EVALUATING THE EFFECT OF CINEMATOGRAPHY ON THE VIEWING EXPERIENCE IN IMMERSIVE ENVIRONMENT

Feilin Han[*], *Ying Zhong, Minxi Zhou*

Department of Film and Television Technology, Beijing Film Academy, China

ABSTRACT

Cinematic Virtual Reality (CVR) is an increasingly popular digital art production technology that could enhance the sense of presence when a viewer explores immersive environments. There are three important viewing-experience-related aspects, attention, sustainability, and guidance, which can be affected by the cinematography principles. Attention indicates whether the viewer is focusing on the storytelling-related region or not. Sustainability refers to viewers' ability to continuously watch the CVR content, and guidance affects the understanding of the narrative. In this paper, we conducted within-subject repeated-measures experiments on 22 participants in an HMD-based immersive environment, to explore the correlation between viewing experience and comprehensive factors. According to experimental results, we suggest an attention-comfort-understanding analysis paradigm for directing the CVR shot, which could help creators effectively attract viewers' attention, minimize the cybersickness, and deepen their understanding of narratives.

Index Terms— Cinematic Virtual Reality, Viewing Experience, Cinematography

1. INTRODUCTION

Cinematic virtual reality (CVR) entertainment infers immersive film-like 360° stereo videos. In common, CVR viewing facilities include head-mounted displays (HMD) and swivel chairs, as shown in Figure 1. It allows the viewer to explore the virtual environment freely, whereas, the traditional cinematic frame is predetermined. The narrative in CVR is human-centered, non-linear, pluralistic, and dynamic. Interactively real-time rendered, viewer-controlled, and individual-diverse experiences straddle the boundary between virtual reality and film.

Traditional film production took decades to develop the language of visual storytelling, such as shot scale, camera movement, and montage to express directorial intention. It is obviously not appropriate to apply them directly to CVR production. Therefore, exploring the proper way of CVR content-making industry-standard techniques and terminology is essential. Furthermore, the immersive narrative of CVR is in a first-person perspective way, which is associated with the viewer's sensation of presence, emotional engagement, and understanding of the story [1]. To decrease interfering factors, we discuss how to improve the viewing experience with 3 aspects, viewer's attention region, viewing sustainability, and viewpoint guidance efficiency.

[*]Corresponding author: hanfeilin@bfa.edu.cn

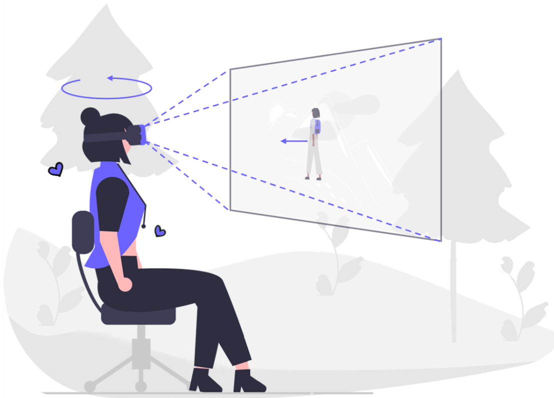


Fig. 1: Concept of CVR viewing facilities and modality.

In this work, we design within-subject repeated-measures experiments in an HMD-based immersive environment, to explore CVR directing principles and implications. Our experiments fill three purposes. One is to explore the scalability and applicability of cinematography principles for CVR production. The second is to suggest a attention-comfort-understanding analysis paradigm for directing the CVR shot, which could help creators effectively attract viewers' attention, minimize the cybersickness, and deepen viewers' understanding of narratives. The last one is to discuss the possibility of CVR as a creative storytelling approach to move the virtual reality artwork production one step forward.

2. RELATED WORK

CVR is widely welcomed by the public with the increasing affordable lightweight VR devices appearing. Engagement with CVR correlates with multiple factors, including the sensation of presence, viewing comfort, and emotional empathy [2]. To explore how the narrative in CVR differs from traditional film storytelling and what matters in improving the viewing experience, in this section, we go through related work in the following research fields.

Field of View (FoV) indicates the range that the human eye can see, which should be overlapped with the storytelling region in the CVR video. The standard FoV for a human is 120°, and when focused, the attentional FoV is about 40° [3]. 360° stereo videos, also known as immersive videos [4], resemble film-like media. Since viewers are free to look around, the FoV changes as the user's interest changes. Special settings in the plot should be designed to cue viewers' attention and attract them to relocate their viewpoint to the narrative Region of Interest (ROI). In immersive film *HELP*, director Lin utilizes directional stereophonic sound and actors' facing direction to guide viewers' viewpoint back to the monster. Moreover, CVR narrative attaches great relation to the amount of ROI , the distance between character and viewer, and dynamism in the scene [5].

