivate an atmosphere of free discussion since they had seen how others were doing and had a certain amount of communication beforehand.

After all the kaikōrero had tried the experience, they were first instructed to freely talk about their feelings and thoughts on the digital system, then moved on to a semi-structured interview, guiding the discussion towards comparing the personalized digital storytelling experience with their knowledge of storytelling face to face in the real world. The conversation during the interview was recorded and then transcribed using an online tool, and manually adjusted to correct Te Reo Māori words and expressions. During the interview, two key questions were:

- *Now imagine that one of the visiting groups you selected during the experience is actually going to visit the marae, and you will tell a story to them face to face. At the last moment, however, they need to cancel their visit because one group member tested positive on Covid. How viable do you think the digital version you just experienced is to use in place of the physical visit for this group? Is this an appropriate replacement?*
- *What do you think is necessary to be added or improved in this digital experience, for it to be actually put into use in the marae, such as a setup for virtual visiting?*

Other flexible questions helped the interview flowing, as well as the open questions at the end for extra suggestions, are not explicitly listed here. Opinions and feedback from transcription were then summarized into four main aspects, presented in next section.

6.6.2 Results

Impression of the CVR experience. All kaikōrero stated that the immersive medium (watching 360-degree video via an HMD) gave them a better experience than conventional media such as TV or mobile phone. They pointed out that the immersion CVR provided gave them a feeling of standing there with the storyteller at the scene while listening to the stories, such as *"...what I particularly liked about that 360 experiences sort of [you] sit there and I actually sat there and really listened to [the stories]..it makes sense as if I am really there..."*. The kaikōrero all stated that they paid much more attention to the content presented in CVR, compared to similar content presented via other types of media such as photos, books or video recordings. One example would be from one interviewee he said *"I find myself it's really really difficult to watch a static [flat video], like even on Zoom, and you are getting zero feedback...but here you get the feedback on face expression and body movements and the environment..."*. Such attention difference were not expected by the kaikōrero since they all thought they were already familiar with the content. Additionally, the fact that the stories were recorded in situ (in this case, in the wharenui where the kaikōrero is describing its carved and woven walls and the stories on them), further helped on enhancing the feeling of spatial presence, as one of the interviewees pointed *"I learned heaps on all sorts of all sorts of content was just so interesting...it was relevant and regarded in that virtual reality, he (referring to the storyteller) actually had things to show us. So that cemented the knowledge that he was talking [about]..."*.

Other than that, because the recording happened at the scene, the storyteller could at the same time interact with the objects being mentioned, such as pointing to and touching the walls. In the viewers' eyes, this also helped to increase the realness and the feeling of being there with the storyteller, mentioned as *"...was the art one (group 2) where he was actually interacting with the wall and touching that thing, stroking the different parts and then pushing on the bits and he was explaining..."* and *"for an art class (group 2), it is glad to see specific patterns/art being explained in their original context location"*. All those are not achievable in an audio recording or conventional 2D video.

One extra thing to point out is that they also discussed how the host talked to a virtual visiting group, instead of just an individual visitor. This observation came from the fact they looked at how the storyteller in the scene did not stare at the camera all the time, but instead acted as if there was a group of people, scanning evenly on all the virtual viewers there, across the camera, like one of from the interview said *"...he (the storyteller) was looking at all the different people who potentially there and you are one of a group..."* and *"...in that virtual world, I think I only saw him looks at me two or three times, but everybody else and his gaze until he looked at the camera straight away..."*. Such observation inspired new thoughts on the use of Action Units, which will be touched later in the discussion section.

Impression of the characteristic of personalization. All kaikōrero who experienced the digital storytelling system were aware of the characteristic of personalization. They all noticed the storyteller's variations when presenting the story to different viewer groups. They think it is an excellent opportunity to see how the knowledge in the stories is structured and stacked differently from one to another, like one of them put that *"..a really good thing I noticed about it is that you could actually stack up different users..and if they could group the content according to that..."*. By cross-referencing the personas and the key features listed, they noticed the variation and re-constructing of knowledge from one group to another was mainly brought on by the differences in ages and visiting intentions of the groups. They especially noticed the relatively obvious change in the topic and theme of the stories from group 2 (a group of young art students) to group 4 (Army veterans visiting on ANZAC day).

Realness of the experience and the use of personalization. The feedback on whether the digital storytelling experience is more similar to a kanohi ki te kanohi experience, after the introduction of personlization, was mainly considered after asking the first question listed above. The kaikōrero's discussion included the following points:

1. Their foremost opinion of the digital experience was that it could not be used as a total alternative, because the physical visit is more than just storytelling. As one of the kaikōrero put during the interview that *"visiting a marae is more than listening to kōrero (story) in the wharenui, there are other experience...when people come to the marae you offer them the kai (sharing food together), the pōwhiri (calling onto the marae and formal welcome with speeches), the chatting, etc. You offer them the full experience...This (referring to the digital storytelling) is only part of the experience..."*. In the eyes of the whānau (family, which in this conversation refers to the people from the marae), welcoming people visiting the marae is about showing manaatikanga (hospitality, the cure for people) and building whanaungatanga (the connection between people) between the visitors and the hosts. They do not think this can be achieved simply by letting visitors put on an HMD and go through the digital system.

2. The kaikōrero also expressed that telling stories face to face is a default scenario for them. In the Māori context, storytelling is a reciprocal sharing process of knowledge. This means that while the storyteller is telling a story to the visitors, s/he is also learning from the visitors. The visitors can also share different opinions, ask questions, challenge the authenticity of a story (as one story will have different disclosures and versions between different iwi and whānau), and interact with the storyteller. Such interactivity was also seen as the process of *building connection* before knowledge sharing, like one interviewee put *"..started to talk about that connection...connection was pivotal before we can even get the content. So the tikanga of that is required to come on to the marae is going to be part of the learning experience as well, not just the fact of content that [storyteller] has given us..."*. Until interactivity is available in the digital experience, they do not think it can be a complete alternative to the physical experience.

3. A partial replacement is possible. Because the system is designed with personalization features, the digital system can provide appropriate content to various visiting groups, such as school and community groups (which are the two types that happen quite frequently, making up to eighty percent of the regular visits). They think it can be used as *"one of the many ways to share knowledge"* and can replace at least the activity of giving the visitors basic Māori kōrero, such as the story of how Ranginui (the sky) and Papatūānuku (the earth) were separated by their children. It was highlighted in the interview as well, e.g. *"...then someone say for example, [name hidden] who has been here for the first time and it has given the overview kind of story (generic version)..."* and *"..ensure that they (the kaikōreo at the marae) were not overburdened [with] these types of [storytelling] and get the concern out of the head as well..."*. Using a digital storytelling system would help to relieve the Kaikōrero from the burden of repeating the same generic story to each visiting group, and allow them to focus on telling stories with detailed content tailored to the specific groups.

In sum, the kaikōrero believe the introduction of personalization helped to improve the feel of realness and content flexibility of the storytelling experience, and that it would be appropriate to use such a digital system as a part of the marae visiting experience for the guests. However, they would not like to propose replacing the physical visiting experience with the digital one. While they believe the physical visit to the marae is a unique experience showing their manaatikanga and building whanaungatanga with people, they did admit that the digital system could be applied as a parallel option to deliver generic and entry-level kōrero to a wider group of visitors, especially in the time of pandemic or where people are scattered in distant locations.

Interaction capability as a viewer during the experience. The kaikōrero feedback converged on viewer interaction during the storytelling experience when asked about future upgrades they would like to see in the system. When asked about the reason behind it, they stated that in Tikanga Māori they treat the storytelling activity a type of bilateral interaction, with viewer involvement in multiple factors, including the pace, topic and one thing unique to their culture, challenging the storyteller. The kaikōrero started by listing the interactions they would like to see between the viewer and the storyteller/host, such as being virtual viewers, they would like to be able to interrupt the storytelling process, raise their hands, as questions, or push the topic towards something they would like to see, e.g. *"...when he (the storyteller) was talking about Māui (the great hero in Māori mythology) and then* *his brothers on one of the carvings I wanted to say 'wait, talk more about Māui...'"* and *"..like he touched on the Tukutuku panels (a distinctive art form of the Māori people, a traditional latticework used to decorate meeting houses) with the art students...he talked about the different materials, how the colors were made...there could be people want to pause and ask follow up questions around around the specifics..."*. These observations all show that they are comparing the personalized digital storytelling experience with a real life one and trying to have the capabilities of a physical real visit experience. One unexpected factor mentioned was the change of the height of the camera. Especially for the children's group, the kaikōrero pointed out that in a marae visit (physical one), it is common for the children to sit on the floor and look at the kaikōrero. Such a scenario was not considered in AUs or phase one of the APC, as maintaining the camera height at eye level was initially recommended in the APC guidelines.

Updates to the AUs were again mentioned. Since the content used in this study is kōrero about the wall carvings and weavings, the kaikōrero also mentioned that they would like to see visual cues (e.g. highlight different zones of the wall, show text prompts) assisting them to quickly identify which part of the wall was currently being introduced by the storyteller. Since AUss were used during capturing, a follow-up question was asked to see if the viewers noticed the storyteller was using pointing gestures to help them locating the ROI. The feedback indicated that they thought the gestures were insufficient in such a scenario, especially when they turned to look at the wall. When the storyteller was not visible in the FOV, they could not follow the pointing gestures and got lost.

They also talked about the necessity of controlling the time length and pace of storytelling. For some groups, such as the rangatahi (children and young youth), ten minutes or more would be too much for them, especially for a passive learning style. But the kaikōrero added that since the system has enabled self-control over the pace, it could mitigate the issue of time length over visitors' attention span. This additional interaction feature would allow the viewer to choose whether to stop after a generic segment or continue listening to a detailed version of the story. Other than that, since it will be a self-guided tour style, if implemented, a pause function would be favorable in case the system user is interrupted by something else. It was later treated as a non-optimal option since philosophically the CVR is still believed to be more like an immersive experience of virtually visiting a place, instead of watching a clip of omnidirectional video. Adding a video play function, such as pause, may lead to a break of presence, as it signals to the viewer that s/he is watching a recorded video.

6.7 Discussion

The APC and its two components were introduced in Chapter 5. In this personalized storytelling project in the Māori storytelling context, both components of the APC have been put into practical use. In this discussion section, three topics will be introduced to reflect all steps throughout the case study, from content preparation to system deployment to evaluation. Those topics are:

- The use of immersive content with APC for Māori storytelling;
- Adaptation of APC to appropriately tell a Māori kōrero in digital media;
- Modifications needed for the APC framework.

6.7.1 Using CVR with APC for Māori Storytelling

The appropriateness of fitting Māori storytelling into the APC framework in terms of applicability was determined in the early co-design sessions that on the one hand, the content of this project would be in the form of 360-degree videos and the topic would be cultural-heritage oriented. The storyteller would be visible in the scene, and content creation guidelines from the APC would be applied during the content capture, including the spatial arrangement of the camera, the storyteller and the key objects in the scene guided by the Triangle System (see Section 3.1), the use of AUs (see Section 3.2) by the storyteller during capture, etc. On the other hand, viewer participation in the storytelling process was also carefully considered and implemented in this project. The viewer interaction guideline from APC recommends that content creators have a plan about the viewer's role in the story and her/his capability and involvement in the story in advance when composing the storytelling experience. Following this recommendation, several discussions and co-design sessions with kaikōrero Māori were conducted around whether the technology of CVR could be used for Māori storytelling, under the guidance of APC. During those discussions, the group focused on the role of a viewer in a real-world storytelling scenario and talked about how the viewer participates in a storytelling process in the Māori context.

During the production stage, the APC was used as a key guideline, helping the kaikōrero and me to decide the locations of the camera and the storyteller, and applied AUs when telling the story. The viewers' impressions of the content, as described in the Section 6.6, have proved that the stories captured in the form of CVR, under the guidance of APC, are usable and effective. The kaikōrero who reviewed the system also liked the design where the viewer could choose whether to move onto a detailed segment or stop the experience after the generic segment was completed. They thought this choice gives the viewer a control of the pace and the amount of time s/he will need to invest into this experience. One can conclude that the design of giving viewers self-control to some aspects of the storytelling process, which was highlighted in the second part of APC, is also effective in this case. In summary the digital storytelling system (CVR content produced with APC guidance) is usable and effective in Māori storytelling, as an immersive medium for sharing of knowledge in the sector of cultural heritage. It also answers the part 2 of the RQ_3 listed in Section 1.2, proves that a digital version of personalized storytelling can give viewers a similar impression of storytelling physically face-to-face.

6.7.2 The Adaptation of APC For Cultural Appropriateness

Although APC can be used as a guideline to produce a CVR experience for Māori storytelling from this case study, its utilization by Māori people (who are the stakeholders and the owners of their storytelling and mātauranga (Māori knowledge)) and its appropriateness requires further exploration and inspection. In this section, part 3 of the RQ_3 listed in Section 1.2, i.e. whether a personalized CVR experience be an appropriate tool for Māori storytelling, is also answered.

The first step was the employment of *personalization* into APC. From the co-design sessions (see Section 6.3.2) it was already clear that in a Māori cultural context the notion of *listeners also participating in the storytelling process* is different from the implementation envisaged during the user study for APC (Chapter 4). Instead of referring to a viewer's specific input, such as gaze or gestures, kaikōrero Māori look at viewer participation at a higher level. They treat the process of storytelling as a way of reciprocal sharing of knowledge (mātauranga). While the viewer is learning knowledge from the story presented by the teller, the teller is also gaining new understanding and thoughts from the feedback from the listeners. They believe that the viewer's personal traits will, from the beginning, influence how the storyteller is going to present the story, such as the choice of what content segments to use around a given topic. Such influence also moves on into the storytelling process. The storyteller will dynamically adjust both the content choices and the style of disclosure of a story based on the viewer's response and feedback. The viewer will also show various attitudes towards content segments based on the teller's performance and skills.

Thus, in this case study, the viewer interaction design guideline is employed and customized for Māori storytelling context. Instead of using the viewer's real time input or gazing activities, the personal traits, through *personas*, were used to emphasize viewer participation in the storytelling. Generally, the APC is not only employed in this project, but also modified to adapt to the unique Māori cultural context of the project. It was also learned that acknowledging the differences between listeners (visiting groups) is a key factor in Māori storytelling activity in the physical world. With the motivation to provide a digital alternative for scenarios when remote whānau members cannot be physically at the marae but still wish to experience a story, the viewer interaction design guidelines from the APC framework were employed during implementation and we chose to implement viewer interaction in the form of *personalization*, from which different types of viewers of the same story will have a slightly different experience and feel the story was tailored specifically for her/him.

The kaikōrero's thoughts on the characteristics of *personalization* of the digital experience were especially highlighted during the evaluation, in order to learn whether this characteristic (comparing to other pre-recorded storytelling experience) had contributed to making kaikōrero think that the digital storytelling system is appropriate enough to be used as an alternative when physical face-to-face storytelling is not viable.

Referring to the MS&T model in Section 6.2.1, the content in the digital system are treated as a bundle of several parallel takes as a dynamic structured system. The kaikōrero were aware of the variation from one version of the story to another, introduced by the different visiting group personas. On the perspective of usability, they thought it could be a good example of involving the visitor (viewer) in the storytelling process, following Tikanga Māori, in which they regard the activity of storytelling as an activity of reciprocal knowledge sharing the interaction, plus sharing corresponding knowledge to a given visiting group.

The implementation of personalization also means creating customized content, in which the local kaikōrero are actively involved. It is a way they prefer to see as the topic *letting the one who knows the story well to tell the story, and tell it appropriately to the viewer* is attracting substantial attention among the Māori communities, indicated by the statements from the interview such as *"...when preparation* *for class materials one of the requirement is we want the content to come as localized...for a content here [Bluff] you would like to have someone from down here to talk about it...", "...we also need to go into the local marae and ask the knowledge holders, ask them what type of mātauranga they would like to share and how they want to share it..."* and *"...because it is the content of mātauranga that the local marae would like to share and they are sharing it [being] really vibrant and different..."*. That feedback points toward the assumption stated in the beginning of this chapter, that the implementation of personalization, letting visitor types trigger different versions of stories, is also a type of viewer (visitor) involvement in creating and delivering stories. It is safe to say that the APC guidelines had been successfully adapted to the cultural backgrounds and unique Tikanga Māori in this case.

However, if one looks deeper into the use of the digital storytelling system powered by APC guidelines in the realm of cultural knowledge transfer, especially from the perspective of kaikōrero Māori, several outstanding observations would emerge. The digital system used here succeeds from the merger of APC guided CVR content with viewer involvement in the form of personalization, and was proven to be at least usable and effective. However, when it comes to whether the digital storytelling experience can be used as a replacement for a physical storytelling experience at the marae, the kaikōrero's opinion turned out to be unexpected. In a short version, they found the digital experience can have its place at the marae, but is definitely unable to replace the physical face-to-face visit. This is because in the Tikanga Māori context, a visit experience to the marae is more than just *visually* looking at the buildings and carvings or *audibly* listening to a story; there are other activities that have to happen face to face physically. It was not the first time the kaikōrero talked about using a digital tool as a *pathway* to physical visits. During the the co-design discussions before capturing the content (see Section 6.3.3), kaikōrero also described their expectation of the digital experience acting as a teaser or preview, and bringing people back to the marae later. The Kaikōrero all commented that the digital experience should not be a replacement but an extra way to deliver some content. One of the significant concerns they showed towards the digital storytelling experience was that if this digital experience becomes available someday in the future, it will stop people from visiting the marae physically. Although they agreed on the point that with the introduction of personalized content the digital system can capture and deliver the kōrero deemed appropriate for each visiting group, the kaikōrero would prefer to see the digital storytelling experience co-exist with physical visits. They would like the digital storytelling experience to help attract more people willing to visit the marae physically, instead of replacing it. Besides, the digital experience can also serve as an *entry level crash course* for the visitors to gain a basic understanding of the kōrero and tikanga Māori. When they later move on to the physical face-to-face storytelling with kaikōrero, the latter will know that those visitors have a certain background knowledge because they had gone through the digital experience. It also helps kaikōrero access the knowledge level of the visitors, to avoid giving the basic kōrero to all groups while not knowing which group needs it and which does not.

We should also realize that the use of personalization and narrative variation in this CVR case study is the central focus of exploration towards the APC framework. The viewer's participation in the storytelling process is moved beyond the simple temporal succession (which was discussed in the original form of APC in Section 5.3.2) with personalization enabled. It is now at the level of alternating the disclosure of the content. The assembly guidelines are extended beyond just viewer-interaction design. The co-design sessions also explored the most appropriate method to use predefined personas to help storytellers bring variation into the presentation and capture the customizable content. It adds to the content creation guidelines for CVR creators. In sum, this case study validates the APC to be usable in Māori storytelling. It also illustrates that adaptation and careful consideration are necessary when using the APC for different storytelling scenarios. More importantly, creators also need to learn from the stories' owners and the authentic people where the stories come from before finalizing the design of a CVR experience, to ensure its appropriateness as a digital tool before handing it over to the end users.

6.7.3 Inspirations for the APC Framework Upgrades

Moving from viewer interaction design (part 2 of the APC guidelines) back to immersive content production guidelines (part 1 of the APC), one could summarize that both the Triangle System and the AUs applied in this project were effective. They helped the viewers acquire the natural feeling of being in the virtual scene with the kaikōrero. However, the variable camera height is a new factor inspired by this study with the kaikōrero discussion. In the original guideline, it was recommended to set the camera height to match the eye height of the storyteller/host (see Section 5.3.1). In common daily scenarios, the eye height of every participant stays the same. But as the kaikōrero pointed out, the marae visiting scenario is quite different from those above. They would like to manually set the camera to different height levels to accommodate different visiting groups, to influence or mimic the relationships between the kaikōrero and the visitors during the storytelling. One example will be the content for rangatahi (youth) visitors should be captured with the 360-degree camera placed lower than normal, closer to the floor. Because the young visitors normally sit on the floor in a real world scenario, so they should also have a lower POV in the digital experience to match the real world experience. It is also believed to be an essential factor to consider when applying the APC framework to an actual cultural-heritage-oriented case.

The feedback on the ineffectiveness of AUs when the kaikōrero is not visible in the viewer's FOV matches the limitation of AUs mentioned in Section 3.5.3. One solution is adding extra visual cues constantly visible to the viewer to make the AUs more effective. In future content capture tasks, it is also practical to teach the storytellers to maintain their pointing gestures longer than what they would normally do in physical face-to-face scenarios. So in case the viewer looks away and missed the gestures once, s/he can turn back to the storyteller to pick up the guidance again. Discussion on AUs also brought attention to the application of group settings of Māori storytelling into the digital experience. When capturing the content, a group visiting scenario was used to match how kaikōrero would normally tell stories. It introduced changes to the use of AUs such as instead of maintaining eye-contact with only the camera, the host needs to look towards left and right sides of the camera, pretending there are more than one person (the viewer) in the scene. The APC guidelines still apply if a group of people is actually present on site (although this did not happen in this case study), with the 360-degree camera placed among them, making the viewer feel like s/he is one of the group when watching. However, a group setting was not applied to the viewer's side. In this study the viewer still watched the content via a HMD as an individual experience. The scenario of *watching the same CVR content together with multiple audience* is beyond the scope of this study. Works done by other researchers showed that extra functions and new technologies will be needed to accommodate a multi-viewer scenario [70, 142].

Suggestions on the need to give viewers control over time and pace were also mentioned in the feedback. They also expect more explicit visual cues to be added as assistance to the AUs implemented by the human character in the scene. Those have been already explained in Sections 4.6.1 and 4.6.2. It will be further discussed in Section 7.3 as potential future work on both the two parts of the APC framework.

Overall, this case study highlights the adaptation of personalization into CVR and the evaluation of its appropriateness for being used as a storytelling tool for the marae. First, the co-design sessions with the kaikōrero gave a clear and abundant look into what factors they consider when telling a story to visitors, plus how they think about the activity of storytelling itself. It led to the formulation of the personas as a tool for digital storytelling variation. Second, although the *visiting group* personas were used as a tool to help guide the kaikōrero to prepare story content before actually telling them, a few processing mechanisms behind it remain unknown. Those unexplored mechanisms include how different knowledge blocks were chosen, assembled and disclosed during the storytelling, how decisions about story variations were made and how those variations were managed. Third, as stated above, the digital version can be used for immersive storytelling and getting visitors to know some basic kōrero of the marae, but remain far from, and probably never will replace, the physical face-to-face experience.

6.8 Conclusion

The personalized immersive storytelling system was implemented using the APC as its guidelines, with Māori storytelling as its context. At the beginning of this project, several co-design sessions were conducted with kaikōrero (Māori storytellers) about storytelling with personalization to elicit the key factors most important for tailoring content toward different visiting groups. During the sessions, *personas* were chosen to present different types of visiting groups. Using the knowledge learned during the co-design sessions, six personas were then made, representing six different types of visiting groups. The personas were reviewed by the kaikōrero and adjusted before settling on their final versions.

After the co-design sessions, I physically traveled to the marae that the kaikōrero have whakapapa to (are genealogically connected to) and conducted phase two of the project. There, two of the kaikōrero telling stories about the whānau wall were digitally captured using 360-degree videos. During the capture, six personas were used to guide the kaikōrero (who told the stories) to imagine scenarios in which they would prepare and tell the story to each of those groups visiting the marae. This led to six versions of the story, corresponding to the six personas. The captured content was then loaded into an immersive storytelling system in a VR headset.

In an evaluation session, a group of kaikōrero were first asked to select a visiting group and watch the version of the captured story linked to it. When all kaikōrero invited had used it, a group discussion was conducted in a semi-structured interview format, in which they evaluated the appropriateness of this system. Open-ended questions were used to explore the digital nature of the experience compared to the face-to-face experience, and the variations based on visitor type. The results indicated that this digital storytelling experience with personalization design can be a viable and appropriate system for sharing cultural knowledge with visitors to the marae. The kaikōrero thought the digital system could share an experience of face-to-face storytelling to people who are remotely located and unable to visit the site physically. However, they would not like to use it as a complete replacement for a physical visit. Rather, they hoped people would be more willing to personally come to the marae after experiencing the digital version of the stories.