Narrative is an storytelling medium for director. In cinematic virtual reality, it is essential to guide attention to follow the narrative. Otherwise, FoV shifting may result in missing key events of the story. Previous work has compared the Mise-en-scéne in CVR with traditional movies, to address the challenge of continuously focusing and re-focusing the target by visual guidance arranged in the scene [6]. Director could also create storyline components, in general, such as the characters in the film's scenes, the movement of objects, and the variation of scene light and shadow, to guarantee the narrative, which could bring a higher level of immersion [7]. Sometimes, CVR directors have to choose non-storyline components to cue attentions, such as arrows signs, and dots of light [8]. The use of non-storyline factors is very effective but could easily disrupt the immersion of the film. Guiding and focusing attention is particularly challenging if ROI is outside of the user's FoV or the video includes multiple dynamic ROI [9].

Viewing Experience in CVR is related to the immersive quality, the sense of presence, and the viewing comfort. The technology developers and media creators are pursuing high quality immersion and presence to effectively convey content [10]. VR sickness, also know as cybersickness, typically results from visual-vestibular conflict. Previous studies focused on finding a positive relationship between FoV and presence [11]. FoV restrictor according to the users eye gaze position could reduce VR sickness [12] . Blocking the perception of peripheral motion by reducing the user's FoV is another effective strategy [13]. Other work indicated that a larger FoV increases presence with additional sickness [14]. As we expected, evaluating the transformation of FoV could provide the measurement basis for quantitative analysis.

3. DESGINING AND IMPLEMENTING AN IMMERSIVE EXPERIMENTAL ENVIRONMENT

3.1. Problem Analysis

There are three important viewing-experience-related aspects, attention, sustainability, and guidance, which can be affected by the cinematography principles. Previous CVR-related research works use subjective study methods in the majority. In this work, we would like to discuss influencing factors in a more quantitative way to find out the paradigm of analyzing the viewing experience in CVR.

Attention indicates whether the viewer is focusing on the storytelling-related region or not. Traditional film frames are presented fixed with predetermined cut-offs, which helps viewers pay attention to subjects and visual elements on the 2D screen plane. However, the virtual world in CVR is informative-rich in 3D space. Catching up viewers' attention is the foundation of building up the story. To evaluate whether the designed ROI successfully attracts the viewer's attention, we could find out if the FoV is located to the ROI and the viewer correctly captures the main subject.

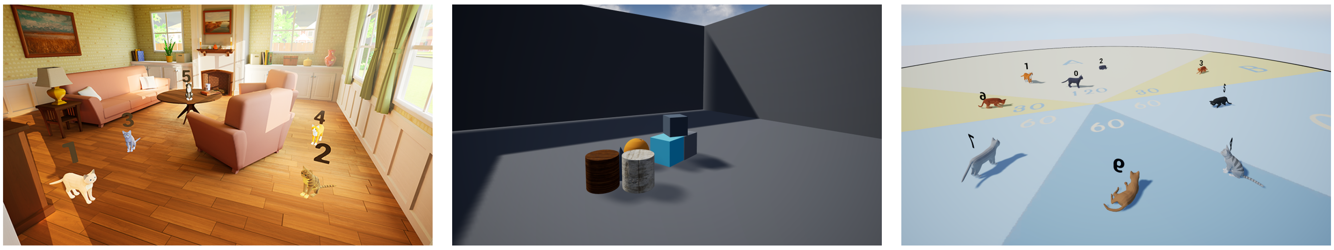
Sustainability refers to the viewer's ability of continuously watching the CVR content. To our knowledge, in traditional cinema, there are lots of factors that affect whether the audience could keep watching the film or leave, such as the story, the viewing condition, and the film duration. In CVR, the camera is controlled by viewers, rather than pre-rendered frames in traditional film. Though former mentioned factors have impacts on viewing experience as well, the inappropriate camera movements could cause physically, not subjectively, uncomfort, which may have a huge impact on the viewing experience. To evaluate this manageable factor, we could analyze viewers' response to variance camera movements, which could help us infer shot designing principles.

The guidance affects the understanding of the narrative. The narrative in immersion is non-linear, highly viewer-controlled, and of diverse viewpoints. Obviously, the traditional narrative techniques will fail if viewers miss the ROI. In CVR, directing viewers facing the storyline contents is crucial to the success of the narrative. For the HMD-based VR viewing condition, we could record the HMD rotating data compared with the subject movements in the virtual environment. The relative variation is a convincing measurement for evaluating the efficiency of guidance intuitively.

3.2. Experiment Design

In this work, we discuss the effect of cinematography on the viewing experience in immersive environment. Three short CVR experiments are designed to evaluate the relation between different cinematography techniques and user experiences, as shown in Figure 2.

Shot Scale Experiment. CVR video titled '*Cats in the Room*' is designed to evaluate the influence of 5 typical shot



(a) *Cats in the Room* (b) *Boxes in the Box* (c) *the Spinning Cat*

Fig. 2: Three short CVR demos were created, named as *Cats in the Room*. *Boxes in the Box*, and *the Spinning Cat*, for proposed experiments, shot scale, camera movements, and full-length shots, from left to right respectively.

scales on attracting users' attention. This experiment compares the effects of different shot scales on the main character's attraction to users. The traditional shot scales are divided by the proportion of the main character in the frame. Whereas, in immersive environments, they are classified by the distance between the camera and the main character. According to this, we created a CVR called '*Cat in the Room*' to test the users' experience of different distances between the camera and the cat. The plot is as follows: there are five cats (numbered from 1 to 5) in the living room with different poses, as shown in Figure 2a.

Camera Movement Experiment. We designed the CVR video titled '*Boxes in the Box*' to compare the effects of different camera movements on participants' comfort. To explore the appropriate way for camera guidance, we created three CVRs with three types of camera movement - dolly, truck, and pedestal. CVRs are designed with the same scene, same objects, and different camera movements. In this experiment, we control the size of objects in the field of view stay on the same scale to ensure the experiment condition is consistent, as shown in Figure 2b. We set the base speed of the camera movement to 1m/s (speed 1), and the rate goes up in multiples, e.g., speed 3 is 3m/s. To avoid the dizziness caused by wearing for a long time, we ask participants to take a rest after each CVR test.

Full-Length Shot Experiment. CVR video titled '*the Spinning Cat*' was produced to analyze how the character's moving intensity affects the efficiency of guiding users' viewpoint. In '*the Spinning Cat*', the main cat signed 0 is designed to guide the users to follow the subject resulting in transferring the scenario and shooting a full-length shot in the CVR. This was used to test the accuracy of different amplitude guidance methods and the users' understanding of the storyline. The plot is as follows: there are eight cats (numbered 1 to 8) in a circle on a plane, each with a different pose and color, and the main cat (numbered 0, cat 0) will move from one area to another, as shown in Figure 2c.

In this CVR, the narrative sphere is divided into four regions A, B, C, and D. Region A is the maximum perspective range of the user without head rotation. Region B is where the user can obtain information with a slight head rotation. Region C allows the user to access the content only when the head turns to the maximum range. In the last region D, users could reach the information only by turning their body. For each CVR, the main cat 0 moves from A to B, A to C, and A to D with different moving intensities, walking, trotting, and jumping. After watching each CVR, participants were asked to recall from memory and answer questions to infer better guidance to improve viewers' understanding of contents.

3.3. Participants and Apparatus

As the audience for VR is mostly aged between 18 and 30, we recruited 23 users in this age group, with 13 females and 10 males. Each user has participated in three experiments, the shot scale experiment, the camera movement experiment, and the full-length shoot experiment. Each experiment is conducted under the same conditions, and we ensure that all viewers have no idea about experiment-related contents before they start the test.

Scenes for three experiments were created using Unreal Engine (4.26.2). We used the HTC Vive Cosmos Elite (HMD screen resolution 2880 x 1770, refresh rate 90Hz, field of view 110 degrees) to display the CVR and obtain the users' head rotation data. The CVRs were performed on Windows 10 OS, i9 CPU desktop computer, with an NVIDIA GeForce RTX 3090 video card.

3.4. Procedure

We designed a questionnaire and 13 questions about the CVRs. Users read questions before the experiments and were asked to answer the questions after viewing the CVRs. The questions include some personal statements like if they have symptoms such as vertigo, periodic migraines, photosensitive epilepsy, or have the experience of watching immersive videos or playing VR games before, which are used as background information for the data analysis.

In ***Shot Scale Experiment***, users will view the CVR five times with a different distance between the camera and the main character, corresponding to full shot, medium-full shot, medium shot, close shot, and close-up in the traditional film. Two questions are desgined for this experiment.