The case study also proved that the APC can also serve as a guideline for building interactive immersive storytelling projects in the cultural realm. However, it is necessary to understand the storytelling activity in the target realm before designing a story, since both the *role of viewer* and *viewer participation* will contain various interpretations. The viewer interactions mentioned in the APC will need to be adapted to those specific interpretations to reach their proper form in each different realm. It is also important to understand the expectations of key stakeholders (such as the kaikōrero at the marae, in this case) and what outcome they would like to have from those storytelling experiences. Once those are established, the interactive CVR with APC support can be a powerful tool to deliver a satisfying experience, where the viewer will find her/himself not only spatially in the virtual scene, but also socially being there as the virtual story world can be influenced by the existence and behaviors of the viewer. Such experience is not achievable via a conventional 2D video or a traditional 360-degree video. The narrative variation is introduced into the content during the production stage. The personalization stories are then delivered when the viewer is experiencing the content later. The creation of narrative variation can be independent of the existing system with which the CVR experience is implemented. In the future, extention to the APC framework is planned to cover not only 360-degree video based CVR, but also 6DOF ones such as volumetric capturing, as personalization storytelling also applies to them once the designer modifies the delivery stage accordingly.

Chapter 7

Overall Conclusions and Future Work

In this chapter, the user studies for this thesis will first be summarized and reviewed, including their results and the reflections gleaned from them throughout this thesis. This is followed by an assessment of the study's limitations, and lastly recommendations for potential research directions and future plans.

7.1 Recap of User Studies in this Work

This thesis investigated and produced guidelines for CVR creators to help them introduce engaging and enjoyable immersive storytelling experiences. These guidelines provide a solution to CVR production challenges such as narrative paradox (NP), fear of missing out (FOMO), the dissonance between the viewer's expectation of agency and the actual level of interactivity, and the lack of acknowledgement of the needs and tendencies of specific individual viewers coming to the experience. Results and reflections from the studies in this thesis were consolidated to develop an Adaptive Playback Control (APC) framework for cinematic virtual reality (CVR). The APC framework emphasizes not only the need for creators to use diegetic and on-site attention guidance cues in immersive media but also the importance of design taking into consideration the balance between the viewer's and creator's role in terms of control over the narrative progression and the visibility of interaction affordances in the immersive storytelling experience. Broadly, the goal of this research was to propose a way forward for balancing the involvement of both the viewers and the creators via the APC framework. It is expected that using this method creators will produce content that considers both viewer engagement and the individual viewer's needs from the first step of content design, and that the future experiences they create will meet viewer expectations as well as be exciting and diverse.

An overview of relevant background information knowledge was given in Chapter 2 of this thesis. The chapter provided an extended description of the current research work around CVR content production and the involvement of viewer participation in immersive media. Areas it covered included early discoveries of the how viewer's role was different in this new medium and new grammars and various solutions adopted to counter the NP issue, when creators started to use VR, mainly 360-degree video, as a storytelling medium. Discussions around viewer interaction and limits to this in an immersive storytelling experience were also covered. The chapter also examined more current topics about viewer involvement in CVR, such as the complex and customizable content format used in approaching the issue and the introduction of viewer personalization.

Chapters 3 and 4 discussed the studies conducted to investigate several unresolved problems in CVR storytelling with viewer interaction. In Chapter 3, Action Units (AUs) were introduced as an addition to the creator's toolkit, serving as attention guidance cues used by on-site actors to direct the viewer's attention onto key story elements, mitigating the issue of FOMO. The use of AUs was compared with two other commonly used synthetic cues, the pointing arrow and the angular shift, demonstrating that with AUs the viewer experienced increased feelings of enjoyment and engagement with the story and reduced cybersickness. AUs were also preferred by viewers for their diegetic characteristics, and by creators as they are easy to implement.

Building upon the content production guidelines that emerged from Chapter 3, exploration of the viewer's participation in CVR storytelling continued in Chapter 4. The two topics discussed were control over the narrative progression and the visibility of interaction affordances. An interactive immersive guided tour was produced with four variants to explore whether the viewers tend to want to take active control of the narrative and interact with the system explicitly. Results and observations indicate that viewer control is recommended for immersive storytelling projects that focus on providing an experience of participation and a feeling of agency. Careful design will be necessary if the creator wants to expose interactors to the viewers and attract them towards more exploratory behaviors.

Then in Chapter 5, insights from these studies were consolidated to develop the APC framework. This framework provides a comprehensive reference system for creators who work on interactive CVR experiences and are willing to involve viewers in the narrative progress. It provides a series of guidelines on content preparation for a better viewer experience with less FOMO for the design of CVR experiences with viewer participation and interaction enabled. The APC framework also covers factors to consider when designing the viewer's capability in CVR experiences with various emphases and focuses.

In Chapter 6, a co-design project with *kaikōrero Māori* (Māori storyteller/s) collaborators was introduced as a case study aimed at creating a digital experience using CVR. The digital experience aimed to offer viewers virtual visits to Te Rau Aroha Marae in Bluff, New Zealand, including a storytelling experience from kaikōrero, done in a way very similar to a real-world, face-to-face experience. My goals with this case study were to see if an APC-guided CVR experience could be used by kaikōrero Māori in their community, be culturally appropriate, and follow Tikanga Māori. The APC framework was applied to guide the production of content, capturing the kaikōrero orally telling stories to different groups of virtual visitors. The concept of *personalized storytelling*, in which disclosure of the story changes according to a viewer's profiles (responding to who the viewer is), was also explored, as it was based on the natural practice of Māori oral storytelling in the real world. This characteristic was incorporated into the design to implement the expected personalized digital experience.

Co-design sessions were conducted with storytellers, using a qualitative approach, to explore the appropriateness of storytelling personalization. Insights from the co-design sessions were used to refine the construction of the digital system. The system was then tested and reviewed by the same group of kaikōrero for user experience and cultural appropriateness, compared to the traditions used in real-world face-to-face storytelling. Feedback indicated that design of the digital storytelling experience (CVR) to include personalization could be used as a viable and appropriate system to share cultural knowledge with visitors to the marae. The APC guidelines helped make the immersive experience more effective than conventional media (such as 2D video on a screen or game-like 3D experiences) in giving the viewer an impression of "being there" in the wharenui with the kaikōrero.

This case study also explored how APC can be used as a guideline for incorporating viewer participation into immersive storytelling projects, especially in a cultural realm with a specific understanding of storytelling. Insights included that although the digital experience can be used in certain scenarios in its ability to provide visitors with a basic and universal storytelling experience, it should not be seen as a complete replacement for physical visits to the marae. On the contrary, kaikōrero expect the digital experience to serve as a teaser to attract people to come to the marae and discover more stories face-to-face. This insight went beyond the initial expectation of the case study, which was the simple determination of whether personalization can be effective. The importance of understanding the storyteller's expectations when an experience is delivered was thus revealed, especially with regard to what they expect the viewer to feel and gain, rather than simply aiming to duplicate a real-life experience with a digital system.

7.2 Contributions

The main contribution of this thesis is the Adaptive Playback Control (APC) framework, as a set of effective storytelling guidelines for creators working on guided-tour type CVR content. The APC framework presented in this thesis provides support for both *content creation with viewer participation in mind* and *viewer interaction design that involves the viewer as a participating character in the storytelling scene*, as a combination of:

- **The immersive content production guidelines** including the *triangular rules* for immersive scene setup and *action units* (AU) as attention guidance techniques, aimed at resolving the issues of narrative paradox (NP) and fear of missing out (FOMO);
- **The viewer interaction design guidelines for CVR** which emphasize the level of interaction the creator can enable for the viewers, balanced against the need for authorial control to remain in the creator's hands, and the visibility of the interactor if one is enabled.

Both guidelines were developed to help CVR content creators build prolonged, immersive, enjoyable, and engaging storytelling experiences.

Besides describing how the APC framework was established and tested, this thesis also presents a case study, which demonstrated the use of APC in a cultural-oriented storytelling project co-designed with Māori research collaborators and kaikōrero. The APC framework was adapted to the Māori storytelling context by adding the feature of personalization to CVR. It shows how the different attributes of viewers can be acknowledged in a CVR storytelling system and narrative variation be employed in response to these attributes. The case study also looks at design implications for content creators in terms of:

- how research can be conducted focusing both on storytelling experiences and cultural heritage preservation of and led by indigenous people;
- how non-Māori creators can adopt western-view-oriented storytelling guidelines (APC) to create Tikanga Māori–guided storytelling experiences using digital tools.

This case study broadly presented a process of modifying the APC framework for a storytelling project with a specific cultural context, as well as the experience design and content production work this entailed.

Overall, the work in this thesis first propose the APC framework for guided-tour-style CVR experiences, but also extendable to other genres of CVR, and to a 6DOF experience in some circumstances. It then provides a demo case for CVR creators showing how the framework can be adopted and used to guide their production work, including choosing the story structure to support their purpose with storytelling, preparing content segments, and assembling them with interactive elements designed for viewers, culminating in an engaging and satisfying immersive storytelling experience with interactivity.

7.3 Study Limitations and Future Work

This thesis's limitations became apparent in reviewing the processes and results of each study. In study one (Chapter 3), in the 360-degree videos presented to the participants, the AUs were not constantly directing the viewers throughout the entire process of search and attention. Since the AUs were implemented by the storytellers, they could only be perceived when the viewer was looking at them. In other words, the AUs were only effective when they were visible in the viewport. Once the viewer turned away, such as to look for a region of interest (ROI), the AUs were no longer visible and their function of attention guidance was suspended. In the previous chapter where the results of time to search (TTS) and time on target (TOT) were analyzed, it became apparent that attention guidance was not instantly effective but rather went through a build-up process from the time of its appearance. Thus when a viewer's total "exposure time" under each guidance cue was counted, those of AUs were not as high as PA. One possible cause of the differences in exposure to cues was that the build-up in effectiveness of AUs was closely dependent upon a viewer's actual behavior when watching the video. The pointing arrow (PA) and angular shift (AS), however, had lower dependencies and faster build-up processes. Future work should look into integrating high-fidelity spatial sound, including simulated reverb effects, to reinforce the effectiveness of AUs in directing viewer attention.

Also, the video clips were all about 90 seconds in length. In some results, the level of enjoyment (E2) and recall rates showed similarities between AU, PA and AS, without any significant difference. One possible reason could be the relatively short time (six minutes) subjects spent watching all three videos, potentially not enough for the differences between conditions to be revealed and detected. In addition, in the materials the PA and AS were both added during the post-production phase while the AUs were implemented during the production phase. It is possible that implementing AUs becomes an extra task added to the storyteller beyond the primary task of “presenting the content”. The increased workload and related effects on the storyteller when using AUs while also rendering the narrative is also worth further exploration. Moreover, although the viewers were assumed to be focused on the narration given by the storyteller and less aware of memorizing the visual details of ROIs when they felt higher engagement and enjoyment, the correlation remained vague and unconfirmed. A concrete conclusion as to whether AUs, as diegetic cues, can be less distracting than the non-diegetic PA and AS also could not be reached due to limited data gathered.

In future studies, the design of AUs will be iterated with additional cues that are constantly visible to the viewer to ensure they are taking effect throughout the viewing experience. More prolonged video footage will also be used as study material, instead of 90-second clips, to ensure longer immersion for participants. New filming locations and ROIs will also be carefully chosen for further study to mitigate the distraction from cues to the viewers when watching. Lessons learned from AUs in swivel-chair VR can potentially be applied to other forms of CVR, where the viewer’s movement within the virtual environment is less constrained. In particular, careful integration and combination of AUs and other cues to “walk” the immersed viewer toward spaces of interest will be applied.

In the second study (Chapter 4), the conditions were relatively simple compared to the broad design space that the APC framework covers. Looking at who should control the narrative progression and what the visibility of interactions should be were only the starting point. In the user study, each participant only experienced one of the four conditions (CC, CR, VI and VE), which may not have captured the differences between the given conditions, i.e., *creator control* (CC), *creator control with randomly rearranged segments* (CR), and *viewer control with* *implicit input* (VI). This might be the reason for the lack of significant differences among the objective measures. The potential relationship between the viewer's tendency toward exploratory behavior and how the tasks were introduced before they started the experience, was also not foreseen during the experiment. The main focuses of the user study were 1) the visibility of viewer interaction methods while limiting the form of viewer input to controller-based laser pointer only, and 2) viewer interaction limited to affecting the narrative progression only. The effects of other parameters with a possible connection to viewer control, such as temporal sequencing, pace, and speed of progression, still need further exploration. We plan to conduct further user studies looking into those topics and add them as pieces to form a more comprehensive APC framework to cover various demands in effective CVR content creation and production.

Another factor is that all materials used in the user studies were virtual guided-tour styled experiences. They all belong to one type of story structure, the "hub-and-spoke." In non-linear storytelling there are other story structures suitable for CVR, including detours, the string-of-pearls, and the gauntlet (unlike hub-and-spoke, these are linear based but contain sub-branches; see [24]). Both viewer control and interactor visibility designs were not tested on these structures. Even within the one hub-and-spoke type used in the study, the situation varied depending on the actual content. The arrangement of the objects in the space captured for the material for that user study is illustrated in Figure 4.2. Before that, a prototype system was implemented based on a series of 360-degree videos captured in a campus greenhouse as a trial run. Unlike the lab's space in Figure 4.2, the greenhouse has a mixture of indoor and outdoor scenarios, as shown in Figure 7.1. During production, the camera was initially placed in the mid yard (the hub) when capturing the introduction clip, then placed in two greenhouses and by the river bank (the spokes) for the main segments. In that prototype experience, the tour started at the outdoor yard in the center of the scene, which is also the hub, as the greenhouse scene is also part of a hub-and-spoke structure. But when the viewer moved into an indoor node via one of the spokes, the view of the scene was blocked and her/his choices were downsized and limited to either elements within that indoor space or returning to the yard. Although the overall structure stayed the same, the availability of choices for a viewer did not. In the lab scene for the user study, all nodes were visible to each other and the viewer's choices remained the same at any node. In this greenhouse prototype, however, the number of choices a viewer was given at a time depended on her/his location inside the scene, due to the spatial features of that space (in this case, the mix of indoor/outdoor spaces). The comparison between the greenhouse prototype and the lab tour used in the study implied that although the story structures stay in the same category and all input and interactor designs remain identical, the freedom on choices a viewer is given can change as the scene changes. In real-life scenarios, the space a creator chooses to create an immersive storytelling experience may differ in real topological relationships, and can affect on the viewer's experience. One solution could be to employ multilevel structures such as a spoke also serving as a hub and leading to further "sub-hubs." This idea was not explored further in this thesis. Whether the guidelines from the APC framework apply to the other types of non-linear structures remains to be evaluated.

The issue of embodiment of the host in guided-tour-style experiences is also worth exploring. In this study, some participants provided feedback that they preferred a narration-only style where the host was not visible in the scene but rather a disembodied voice-over. Because they thought the host would attract their attention and they would therefore be less likely to explore for fear that doing so would be seen as disrespectful, voice-over might lead to less-restrictive behavior. Although the voice-over style is out of the scope of the current APC framework, this does point out the necessity of considering the other side of the attention-directing methods being discussed. Since the embodiment of the storyteller is not restricted to 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, thus avoiding the feeling of "fear of insulting the host." These will all contribute to constructing an effective APC framework.

Later, when the APC framework takes shape, it is expected to give creators a sense of familiarity akin to scripting for conventional videos and provide them with the confidence of authorial control, along with satisfactory outcomes throughout production, delivery, and later consumption. Thus, creators who apply the APC framework as their production reference will at least be more assured when facing the challenge of making immersive content, knowing their 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 creator's control while freedom of interaction is placed in the viewers' hands.

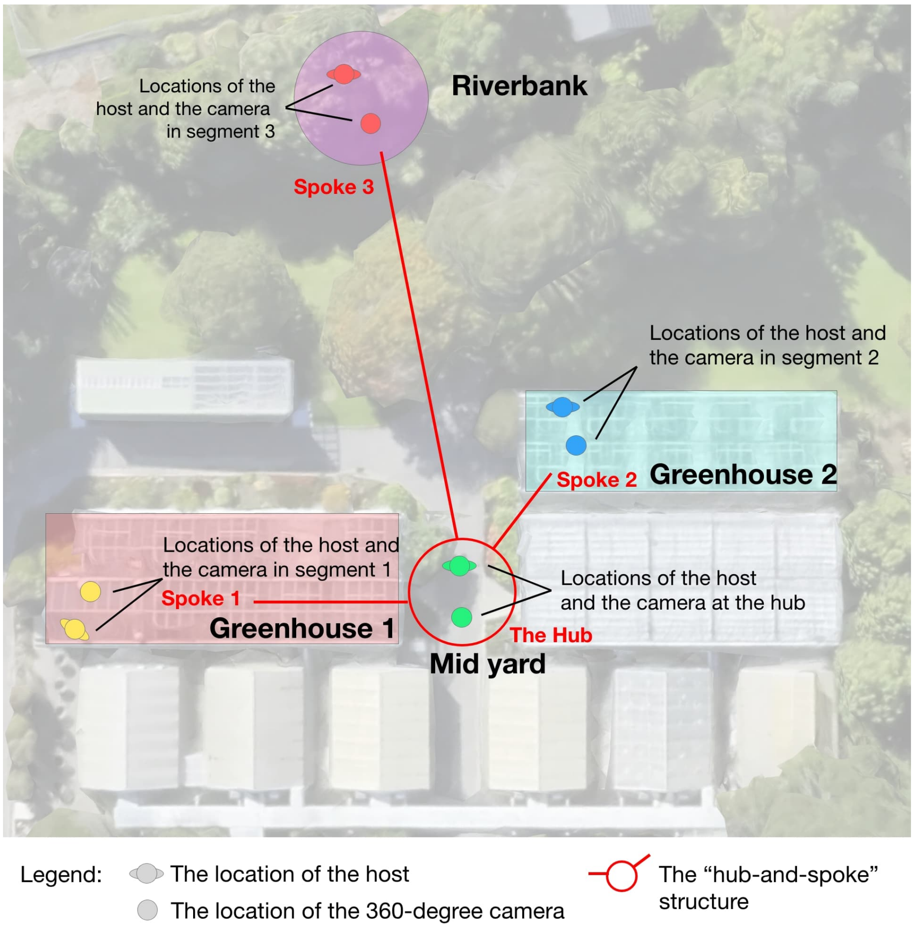


FIGURE 7.1: The layout illustration with a top-down view of the greenhouse where the prototype content was captured, overlaid with a satellite image from Google Maps and a hub-and-spoke structure (the red circle and lines). Each spoke is shown in a different color. The placement of the camera and the host when capturing each spoke segment are also marked. Because of the mixture of indoor and outdoor scenarios and occlusion from buildings, the spokes are not visible to each other. The viewer can only see items inside/around a spot and the exit back to the mid yard.

My first trial was to apply the APC framework in a Māori storytelling project, as discussed in Chapter 6. In that project the second part of the APC, the *viewer interaction design guideline*, was modified to adapt to Tikanga Māori, shifting from looking at specific interaction methods viewers will use to influence the narrative to seeing viewers' traits and features as parameters affecting the disclosure of the story (known as the personalized storytelling). However, due to limited time and resources, only the appropriateness of the personalized storytelling implemented in CVR was evaluated, with six validated personas. The project did not go so far as to produce a robust system to dynamically collect information from individual visitors and match them to a persona. Six personas were also far from exhaustive in terms of the possible types of visitors to a Māori marae. From the perspective of *a possible personalized storytelling experience*, the system was viable to verify the concept of *personalization* in CVR but not yet to act as a fully functional prototype for end viewers to immerse themselves in.

It is undeniable that different storytelling methods exist for different user scenarios besides personalized CVR. There are creators working in areas outside the cultural heritage or GLAM sectors who are beginning to work with newer CVR content using different styles of 360-degree video or other media, as mentioned above. There are also various types of immersive media forms for CVR other than prerecorded 360-degree video, such as computer-generated 3D scenes and captured voxel videos, both of which support 6DOF for viewers to walk freely within the virtual scene. Although the CVR content produced and used in this thesis is limited to the guided-tour style (in which the host usually appears in the scene and introduces the content orally), the applicability of the derived APC framework to other genres of CVR or even 6DOF experiences is still worth investigating.

In terms of the CVR genre, other types can be assumed by looking at the viewer's role in the scene besides the guided-tour style used in this thesis. If the viewer is seen as a member of a group of characters in the scene and is recognized by others, the experience he is having belongs to a *group* type. If the viewer is in the scene but not considered a character, then the experience is categorized as a *bystander* type (often seen in concerts, performances, and documentaries). There is also another scenario where there is no character in the scene, with only voice-over applied, namely a *blank* type. To determine whether the APC is applicable to any of these genres, each of the two parts of the framework should be examined. The APC's immersive production guidelines clearly apply to the group and bystander types as long as there are human characters in the scene. Attention guidance (i.e., AUs) applies regardless of whether or not the viewer is acknowledged as a part of the scene. The AUs can directly guide the viewer or guide other characters in the scene and then let the viewer follow them. However, they will not apply to the blank genre since no human or human-like character is present in the scene. The viewer interaction design guideline (part 2 of the APC framework), on the other hand, is applicable to a scenic tour, a documentary, or a dramatic story, as long as the plot has a narrative element (which is also part of the definition of CVR throughout the thesis) and the story structure is designed to allow various non-linear disclosure. It will fail to be effective if the creator specifically designs the plot to be strictly linear and does not allow any alteration.

Moving away from content type, the discussion around the application of the APC to an immersive experience with 6DOF interactivity is closely linked to the evolving definition of a CVR experience, discussed in Section 2.2.2. First it is clear that the attention guidance rules apply to a 6DOF CVR experience since regardless of how the viewer will interact with the virtual scene, the issues of NP and FOMO remain (possibly worsening due to the increment in interactivity). Secondly, the situation around the viewer interaction design guideline gets complicated with a 6DOF CVR experience. The applicability of the APC will closely depend on what extra interactions the viewer is given access to in the CVR experience. It is assumed that the implicit interaction design can still be applied if the additional interactivity is limited to the ability to move around. The creator, however, now needs to consider the data to be extracted from the viewer. On top of viewpoint, the viewer's location, speed of movement, body orientation, etc. are now all in play since the experience moves to 6DOF. If the viewer is given yet another level of freedom of interactivity and can now interact with detailed objects or elements in the scene, the creator will need to further consider how such interactivity can be coupled with with narrative progression and alteration and how these can be embedded into the context of the story. These newly raised questions are yet to be explored.

An example of a 6DOF CVR project is the Ātea Presence experience, relevant to the Māori storytelling case study presented above (see [109], also mentioned in Section 2.3.5). In this free-roaming experience, a voxelized avatar of a real person from the marae in the physical world acts as the host in a computer-generated environment of the marae. The host delivers narratives while the 6DOF viewer is freely roaming in the scene. Instead of capturing the marae using 360-degree video, three-dimensional clips of a real storyteller (voxelvideos) are rendered (visually and acoustically) in a virtual marae environment. The virtual environment is photogrammetry synthesized. The virtual storyteller introduces artifacts and decorations in the marae, using natural gestures, as shown in Figure 7.2. The voxel video playback can be stopped and started through explicit viewer interaction, similar to the UI controls described in Section 2.4.1. In addition, the voxelvideo host avatar is capable of actively establishing eye contact with the viewer and initiating the introduction when within a certain proximity, maintaining eye contact when speaking to the viewer. When the viewer steps away from the gaze-maintaining proximity, the host avatar returns to the initial pose (of not attending the viewer) and pauses the speech.

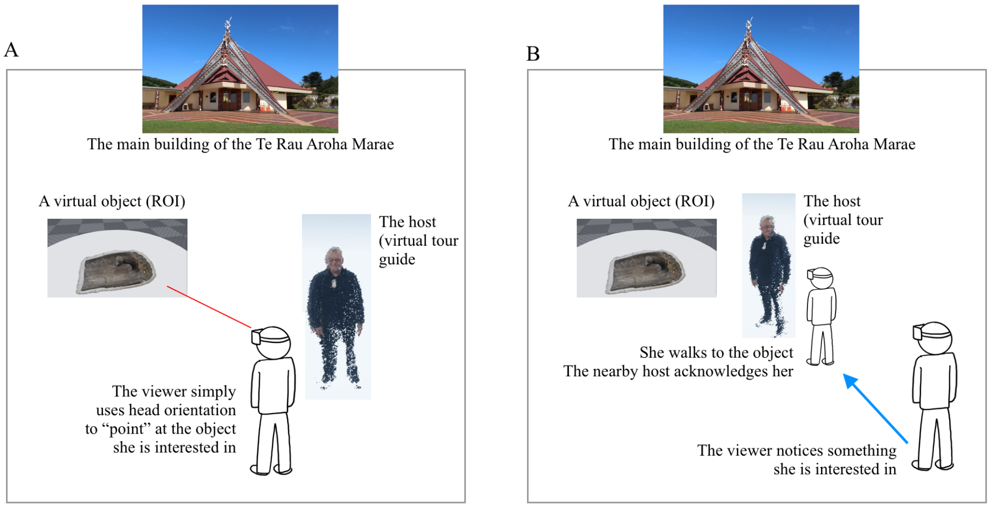


FIGURE 7.2: Two illustrations of Te Rau Aroha Marae's digital marae experience (source: Park et al. [109]). In **setup A** on the left, the viewer uses head orientation to "point" to one of the objects (ROIs) she is interested in, in this case, a tatā or tīrehu (bailer). The virtual host (voxelized avatar) stands beside the viewer to deliver the narrative. In **setup B** on the right, when the viewer notices an object she is interested in, she moves (or teleports) to proximity to that object. The virtual host stands next to the object, acknowledges the viewer, and delivers the narrative.