- Question 1: Which cat is the main character?
- Question 2: Which shot scale above would you prefer to show the main character?

In this experiment, the users were asked to answer question 1 after viewing each CVR. We estimated whether that shot scale effectively drew the users' attention to the main character according to the accuracy of the question. At the end of the experiment, the users were asked to answer question 2. Question is based on the assumption that the viewer is the director, and in their subjective opinion, which shot scale could be better for presenting the main subject.

In ***Camera Movement Experiment***, each participant will watch three CVRs five times, with different camera movements and at different speeds. After watching each CVR, users were asked to rate their dizziness on different movements at different speeds. This was judged on a five-point Likert scale, with 1 being no dizziness at all, 2 being no dizziness, 3 being fair, 4 being dizzy, and 5 being extremely dizzy. The ratings of the 23 users were used to analyze the impact of different movements and speeds on users' viewing experience.

In ***Full-Length Shot Experiment***, users are required to answer questions related to the content. During the experiment, we record the HMD Yaw axis data. After their watching, participants choose their preferred CVR guidance method. To get the benchmark rotation data, we set a virtual HMD to follow cat 0. In this experiment, we ask participants to watch 9 CVRs, with cat 0 as the main subject. The Yaw axis rotation data of users' HMDs will be compared to the benchmark data to explore the effect of different guidance on the users' attention. To infer their understanding of stories, we analyze the accuracy of their answers under nine types of motor behavior.

4. EXPERIMENTAL RESULTS

4.1. Shot Scale Experiment

We collected the answers to questions 1 and question 2 and calculated the accuracy for question 1. The result is shown in Table 1. To our knowledge, it is pretty obvious that a closer shot will better indicate the main character, where a greater percentage of correct answers can be seen from the table for close shots and close-ups. In the medium shot, 45.5% of users chose the wrong answer cat 3 as the main character because it was much closer to the camera.

However, in question 2, 40.9% of users preferred to use bigger shot scales (full shot and medium-full shot) to show the main character. 31.8% choose closer shot scales (close-up and close shot), and 27.2% chose medium shot. We found

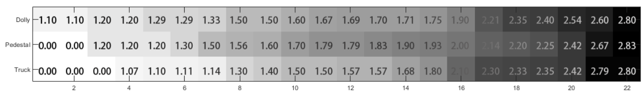
Table 1: The correct rate of answers in shot scale experiment.

Basic kinds of shot scale	Full shot	Medium full shot	Medium shot	Close shot	Close-up
Accuracy(%)	47.8	68.2	45.5	90.9	100.0

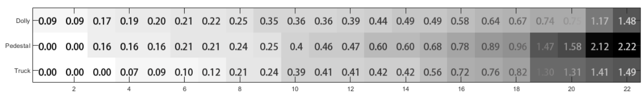
that most users were influenced by traditional cinematography techniques and considered photography composition, favoring shot scales that showed characters' relationships. But this is opposite to our results mentioned above. Therefore, when making a CVR, the artist has to consider the distance between the camera and the main character, whereas to prevent the CVR shot designing from stereotypes of traditional principles.

4.2. Camera Movement Experiment

We calculated the variance and average of each user's ratings for the three types of movement: pedestal, truck, and dolly, as shown in Figure 3. We use the average as a criterion to infer the degree of users' dizziness. When the value goes higher, it means the participant states more dizzy and uncomfortable under this camera movement. The variance was utilized to determine users' sensitivity of movement. When the variance greater, it means the user was more sensitive to the speed changes of that camera movement. For the pedestal, participants rate higher than the truck in Figure 3a. After the experiment, some users said that the pedestal features caused a sense of weightlessness. While the dolly ratings are more numerically stable in Figure 3b, which means they are less sensitive to speed changes.



(a) average



(b) variance

Fig. 3: The users' rating grades of comfort in camera movement experiments.

We believe that the pedestal movements produce the most vital sense of dizziness in immersive environments, where viewers are sensitive to the speed changes. Therefore, the CVR creater needs to manage the length and speed of pedestal shot. Truck are more comfortable to the viewer and are suitable for longer CVRs viewing. The speed of the dolly does not have a significant impact on the viewer's experience. However, compared to other camera movements, the uncomfort of viewers is relatively intense. Therefore, we suggest CVR makers should carefully consider their dolly shot at length and speed.