Although this experience is not a 360-degree-video-based CVR, it still follows some of the APC guidelines, such as acknowledging the viewer/player as a character in the scene and enabling viewer control over the progression of the narrative. The enhanced digital experience provides the viewer with the impression of interacting with a real-person-like host when going through the marae, and that the host is giving a tour especially for her, instead of the impression of a holographic recording that is played a button is pressed.

The digital marae experience shows that the APC framework has the potential to go beyond 360-degree video to also cover other types of CVR content, including ones with computer-generated scenes and 6DOF interaction design. However, as interaction techniques and viewer involvement with the system vary with different types of content and system configurations, more investigation, tweaking, and modification will be required when adapting the APC framework to any specific project. The digital marae experience also inspired the idea of blending CVR content with computer-generated virtual environments (such as a 3D model of a wharenui). On the one hand, this would maintain the easy-to-access characteristic of CVR. On the other hand, the free-roaming feature can be applied for a 6DOF virtual visit. At the time of writing, several projects are being carried out with the collaborators at the Bluff marae, all using APC-guided CVR as part of their foundation. One of these projects is looking at using CVR as a medium to bring a realistic visiting experience to a remote island (Stewart Island/Rakiura; see Section 6.1) to people who live far away and are seeking reconnection with their ancestors. Another one is attempting to use CVR with viewer interactivity to enhance awareness of Māori astronomy for the general public. The third aims to use CVR as a technology to enable Māori wānanga (a traditional meeting or seminar where chosen participants meet to discuss, deliberate over, and consider specific topics) in the digital space. These projects are all aimed at further exploring how the APC framework could be employed as a guideline for broader types of CVR content (as well as hybrid content), various viewer interactivity settings, and even multi-viewer CVR experiences. Unknowns such as the usability of such a digital system, the level of presence and immersion viewers will achieve in such an experience, and how close it will be to a real-world visit are all questions for future exploration.

7.4 Conclusions

The Adaptive Playback Control (APC) framework is a set of design guidelines for cinematic virtual reality (CVR) content creators. It has two main components, the *content production guidelines* and the *viewer interaction design guidelines*. This framework is a response to the research questions proposed in Chapter 1. The process of development and validation of the framework affirmed that when moving from flat screen to immersive media, the viewer must be considered as a character of the story, not as a passive spectator beyond the fourth wall. The case study in Chapter 6 proved that applying the APC framework to CVR production is capable of yielding a CVR experience that is immersive, engaging, and effective in sharing the knowledge stakeholders would like their viewers to receive.

It should be reasserted that the process of formulating the APC framework demonstrates that ensuring viewer participation in the narrative process does not mean giving the viewer full control with the creator receding into the background and losing authority. In CVR, the creator and the viewer are not two conflicting, opposing parties; the creator retains authorial control. However, the power of control must be cleverly used to provide a comfortable and safe space to accommodate viewer curiosity, so that viewers can explore, interact, and alter the outcome of events within the boundary of their comprehension. In the larger picture, the creator can ensure that the overall narrative develops in the intended direction as designed.

The case study in Chapter 6 also reminded me of the importance of understanding the lasting feeling that stakeholders expect a visitor to have after exposure to an immersive digital experience. Insights from this study shed light on the relationship between the digital experience, the user of the digital tool, and the storytellers holding the knowledge and content and considering the passing on of this knowledge to future generations. These considerations all exists on a different plane that goes beyond simply whether the particular digital experience achieved the authenticity one expected. I see my PhD work as part of the first steps in this direction. Many more topics further along the way await exploration.

References

[1] Moneerah Almeshari, John Dowell, and Julianne Nyhan. Museum mobile guide preferences of different visitor personas. *Journal on Computing and Cultural Heritage*, 14(1):9:1–9:13. ISSN 1556-4673. doi: 10.1145/3423186. URL https://doi.org/10.1145/3423186.

[2] Liliana Ardissono, Tsvi Kuflik, and Daniela Petrelli. Personalization in cultural heritage: the road travelled and the one ahead. *User Modeling and User-Adapted Interaction*, 22(1):73–99. ISSN 1573-1391. doi: 10.1007/s11257-011-9104-x. URL https://doi.org/10.1007/s11257-011-9104-x.

[3] Daniel Arijon. *Grammar of the film language*. Focal Press, first edition.

[4] R. Aylett and S. Louchart. Towards a narrative theory of virtual reality. *Virtual Reality*. ISSN 1359-4338. doi: 10.1007/s10055-003-0114-9.

[5] Ruth Aylett. Narrative in virtual environments - towards emergent narrative. In *AAAI fall symposium on narrative intelligence*, pages 83–86. URL http://www.aaai.org/Papers/Symposia/Fall/1999/FS-99-01/FS99-01-014.pdf.

[6] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments*, 10(6):583–598. ISSN 1054-7460. doi: 10.1162/105474601753272844. URL https://www.mitpressjournals.org/doi/abs/10.1162/105474601753272844.

[7] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. Subtle gaze direction. *ACM Transactions on Graphics*. ISSN 0730-0301. doi: 10.1145/1559755.1559757.

[8] Baobab Studios. Invasion! —baobab studios. URL https://www.baobabstudios.com/invasion/.

[9] Oscar Bastiaens. Dyskinetic (2014) - IMDb. URL https://www.imdb.com/title/tt3686216/. Accessed on 09/08/2019.

[10] BBC Worldwide. Bear island. URL https://www.oculus.com/experiences/go/1330308093710823/.

[11] Leslie Bedford. Storytelling: The real work of museums. *Curator: The Museum Journal*, 44(1):27–34. ISSN 2151-6952. doi: 10.1111/j.2151-6952.2001.tb00027.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2151-6952.2001.tb00027.x.

[12] Stuart Bender. Headset attentional synchrony: tracking the gaze of viewers watching narrative virtual reality. *Media Practice and Education*, 20(3):277–296. doi: 10.1080/25741136.2018.1464743.

[13] Chris Bevan, David Philip Green, Harry Farmer, Mandy Rose, Kirsten Cater, Danaë Stanton Fraser, and Helen Brown. 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*, pages 1–12. Association for Computing Machinery. ISBN 978-1-4503-5970-2. URL https://doi.org/10.1145/3290605.3300736.

[14] Judith Binney. Maori oral narratives, pakeha written texts : two forms of telling history. *New Zealand journal of history*, 38(2):203–215. URL https://natlib.govt.nz/records/21094689. Publisher: New Zealand journal of history,.

[15] M. Blythe, J. Reid, P. Wright, and E. Geelhoed. Interdisciplinary criticism: Analysing the experience of riot! a location-sensitive digital narrative. *Behaviour & Information Technology*, 25(2):127–139. ISSN 0144-929X, 1362-3001. doi: 10.1080/01449290500331131. URL http://www.tandfonline.com/doi/abs/10.1080/01449290500331131.

[16] Matt Boggus and Roger Crawfis. Distance field illumination: A rendering method to aid in navigation of virtual environments. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. ISBN 3-642-17273-3. doi: 10.1007/978-3-642-17274-8_49.

[17] David Bordwell and Kristin Thompson. *Film art: an introduction*. McGraw-Hill, 10th edition. ISBN 9780073535104;0073535109;.

[18] Signe Brewster. Designing for real feelings in a virtual reality. URL https://medium.com/s/designing-for-virtual-reality/designing-for-real-feelings-in-a-virtual-reality-41f2a2c7046. Accessed on 09/10/2018.

[19] A. Brown, A. Sheikh, M. Evans, and Z. Watson. Directing attention in 360-degree video. In *IBC 2016 Conference*, pages 29 (9 .)–29 (9 .). ISBN 978-1-78561-343-2. doi: 10.1049/ibc.2016.0029.

[20] Warren Buckland, editor. *Puzzle films: complex storytelling in contemporary cinema*. Wiley-Blackwell, Chichester, West Sussex, U.K. ; Malden, MA, 2009. ISBN 978-1-4051-6861-8 978-1-4051-6862-5. OCLC: ocn230183112.

[21] Matt Burdette. The swayze effect | oculus. URL https://www.oculus.com/story-studio/blog/the-swayze-effect/. Accessed on 09/10/2020.

[22] Stefano Burigat and Luca Chittaro. Navigation in 3d virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human Computer Studies*. doi: 10.1016/j.ijhcs.2007.07.003.

[23] Cameron Grey. Cameron grey - never bout us VR - YouTube. URL https://www.youtube.com/watch?v=_dw-Q6seO0k. Accessed on 05/06/2019.

[24] Elin Carstensdottir, Erica Kleinman, and Magy Seif El-Nasr. Player interaction in narrative games: structure and narrative progression mechanics. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*, FDG '19, pages 1–9. Association for Computing Machinery. ISBN 978-1-4503-7217-6. doi: 10.1145/3337722.3337730. URL https://doi.org/10.1145/3337722.3337730.

[25] Marc Cavazza and Fred Charles. Dialogue generation in character-based interactive storytelling. In *Proceedings of the First AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, AIIDE'05, pages 21–26. AAAI Press.

[26] Marc Cavazza, Jean-Luc Lugrin, David Pizzi, and Fred Charles. Madame bovary on the holodeck: immersive interactive storytelling. In *Proceedings of the 15th international conference on Multimedia - MULTIMEDIA '07*, page 651. ACM Press. ISBN 978-1-59593-702-5. doi: 10.1145/1291233.1291387. URL http://portal.acm.org/citation.cfm?doid=1291233.1291387.

[27] Chingching Chang. Repetition variation strategies for narrative advertising. *Journal of Advertising*, 38(3):51–66, . ISSN 0091-3367. doi: 10.2753/JOA0091-3367380304. URL https://doi.org/10.2753/JOA0091-3367380304. Publisher: Routledge _eprint: https://doi.org/10.2753/JOA0091-3367380304.

[28] Wooksang Chang. Virtual reality filmmaking methodology (animation producing). *TECHART: Journal of Arts and Imaging Science 3*, 3(3), .

[29] Jaehee Cho, Tsung-Han Lee, Joel Ogden, Amy Stewart, Tsung-Yu Tsai, Junwen Chen, and Ralph Vituccio. Imago: presence and emotion in virtual reality. *ACM SIGGRAPH 2016 VR Village*, pages 1–2. ISSN 9781450343770;1450343775;. doi: 10.1145/2929490.2931000. Publisher: ACM.

[30] Barbara Clucas, John M. Marzluff, David Mackovjak, and Ila Palmquist. Do american crows pay attention to human gaze and facial expressions? *Ethology*. ISSN 0179-1613. doi: 10.1111/eth.12064.

[31] Robert Cohen and John Von Szeliski. *Theatre: Brief Edition*. Mayfield Publishing Company. ISBN 0-87484-585-8.

[32] Google Arts Culture. Bagan.

[33] Larry Cutler, Eric Darnell, Nathaniel Dirksen, Michael Hutchinson, Scott Peterson, Robert Schiewe, and Wei Wang. Making you matter: creating interactive VR narratives through experimentation and learning. In *Proceedings of the 2019 Digital Production Symposium*, DigiPro '19, pages 1–9. Association for Computing Machinery. ISBN 978-1-4503-6799-8. doi: 10.1145/3329715.3338883. URL https://doi.org/10.1145/3329715.3338883.

[34] James E. Cutting. Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin and Review*. ISSN 9781405163347. doi: 10.3758/s13423-016-1051-4.

[35] Ahmed Dahroug, Andreas Vlachidis, Antonios Liapis, Antonis Bikakis, Martín López-Nores, Owen Sacco, and José Juan Pazos-Arias. Using dates as contextual information for personalised cultural heritage experiences. *Journal of Information Science*, 47(1):82–100. ISSN 0165-5515. doi: 10.1177/0165551519871823. URL https://doi.org/10.1177/0165551519871823. Publisher: SAGE Publications Ltd.

[36] Edirlei Soares de Lima, Bruno Feijó, Simone D.J. Barbosa, Antonio L. Furtado, Angelo E.M. Ciarlini, and Cesar T. Pozzer. Draw your own story: Paper and pencil interactive storytelling. *Entertainment Computing*, 5(1):33–41. ISSN 18759521. doi: 10.1016/j.entcom.2013.06.004. URL https://linkinghub.elsevier.com/retrieve/pii/S1875952113000104.

[37] Jan Decock, Jan Van Looy, Lizzy Bleumers, and Philippe Bekaert. The pleasure of being (there?): an explorative study into the effects of presence and identification on the enjoyment of an interactive theatrical performance using omnidirectional video. *AI & SOCIETY*, 29(4):449–459. ISSN 0951-5666. doi: 10.1007/s00146-013-0487-6. Place: London Publisher: Springer London.

[38] Kath Dooley. A question of proximity: exploring a new screen grammar for 360-degree cinematic virtual reality. *Media Practice and Education*, 21, 2020 (2):81–96, . doi: 10.1080/25741136.1641005.

[39] Kath Dooley. Storytelling with virtual reality in 360-degrees: a new screen grammar. *Studies in Australasian Cinema*, . doi: 10.1080/17503175.2017.1387357.

[40] Bonnie Jean Eckard. The viewpoints book: A practical guide to viewpoints and composition (review). *Theatre Topics*, 16(2):194–195. ISSN 1054-8378;1086-3346;. doi: 10.1353/tt.2006.0017. Place: Baltimore Publisher: The Johns Hopkins University Press.

[41] Kieran Egan. *Teaching as Story Telling: An Alternative Approach to Teaching and Curriculum in the Elementary School*. University of Chicago Press. ISBN 978-0-226-19032-7. Google-Books-ID: 5R2Ubo_guUcC.

[42] Magy Seif El-Nasr, Athanasios Vasilakos, Chinmay Rao, and Joseph Zupko. Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes. *IEEE Transactions on Computational Intelligence and AI in Games*. doi: 10.1109/TCIAIG.2009.2024532.

[43] Ahmed Elmezeny, Nina Edenhofer, and Jeffrey Wimmer. Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion. *Journal For Virtual Worlds Research*, 11(1). doi: 10.4101/jvwr.v11i1.7298.

[44] Barrett Ens, Joel Lanir, Anthony Tang, Scott Bateman, Gun Lee, Thammathip Piumsomboon, and Mark Billinghurst. Revisiting collaboration through

mixed reality: The evolution of groupware. *International Journal of Human-Computer Studies*, 131:81–98. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2019.05.011. URL https://www.sciencedirect.com/science/article/pii/S1071581919300606.

[45] Colm O. Fearghail, Cagri Ozcinar, Sebastian Knorr, and Aljosa Smolic. Director's cut-analysis of VR film cuts for interactive storytelling. In *2018 International Conference on 3D Immersion, IC3D 2018 - Proceedings*. ISBN 978-1-5386-7590-8. doi: 10.1109/IC3D.2018.8657901.

[46] Felix and Paul Studios. MIYUBI on gear VR. URL https://www.oculus.com/experiences/gear-vr/1307176355972455/. Accessed on 24/10/2018.

[47] Chris Ferguson, Egon L. van den Broek, and Herre van Oostendorp. On the role of interaction mode and story structure in virtual reality serious games. *Computers and Education*, 143:103671–103671. doi: 10.1016/j.compedu.2019.103671. URL https://doi.org/10.1016/j.compedu.2019.103671. Publisher: Elsevier.

[48] Catherine S. Fichten, Vicki Tagalakis, Darlene Judd, John Wright, and Rhonda Amsel. Verbal and nonverbal communication cues in daily conversations and dating. *Journal of Social Psychology*, 132(6):751–769. ISSN 00224545. doi: 10.1080/00224545.1992.9712105.

[49] Rui Figueiredo and Ana Paiva. "i want to slay that dragon!" - influencing choice in interactive storytelling. In Ruth Aylett, Mei Yii Lim, Sandy Louchart, Paolo Petta, and Mark Riedl, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 26–37, Berlin, Heidelberg, 2010. Springer. ISBN 978-3-642-16638-9. doi: 10.1007/978-3-642-16638-9_6.

[50] P Flanagan, K I McAnally, R L Martin, J W Meehan, and S R Oldfield. Aurally and visually guided visual search in a virtual environment. *Human Factors*.

[51] Foundry. Meet the creators: content creation in the virtual age| foundry. URL https://www.foundry.com/insights/vr-ar-mr/virtual-age. Accessed on 09/12/2021.

[52] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '01*. ISBN 1-58113-327-8. doi: 10.1145/365024.365121.

[53] Stephen Gilroy, Julie Porteous, Fred Charles, and Marc Cavazza. Exploring passive user interaction for adaptive narratives. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*, IUI '12, pages 119–128. Association for Computing Machinery. ISBN 978-1-4503-1048-2. doi: 10.1145/2166966.2166990. URL http://doi.org/10.1145/2166966.2166990.

[54] Neil Gramopadhye, Kapil Chalil Madathil, Jeffery Bertrand, and Christopher Shannon Blair. An investigation of the effectiveness of navigational aids in a virtual environment. In *Proceedings of the Human Factors and Ergonomics Society*. ISBN 978-0-945289-45-6. doi: 10.1177/1541931214581407.

[55] Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. SwiVRChair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ISSN 9781450333627. doi: 10.1145/2858036.2858040.

[56] Anton Gustafsson, Cecilia Katzeff, and Magnus Bang. Evaluation of a pervasive game for domestic energy engagement among teenagers. *Computers in Entertainment*, 7(4):54:1–54:19. doi: 10.1145/1658866.1658873. URL https://doi.org/10.1145/1658866.1658873.

[57] Michael Gödde, Frank Gabler, Dirk Siegmund, and Andreas Braun. Cinematic narration in VR –rethinking film conventions for 360 degrees. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. ISSN 9783319915838. doi: 10.1007/978-3-319-91584-5_15.

[58] Jacob Habgood, David Moore, Sergio Alapont, Chris Ferguson, and Herre van Oostendorp. The REVEAL educational environmental narrative framework for PlayStation VR. pages 175–183.

[59] Robert Hassan. Digitality, virtual reality and the 'empathy machine'. *Digital Journalism*, 8(2):195–212. ISSN 2167-0811. doi: 10.1080/21670811.2018.1517604. URL https://doi.org/10.1080/21670811.2018.1517604. Publisher: Routledge _eprint: https://doi.org/10.1080/21670811.2018.1517604.

[60] J. Ibanez, R. Aylett, and R. Ruiz-Rodarte. Storytelling in virtual environments from a virtual guide perspective. *Virtual Reality*. doi: 10.1007/s10055-003-0112-y.

[61] Yannis Ioannidis, Katerina El Raheb, Eleni Toli, Akrivi Katifori, Maria Boile, and Margaretha Mazura. One object many stories: Introducing ICT in museums and collections through digital storytelling. In *2013 Digital Heritage International Congress (DigitalHeritage)*, volume 1, pages 421–424. doi: 10.1109/DigitalHeritage.2013.6743772.

[62] Horace Ho Shing Ip, Chen Li, Selena Leoni, Yangbin Chen, Ka-Fai Ma, Calvin Hoi-to Wong, and Qing Li. Design and evaluate immersive learning experience for massive open online courses (MOOCs). *IEEE Transactions on Learning Technologies*, 12(4):503–515. ISSN 1939-1382. doi: 10.1109/TLT.2018.2878700. Conference Name: IEEE Transactions on Learning Technologies.

[63] Pekka Kallioniemi, Tuuli Keskinen, Ville Mäkelä, Jussi Karhu, Kimmo Ronkainen, Arttu Nevalainen, Jaakko Hakulinen, and Markku Turunen. Hotspot interaction in omnidirectional videos using head-mounted displays. In *Proceedings of the 22nd International Academic Mindtrek Conference*, pages 126–134. ACM. ISBN 978-1-4503-6589-5. doi: 10.1145/3275116.3275148. URL https://dl.acm.org/doi/10.1145/3275116.3275148.

[64] Hermann Kappelhoff and Cornelia Müller. Embodied meaning construction: Multimodal metaphor and expressive movement in speech, gesture, and feature film. *Metaphor and the Social World*, 1(2):121–153. ISSN 9788578110796. doi: 10.1017/CBO9781107415324.004.

[65] Akrivi Katifori, Manos Karvounis, Vassilis Kourtis, Marialena Kyriakidi, Maria Roussou, Manolis Tsangaris, Maria Vayanou, Yannis Ioannidis, Olivier Balet, Thibaut Prados, Jens Keil, Timo Engelke, and Laia Pujol. CHESS: Personalized storytelling experiences in museums. In Alex Mitchell, Clara Fernández-Vara, and David Thue, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 232–235. Springer International Publishing. ISBN 978-3-319-12337-0. doi: 10.1007/978-3-319-12337-0_28.

[66] Peter De Keijzer. Effectively browsing 360-degree videos. *Master's Thesis, Universiteit Utrecht*.

[67] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*. ISSN 0001-6918. doi: 10.1016/0001-6918(67)90005-4.

[68] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*. doi: 10.1207/s15327108ijap0303_3.

[69] Tuuli Keskinen, Ville Makela, Pekka Kallioniemi, Jaakko Hakulinen, Jussi Karhu, Kimmo Ronkainen, John Makela, and Markku Turunen. The effect of camera height, actor behavior, and viewer position on the user experience of 360° videos. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 423–430. IEEE. ISBN 978-1-72811-377-7. doi: 10.1109/VR.2019.8797843. URL https://ieeexplore.ieee.org/document/8797843/.

[70] Hyejin Kim, Jacob Young, Daniel Medeiros, Stephen Thompson, and Taehyun Rhee. TeleGate: Immersive Multi-User Collaboration for Mixed Reality 360°Video. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 532–533, March 2021. doi: 10.1109/VRW52623.2021.00148.

[71] Tina Kjær, Christoffer B. Lillelund, Mie Moth-Poulsen, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. Can you cut it? an exploration of the effects of editing in cinematic virtual reality. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. ISBN 978-1-4503-5548-3. doi: 10.1145/3139131.3139166.

[72] Mark L Knapp and Judith A Hall. *Nonverbal Communication in Human Interaction*. ISBN 978-0-495-56869-8. Publication Title: Dress as NonVerbal Communication.

[73] Hartmut Koenitz. *Towards a Specific Theory of Interactive Digital Narrative*. Routledge. ISBN 978-1-315-76918-9. doi: 10.4324/9781315769189-8. Pages: 91-105 Publication Title: Interactive Digital Narrative.

[74] Willy Christian Kriz. A Systemic-Constructivist Approach to the Facilitation and Debriefing of Simulations and Games. *Simulation & Gaming*, 41(5):663–680, October 2010. ISSN 1046-8781. doi: 10.1177/1046878108319867. URL https://doi.org/10.1177/1046878108319867. Publisher: SAGE Publications Inc.

[75] Kalata Kurt. A japanese rpg primer: The essential 20, March 2008. URL https://www.gamedeveloper.com/design/a-japanese-rpg-primer-the-essential-20. Accessed on 17/08/2022.

[76] Josiah Lebowitz and Chris Klug. *Interactive storytelling for video games: a player-centered approach to creating memorable characters and stories*. Focal Press, Burlington, MA, 2011. ISBN 978-0-240-81717-0.

[77] Chen Li, Horace H. S. Ip, Yuen M. Wong, and Wing S. Lam. An empirical study on using virtual reality for enhancing the youth's intercultural sensitivity in hong kong. *Journal of Computer Assisted Learning*, 36(5):625–635. ISSN 1365-2729. doi: 10.1111/jcal.12432. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jcal.12432. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jcal.12432.

[78] J.J.-W. Lin, H.B.L. Duh, D.E. Parker, H. Abi-Rached, and T.A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. *Proceedings of the IEEE Virtual Reality 2002*, pages 164–171, . doi: 10.1109/vr.2002.996519.

[79] Justin Lin, Lauren Compton, and Sung Kang. Help, . IMDb ID: tt4794550 event-location: United States.