4.3. Full-Length Shot Experiment

We compared the accuracy of questions in Table 2, where answers related to the movement, size, and counting cats in CVR were higher (more than 60%). However, when the questions were about the color and labeled number, the correctness was lower than 35%. We suggest that, when viewing the full-length shot in CVR, users' eyepoints were constantly moving, so they were more impression on moving and conspicuous objects rather than details on their appearance.

Table 2: The accuracy of content-related answers in full-length shot experiments.

Question No.	1	2	3	4	5
Accuracy(%)	100.0	8.7	95.7	60.9	34.8
Question No.	6	7	8	9	
Accuracy(%)	65.0	13.0	100.0	39.1	

According to HMDs Yaw axis rotation data, we compared them with the benchmark data, which is highlighted as the red line in Figure 4. The offset between the user's and benchmark data indicates the efficiency of following main subject. It can be concluded from there plots that, when subject crosses less regions, the faster and intenser guidance tends to attract the audience's attention more efficiently, while subject crosses more regions, the slower and gentler guidance performs better. As seen in the table, we find that viewers are more concentrated under the trot guidance, which shows that moderate guidance could benefit to attracting and maintaining attention.

During the experiments, when the subject goes from zone A to D, some users turns around from another direction to observe it. Though it is considered as a failure of guidance in the data, the user's attention is still successfully directed to the story area. Finally, 43.48% of the participants prefer the jumping guidance due to its dynamic performance. Meanwhile, 34.78% chose walking for the reason that it is easier to observe. In CVR, a full-length shot can successfully direct the viewer's attention and their FoV. To design an appropriate camera shot, the speed of guidance should be considered more comprehensively, where fast speed brings users' discomfort, while a slow guidance may lose their interests.

5. CONCLUSION

In this work, we explore the effect of cinematography on the CVR viewing experience in an immersive environment. Three within-subject repeated-measures experiments were designed to investigate the correlation between viewing experience and comprehensive factors, such as attention, sustainability, and guidance. To help CVR creaters effectively attract viewers' attention, minimize the cybersickness, and deepen viewers' understanding of contents, we suggested applicable principles for CVR-making. In future work, we are willing to make comparative experiments of traditional cinematography vs. CVR with more cinematography techniques. While immersive viewing presents many challenges for evaluation, proposing diverse subjective and objective testing methods could help us have a better understanding of user experiences, improve the creation of content, and engage CVR entertainment enjoyably.

6. ACKNOWLEDGEMENT

This work was supported by Beijing Natural Science Foundation (No. 4214073) and the National Social Science Fundation Art Project (No. 20BC040).

7. REFERENCES

[1] Kata Szita, Pierre Gander, and David Wallstén, "The Effects of Cinematic Virtual Reality on Viewing Experience and the Recollection of Narrative Elements," *PRESENCE: Virtual and Augmented Reality*, vol. 27, no. 4, pp. 410–425, 07 2021.

[2] Rosa María Baños, Cristina Botella, Mariano Alcañiz, Víctor Liaño, Belén Guerrero, and Beatriz Rey, "Immersion and emotion: their impact on the sense of presence," *Cyberpsychology & behavior*, vol. 7, no. 6, pp. 734–741, 2004.

[3] Paulo Bala, Ian Oakley, Valentina Nisi, and Nuno Nunes, "Dynamic field of view restriction in 360? video: Aligning optical flow and visual slam to mitigate vims," *Conference on Human Factors in Computing Systems - Proceedings*, 2021.