[80] Yen-chen Lin, Yung-ju Chang, Hou-ning Hu, Hsien-tzu Cheng, Chi-wen Huang, and Min Sun. Tell me where to look: Investigating ways for assisting focus in 360° video. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, . ISSN 9781450346559. doi: 10.1145/3025453.3025757.

[81] Robert W Lindeman. Tell me about antarctica: Guidelines for in situ capture and viewing of 360-degree video highlighting antarctic science. *Unpublished final research project report for the Postgraduate Certificate in Antarctic Studies (2017/18). Christchurch, New Zealand: University of Canterbury*, 20. URL https://ir.canterbury.ac.nz/handle/10092/15836.

[82] Logan Dwight. These VR film tips show how to direct audience attention. URL https://uploadvr.com/vr-film-tips-guiding-attention/.

[83] Robert F. Lorch, Elizabeth P. Lorch, and W. Elliot Inman. Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, 85(2):281–290. ISSN 1939-2176(Electronic),0022-0663(Print). doi: 10.1037/0022-0663.85.2.281. Place: US Publisher: American Psychological Association.

[84] Stephanie M. Lukin and Marilyn A. Walker. Narrative variations in a virtual storyteller. In Willem-Paul Brinkman, Joost Broekens, and Dirk Heylen,

editors, *Intelligent Virtual Agents*, Lecture Notes in Computer Science, pages 320–331. Springer International Publishing. ISBN 978-3-319-21996-7. doi: 10.1007/978-3-319-21996-7_34.

[85] Patricia Bianca Lyk, Gunver Majgaard, Lotte Vallentin-Holbech, Julie Dalgaard Guldager, Timo Dietrich, Sharyn Rundle-Thiele, and Christiane Stock. Co-designing and learning in virtual reality: Development of tool for alcohol resistance training. *Electronic Journal of e-Learning*, 18(3):219–234. ISSN 1479-4403. URL https://eric.ed.gov/?id=EJ1267513. Publisher: Academic Conferences Limited.

[86] Andrew MaCquarrie and Anthony Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *Proceedings - IEEE Virtual Reality*, pages 45–54. ISBN 978-1-5090-6647-6. doi: 10.1109/VR.2017.7892230.

[87] Jamie Madigan. The psychology of immersion in video games, 2010. URL https://www.psychologyofgames.com/2010/07/the-psychology-of-immersion-in-video-games/. Accessed on 17/08/2022.

[88] Angel Manuel Soto. Dinner party. URL https://www.with.in/watch/dinner-party. Accessed on 01/03/2021.

[89] Maria, Vayanou, Manos, Karvounis, Marialena, Kyriakidi, Akrivi, Katifori, Natalia, Manola, Roussou, Yannis, and Ioannidis. Towards personalized storytelling for museum visits. In *6th Intl Workshop on Personalized Access, Profile Management, and Context Awareness in Databases-PersDB*.

[90] Marie-Laure Ryan. From narrative games to playable stories: Toward a poetics of interactive narrative. *StoryWorlds: A Journal of Narrative Studies*, 1(1): 43–59. ISSN 2156-7204. doi: 10.1353/stw.0.0003. URL https://muse.jhu.edu/content/crossref/journals/storyworlds/v001/1.ryan.html.

[91] Mark T. Marshall, Daniela Petrelli, Nick Dulake, Elena Not, Michele Marchesoni, Elisa Trenti, and Anna Pisetti. Audio-based narratives for the trenches of world war i: Intertwining stories, places and interaction for an evocative experience. *International Journal of Human-Computer Studies*, 85:27–39. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2015.08.001. URL https://www.sciencedirect.com/science/article/pii/S1071581915001251.

[92] Michael Mateas and Andrew Stern. Towards integrating plot and character for interactive drama. *Socially Intelligent Agents*, pages 221–228. doi: 10.1007/0-306-47373-9_27.

[93] John Mateer. Directing for cinematic virtual reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice*. doi: 10.1080/14682753.2017.1305838.

[94] Celso De Melo and Ana Paiva. Environment expression: Expressing emotions through cameras, lights and music. *Affective Computing and Intelligent Interaction*. ISSN 3540305114. doi: 10.1007/11590361_15.

[95] David E. Millard, Charlie Hargood, Michael O. Jewell, and Mark J. Weal. Canyons, deltas and plains: towards a unified sculptural model of location-based hypertext. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media - HT '13*, pages 109–118. ACM Press. ISBN 978-1-4503-1967-6. doi: 10.1145/2481492.2481504. URL http://dl.acm.org/citation.cfm?doid=2481492.2481504.

[96] Nick Montfort. *Twisty little passages: an approach to interactive fiction*. MIT Press. ISBN 978-0-262-63318-5. OCLC: 799016115.

[97] Paul Mulholland, Annika Wolff, Zdenek Zdrahal, Ning Li, and Joseph Corneli. Constructing and connecting storylines to tell museum stories. In Hartmut Koenitz, Tonguc Ibrahim Sezen, Gabriele Ferri, Mads Haahr, Digdem Sezen, and Güven Çatak, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 121–124. Springer International Publishing. ISBN 978-3-319-02756-2. doi: 10.1007/978-3-319-02756-2_14.

[98] Janet H. Murray. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. 2nd, updated edition. ISBN 978-0-262-53348-5.

[99] Lasse T Nielsen, Matias B. Moller, Sune D Hartmeyer, Troels C M Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*. ISSN 9781450344913. doi: 10.1145/2993369.2993405.

[100] Niels Chr Nilsson, Rolf Nordahl, and Stefania Serafin. Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, 12(2):108–134. ISSN 1795-6889. Publisher: University of Jyväskylä, Agora Center.

[101] Valentina Nisi, Alison Wood, Glorianna Davenport, and Ian Oakley. Hopstory: An interactive, location-based narrative distributed in space and time. In Stefan Göbel, Ulrike Spierling, Anja Hoffmann, Ido Iurgel, Oliver Schneider, Johanna Dechau, and Axel Feix, editors, *Technologies for Interactive Digital Storytelling and Entertainment*, Lecture Notes in Computer Science, pages 132–141. Springer. ISBN 978-3-540-27797-2. doi: 10.1007/978-3-540-27797-2_18.

[102] Elena Not and Daniela Petrelli. Blending customisation, context-awareness and adaptivity for personalised tangible interaction in cultural heritage. *International Journal of Human-Computer Studies*, 114:3–19, . ISSN 1071-5819. doi: 10.1016/j.ijhcs.2018.01.001. URL https://www.sciencedirect.com/science/article/pii/S1071581918300016.

[103] Elena Not and Daniela Petrelli. Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences. *User Modeling and User-Adapted Interaction*, 29(1):67–120, . ISSN 1573-1391. doi: 10.1007/s11257-019-09224-9. URL https://doi.org/10.1007/s11257-019-09224-9.

[104] Acushla Deanne O'carroll. Virtual whanaungatanga: Māori utilizing social networking sites to attain and maintain relationships. *AlterNative: An International Journal of Indigenous Peoples*, 9(3):230–245. ISSN 1177-1801. doi: 10.1177/117718011300900304. URL https://doi.org/10.1177/117718011300900304. Publisher: SAGE Publications Ltd.

[105] Oculus Story Studio. Henry | oculus, . URL https://www.oculus.com/experiences/rift/1043217179077622/.

[106] Oculus Story Studio. Lost | oculus, . URL https://www.oculus.com/experiences/rift/1016967501697367/.

[107] Heather L. O' Brien, Paul Cairns, and Mark Hall. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112:28–39. ISSN 1071-5819. doi: 10.1016/j.ijhcs.2018.01.004. URL https://www.sciencedirect.com/science/article/pii/S1071581918300041.

[108] Toni Pakkanen, Jaakko Hakulinen, Tero Jokela, Ismo Rakkolainen, Jari Kangas, Petri Piippo, Roope Raisamo, and Marja Salmimaa. Interaction with WebVR 360° video player: Comparing three interaction paradigms. In *Proceedings - IEEE Virtual Reality*, pages 279–280. ISBN 978-1-5090-6647-6. doi: 10.1109/VR.2017.7892285.

[109] J. N. Park, S. Mills, H. Whaanga, P. Mato, R. W. Lindeman, and H. Regenbrecht. Towards a māori telepresence system. In *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 1–6. doi: 10.1109/IVCNZ48456.2019.8961016. ISSN: 2151-2205.

[110] Tabitha C. Peck, Henry Fuchs, and Mary C. Whitton. Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Transactions on Visualization and Computer Graphics*. ISSN 9781424419715. doi: 10.1109/TVCG.2008.191.

[111] Sara Perry, Maria Roussou, Maria Economou, Hilary Young, and Laia Pujol. Moving beyond the virtual museum: Engaging visitors emotionally. In *2017 23rd International Conference on Virtual System Multimedia (VSMM)*, pages 1–8. doi: 10.1109/VSMM.2017.8346276. ISSN: 2474-1485.

[112] Daniela Petrelli and Elena Not. User-centred design of flexible hypermedia for a mobile guide: Reflections on the HyperAudio experience. *User Modeling and User-Adapted Interaction*, 15(3):303–338. ISSN 1573-1391. doi: 10.1007/s11257-005-8816-1. URL https://doi.org/10.1007/s11257-005-8816-1.

[113] Benjamin Petry and Jochen Huber. Towards effective interaction with omnidirectional videos using immersive virtual reality headsets. In *Proceedings of the 6th Augmented Human International Conference*, pages 217–218. ACM. ISBN 978-1-4503-3349-8. doi: 10.1145/2735711.2735785. URL https://dl.acm.org/doi/10.1145/2735711.2735785.

[114] Andrea C. Pierno, Andrea Caria, and Umberto Castiello. Comparing effects of 2-d and 3-d visual cues during aurally aided target acquisition. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(4):728–737. doi: 10.1518/hfes.46.4.728.56815.

[115] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. CoVAR: Mixed-platform remote collaborative augmented and

virtual realities system with shared collaboration cues. In *Adjunct Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2017*, pages 218–219. Institute of Electrical and Electronics Engineers Inc. ISBN 978-0-7695-6327-5. doi: 10.1109/ISMAR-Adjunct.2017.72.

[116] Vanessa C. Pope, Robert Dawes, Florian Schweiger, and Alia Sheikh. The geometry of storytelling: Theatrical use of space for 360-degree videos and virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4468–4478. ACM. ISBN 978-1-4503-4655-9. doi: 10.1145/3025453.3025581. URL https://dl.acm.org/doi/10.1145/3025453.3025581.

[117] Gerald Prince. *A Grammar of Stories: An Introduction*. Walter de Gruyter. ISBN 978-90-279-2535-0. Google-Books-ID: UDY6pRQSFAYC.

[118] Laia Pujol, Maria Roussou, Stavrina Poulou, Olivier Balet, Maria Vayanou, and Yannis Ioannidis. Personalizing interactive digital storytelling in archaeological museums: the CHESS project. *Archaeology in the Digital Era. Papers from the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA)*, pages 77–90. ISSN 978 90 8964 663 7. URL dare.uva.nl/cgi/arno/show.cgi?fid=545855.

[119] Brian Ravenet, Catherine Pelachaud, Chloé Clavel, and Stacy Marsella. Automating the production of communicative gestures in embodied characters. *Frontiers in Psychology*, 9. doi: 10.3389/fpsyg.2018.01144.

[120] Holger Regenbrecht and Thomas Schubert. Real and illusory interactions enhance presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 11(4):425–434. ISSN 1054-7460. doi: 10.1162/105474602760204318. URL https://doi.org/10.1162/105474602760204318.

[121] Mar¡a Cecilia Reyes. Screenwriting framework for an interactive virtual reality film. doi: 10.3217/978-3-85125-530-0-15. URL https://openlib.tugraz.at/download.php?id=5e6b52c99fbfa&location=medra. ISBN: 9783851255300 Publisher: Verlag der Technischen Universität Graz.

[122] Anna Marie Rezk and Mads Haahr. The case for invisibility: Understanding and improving agency in black mirror's bandersnatch and other interactive digital narrative works. In Anne-Gwenn Bosser, David E. Millard, and Charlie

Hargood, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 178–189. Springer International Publishing. ISBN 978-3-030-62516-0. doi: 10.1007/978-3-030-62516-0_16.

[123] Elaine Rich. Users are individuals: individualizing user models. *International Journal of Man-Machine Studies*, 18(3):199–214. ISSN 0020-7373. doi: 10.1016/S0020-7373(83)80007-8. URL https://www.sciencedirect.com/science/article/pii/S0020737383800078.

[124] Clair Rogers. Māori storytelling - the oral tradition. URL http://maoristorytelling.weebly.com/oral-tradition.html. Accessed on 01/09/2021.

[125] Christian Roth and Hartmut Koenitz. Bandersnatch, yea or nay? reception and user experience of an interactive digital narrative video. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '19, pages 247–254. Association for Computing Machinery. ISBN 978-1-4503-6017-3. doi: 10.1145/3317697.3325124. URL https://doi.org/10.1145/3317697.3325124.

[126] Christian Roth, Tom van Nuenen, and Hartmut Koenitz. Ludonarrative hermeneutics: A way out and the narrative paradox. In Rebecca Rouse, Hartmut Koenitz, and Mads Haahr, editors, *Interactive Storytelling*, volume 11318, pages 93–106. Springer International Publishing. ISBN 978-3-030-04027-7 978-3-030-04028-4. doi: 10.1007/978-3-030-04028-4_7. URL http://link.springer.com/10.1007/978-3-030-04028-4_7. Series Title: Lecture Notes in Computer Science.

[127] Sylvia Rothe and Heinrich Hussmann. Spaceline: A concept for interaction in cinematic virtual reality. *International Conference on Interactive Digital Storytelling*, pages 115–119. ISSN 9783030338930. doi: 10.1007/978-3-030-33894-7_12.

[128] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1):19–19, . doi: 10.3390/mti3010019.

[129] Sylvia Rothe, Boris Kegeles, Mathias Allary, and Heinrich Hußmann. The impact of camera height in cinematic virtual reality. In *Proceedings of the*

24th ACM Symposium on Virtual Reality Software and Technology, pages 1–2. ACM, . ISBN 978-1-4503-6086-9. doi: 10.1145/3281505.3283383. URL https://dl.acm.org/doi/10.1145/3281505.3283383.

[130] Sylvia Rothe, Pascal Pothmann, Heiko Drewe, and Heinrich Hussmann. Interaction techniques for cinematic virtual reality. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings*, pages 1733–1737, . ISSN 9781728113777. doi: 10.1109/VR.2019.8798189.

[131] Maria Roussou and Akrivi Katifori. Flow, staging, wayfinding, personalization: Evaluating user experience with mobile museum narratives. *Multimodal Technologies and Interaction*, 2(2):32. doi: 10.3390/mti2020032. URL https://www.mdpi.com/2414-4088/2/2/32. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.

[132] Maria Roussou, Akrivi Katifori, Laia Pujol, Maria Vayanou, and Stefan J. Rennick-Egglestone. A life of their own: museum visitor personas penetrating the design lifecycle of a mobile experience. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 547–552. Association for Computing Machinery. ISBN 978-1-4503-1952-2. doi: 10.1145/2468356.2468453. URL https://doi.org/10.1145/2468356.2468453.

[133] Marie-Laure Ryan. Interactive narrative, plot types, and interpersonal relations. In Ulrike Spierling and Nicolas Szilas, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 6–13. Springer. ISBN 978-3-540-89454-4. doi: 10.1007/978-3-540-89454-4_2.

[134] Anastasia Schmitz, Andrew Macquarrie, Simon Julier, Nicola Binetti, and Anthony Steed. Directing versus attracting attention: Exploring the effectiveness of central and peripheral cues in panoramic videos. *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020*, pages 63–72. ISSN 9781728156088. doi: 10.1109/VR46266.2020.1581102716289.

[135] Magy Seif El-Nasr, David Milam, and Tony Maygoli. Experiencing interactive narrative: A qualitative analysis of façade. *Entertainment Computing*, 4(1): 39–52. ISSN 18759521. doi: 10.1016/j.entcom.2012.09.004. URL https://linkinghub.elsevier.com/retrieve/pii/S187595211200016X.

[136] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. Movie editing and cognitive event segmentation

in virtual reality video. *ACM Transactions on Graphics 36, 4, Article 47 (July 2017)*, 36(4). doi: 10.1145/3072959.3073668. URL http://arxiv.org/abs/1806.04924%0Ahttp://dx.doi.org/10.1145/3072959.3073668.

[137] Syed Hammad Hussain Shah, Kyungjin Han, and Jong Weon Lee. Real-time application for generating multiple experiences from 360° panoramic video by tracking arbitrary objects and viewer ' s orientations. *Applied Sciences*, 10(7):2248. doi: 10.3390/app10072248. URL https://www.mdpi.com/2076-3417/10/7/2248. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.

[138] Islam Sharaha and Amal AL Dweik. Digital interactive storytelling approaches : A systematic review. In *Computer Science & Information Technology (CS & IT)*, pages 21–30. Academy & Industry Research Collaboration Center (AIRCC). ISBN 978-1-921987-55-7. doi: 10.5121/csit.2016.61002. URL http://airccj.org/CSCP/vol6/csit65702.pdf.

[139] Aman Sium and Eric Ritskes. Speaking truth to power: Indigenous storytelling as an act of living resistance. *Decolonization: Indigeneity, Education & Society*, 2(1). ISSN 1929-8692. URL https://jps.library.utoronto.ca/index.php/des/article/view/19626. Number: 1.

[140] David Slade, Fionn Whitehead, Craig Parkinson, and Alice Lowe. Black mirror: Bandersnatch. URL https://www.imdb.com/title/tt9495224/. Accessed on 01/12/2019.

[141] Thomas P B Smith, Marianna Obrist, and Peter Wright. Live-streaming changes the (video) game. pages 131–138. ISSN 9781450319515.

[142] Callum Spawforth, Nicholas Gibbins, and David E. Millard. StoryMINE: A System for Multiplayer Interactive Narrative Experiences. In Rebecca Rouse, Hartmut Koenitz, and Mads Haahr, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 534–543, Cham, 2018. Springer International Publishing. ISBN 978-3-030-04028-4. doi: 10.1007/978-3-030-04028-4_62.

[143] Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krúger. Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*, pages 1–12. ACM, . ISBN 978-1-4503-6017-3.

doi: 10.1145/3317697.3323350. URL https://dl.acm.org/doi/10.1145/3317697.3323350.

[144] Maximilian Speicher, Jingchen Cao, Ao Yu, Haihua Zhang, and Michael Nebeling. 360anywhere: Mobile ad-hoc collaboration in any environment using 360 video and augmented reality. *Proceedings of the ACM on Human-Computer Interaction*, 2:1–20, . doi: 10.1145/3229091.

[145] Baobab Studios. ASTEROIDS! —baobab studios. URL https://www.baobabstudios.com/asteroids. Accessed on 20/04/2018.

[146] Hannah Syrett, Licia Calvi, and Marnix van Gisbergen. 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, Revised Selected Papers*. ISBN 978-3-319-49616-0. doi: 10.1007/978-3-319-49616-0_19.

[147] Sercan Şengün. Silent hill 2 and the curious case of invisible agency. In Hartmut Koenitz, Tonguc Ibrahim Sezen, Gabriele Ferri, Mads Haahr, Digdem Sezen, and Güven Çatak, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 180–185. Springer International Publishing. ISBN 978-3-319-02756-2. doi: 10.1007/978-3-319-02756-2_22.

[148] Ryohei Tanaka, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. Attracting user's attention in spherical image by angular shift of virtual camera direction. *Proceedings of the 3rd ACM Symposium on spatial user interaction*, pages 61–64. ISSN 9781450337038;1450337031;. doi: 10.1145/2788940.2788951. Publisher: ACM.

[149] L. Te Morenga, C. Pekepo, C. Corrigan, L. Matoe, R. Mules, D. Goodwin, J. Dymus, M. Tunks, J. Grey, G. Humphrey, A. Jull, R. Whittaker, M. Verbiest, R. Firestone, and C. Ni Mhurchu. Co-designing an mHealth tool in the new zealand māori community with a "kaupapa māori" approach. *AlterNative*, 14(1): 90–99. ISSN 1177-1801. doi: 10.1177/1177180117753169.

[150] Theophilus Teo, Gun A. Lee, Mark Billinghurst, and Matt Adcock. Hand gestures and visual annotation in live 360 panorama-based mixed reality remote collaboration. *ACM International Conference Proceeding Series*. ISSN 9781450361880. doi: 10.1145/3292147.3292200.

[151] David R. Thomas. A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation*, 27(2):237–246. ISSN 1098-2140. doi: 10.1177/1098214005283748. URL https://doi.org/10.1177/1098214005283748. Publisher: SAGE Publications Inc.

[152] Michael Tomasello, Malinda Carpenter, and Ulf Liszkowski. A new look at infant pointing. *Child Development*. ISSN 1467-8624. doi: 10.1111/j.1467-8624.2007.01025.x.

[153] Lingwei Tong, Sungchul Jung, Richard Chen Li, Robert W. Lindeman, and Holger Regenbrecht. Action units: Exploring the use of directorial cues for effective storytelling with swivel-chair virtual reality. In *32nd Australian Conference on Human-Computer Interaction*, pages 45–54. ACM, . ISBN 978-1-4503-8975-4. doi: 10.1145/3441000.3441063. URL https://dl.acm.org/doi/10.1145/3441000.3441063.

[154] Lingwei Tong, Robert W. Lindeman, and Holger Regenbrecht. Viewer's role and viewer interaction in cinematic virtual reality. *Computers*, 10 (5):66, . doi: 10.3390/computers10050066. URL https://www.mdpi.com/2073-431X/10/5/66. Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.

[155] Marian F. Ursu, Maureen Thomas, Ian Kegel, Doug Williams, Mika Tuomola, Inger Lindstedt, Terence Wright, Andra Leurdijk, Vilmos Zsombori, Julia Sussner, Ulf Myrestam, and Nina Hall. Interactive TV narratives: Opportunities, progress, and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 4(4):25:1–25:39. ISSN 1551-6857. doi: 10.1145/1412196.1412198. URL https://doi.org/10.1145/1412196.1412198.

[156] K. R. Veale. Interactive cinema is an oxymoron, but may not always be. *Game Studies: the international journal of computer game research*, 12(1), September 2012. ISSN 1604-7982. URL https://mro.massey.ac.nz/handle/10179/10027. Publisher: Game Studies.

[157] Renato Verdugo, Miguel Nussbaum, Pablo Corro, Pablo Nuñnez, and Paula Navarrete. Interactive films and coconstruction. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 7(4):39:1–39:24. ISSN 1551-6857. doi: 10.1145/2043612.2043617. URL http://doi.org/10.1145/2043612.2043617.

[158] Mirjam Vosmeer and Ben Schouten. Interactive cinema: Engagement and interaction. In Alex Mitchell, Clara Fernández-Vara, and David Thue, editors, *Interactive Storytelling*, Lecture Notes in Computer Science, pages 140–147. Springer International Publishing. ISBN 978-3-319-12337-0. doi: 10.1007/978-3-319-12337-0_14.

[159] Shayne Walker, Anaru Eketone, and Anita Gibbs. An exploration of kaupapa maori research, its principles, processes and applications. *International Journal of Social Research Methodology*, 9(4): 331–344. ISSN 1364-5579. doi: 10.1080/13645570600916049. URL https://doi.org/10.1080/13645570600916049. Publisher: Routledge _eprint: https://doi.org/10.1080/13645570600916049.

[160] Huaxin Wei, Jim Bizzocchi, and Tom Calvert. Time and space in digital game storytelling. *International Journal of Computer Games Technology*. doi: 10.1155/2010/897217.

[161] Glenn Joseph Winters and Jichen Zhu. Guiding players through structural composition patterns in 3d adventure games. *Foundations of Digital Games*, pages 1–8. URL http://www.fdg2014.org/papers/fdg2014_paper_35.pdf.

[162] Jonathan Wonner, Jérôme Grosjean, Antonio Capobianco, and Dominique Bechmann. Bubble bee, an alternative to arrow for pointing out directions. doi: 10.1145/2503713.2503753.

[163] Nelson Zagalo, Anthony Barker, and Vasco Branco. Story reaction structures to emotion detection. In *Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, SRMC '04, pages 33–38. Association for Computing Machinery, . ISBN 978-1-58113-931-0. doi: 10.1145/1026633.1026641. URL http://doi.org/10.1145/1026633.1026641.

[164] Nelson Zagalo, Ana Torres, and Vasco Branco. Emotional spectrum developed by virtual storytelling. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 3805 LNCS, pages 105–114, . ISBN 3-540-30511-4. doi: 10.1007/11590361_12.