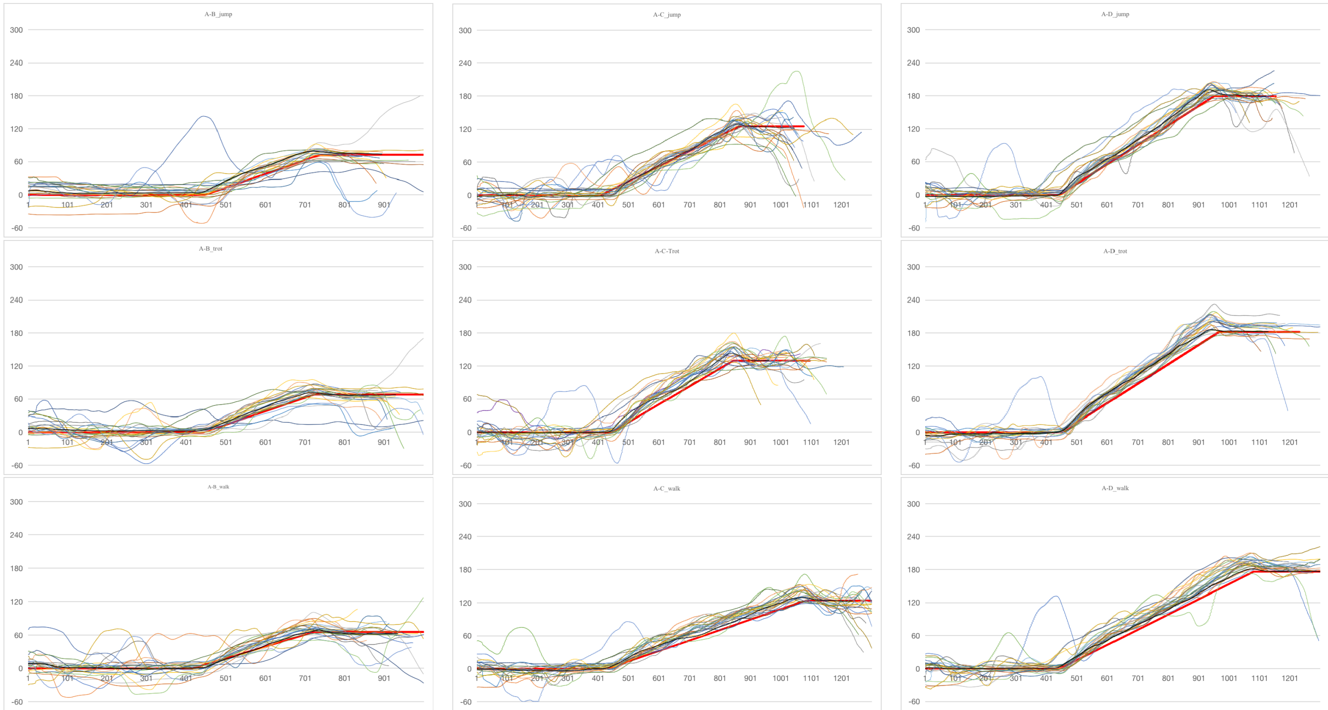
[4] Tina Kjær, Christoffer B. Lillelund, Mie Moth-Poulsen, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin, "Can you cut it? An exploration of the effects of editing in cinematic virtual reality," *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, vol. Part F131944, 2017.

[5] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen, "Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, 2017, pp. 255–265.

[6] Adam Kvisgaard, Sune Øllgaard Klem, Thomas Nielsen, Eoin Ivan Rafferty, Niels Christian Nilsson, Emil Rosenlund Høeg, and Rolf Nordahl, "Frames to zones: Applying mise-en-scène techniques in cinematic virtual reality," in *2019 IEEE 5th Workshop on Everyday Virtual Reality (WEVR)*, 2019, pp. 1–5.

[7] John Bucher, *Storytelling for Virtual Reality*, 2017.

[8] Chong Cao, Zhaowei Shi, and Miao Yu, "Automatic Generation of Diegetic Guidance in Cinematic Virtual



(a) A to B (b) A to C (c) A to D

Fig. 4: The HMDs Yaw rotation data of participants and benchmarks under nine guidance modules in camera movement experiments.

Reality," *Proceedings - 2020 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2020*, vol. 1, pp. 600–607, 2020.

[9] Morteza Malekmakan, Wolfgang Stuerzlinger, and Bernhard Riecke, "Analyzing the Trade-off between Selection and Navigation in VR," *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*, pp. 10–12, 2020.

[10] James J Cummings and Jeremy N Bailenson, "How immersive is enough? a meta-analysis of the effect of immersive technology on user presence," *Media Psychology*, vol. 19, no. 2, pp. 272–309, 2016.

[11] JJ-W Lin, Henry Been-Lirn Duh, Donald E Parker, Habib Abi-Rached, and Thomas A Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Proceedings ieee virtual reality 2002*. IEEE, 2002, pp. 164–171.

[12] Isayas Berhe Adhanom, Nathan Navarro Griffin, Paul Macneilage, and Eelke Folmer, "The Effect of a Foveated Field-of-view Restrictor on VR Sickness," *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020*, , no. February, pp. 645–652, 2020.

[13] Ajoy S Fernandes and Steven K Feiner, "Combating vr sickness through subtle dynamic field-of-view modification," in *2016 IEEE symposium on 3D user interfaces (3DUI)*. IEEE, 2016, pp. 201–210.

[14] A Fleming Seay, David M Krum, Larry Hodges, and William Ribarsky, "Simulator sickness and presence in a high fov virtual environment," in *Proceedings IEEE Virtual Reality 2001*. IEEE, 2001, pp. 299–300.