Action Units: Exploring the Use of Directorial Cues for Effective Storytelling with Swivel-chair Virtual Reality

Lingwei Tong
HIT Lab NZ, University of Canterbury
Christchurch, New Zealand
kris.tong@pg.canterbury.ac.nz

Sungchul Jung
HIT Lab NZ, University of Canterbury
Christchurch, New Zealand
sungchul.jung@canterbury.ac.nz

Richard C. Li
HIT Lab NZ, University of Canterbury
Christchurch, New Zealand
richard.li@canterbury.ac.nz

Robert W. Lindeman
HIT Lab NZ, University of Canterbury
Christchurch, New Zealand
gogo@hitlabnz.org

Holger Regenbrecht
Department of Information Science,
University of Otago
Dunedin, New Zealand
holger.regenbrecht@otago.ac.nz

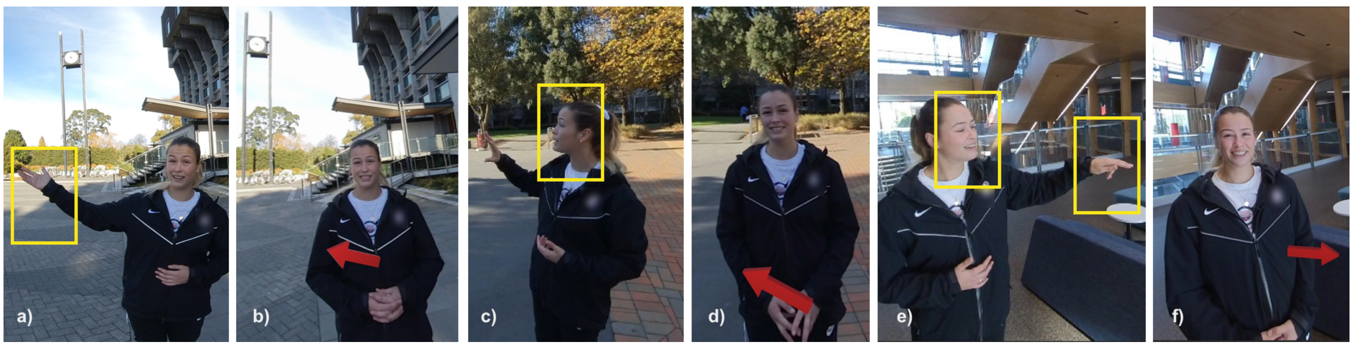

Figure 1: Snapshots from the campus tour video with attention guidance methods applied. The AU cues used (highlighted in yellow boxes) were a) pointing only, c) pointing then looking back, and e) looking and pointing at the same time. Another method "Pointing Arrow" (red arrows pointing to the targets), being compared in this study, are shown in b), d) and f), with the scenes matching to a), c), and e), respectively.

ABSTRACT

The popularity of 360-degree video storytelling has been increasing. However, effectively guiding viewer's attention remains challenging. In this paper, we propose the notion of Action Units (AU) as a guidance method. It aims to improve the user experience of 360-degree video for seated viewers with swivel chairs, as we call "Swivel-chair VR". We first conducted a pilot study with four subject-matter experts to verify the practicality of AU. We then conducted a formal user study to compare the AU with two commonly used attention guidance techniques, namely "Pointing Arrow" (PA) and "Angular Shift" (AS), as well as the baseline without any guidance (BK). We applied them to a virtual tour and measured their effects on engagement, enjoyment, memory, viewers behaviors, and cybersickness. The results indicated that AU is an effective guidance as it increased the levels of engagement and enjoyment, reduced the level of cybersickness and helped users focus on target faster. Users also preferred AU for its diegetic aspects.

OzCHI '20, December 2–4, 2020, Sydney, NSW, Australia

ACM ISBN 978-1-4503-8975-4/20/12…$15.00
https://doi.org/10.1145/3441000.3441063

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; *Visualization systems and tools.*

KEYWORDS

Virtual Reality, 360-degree video, storytelling, social cues

ACM Reference Format:
Lingwei Tong, Sungchul Jung, Richard C. Li, Robert W. Lindeman, and Holger Regenbrecht. 2020. Action Units: Exploring the Use of Directorial Cues for Effective Storytelling with Swivel-chair Virtual Reality. In *32ND AUSTRALIAN CONFERENCE ON HUMAN-COMPUTER INTERACTION (OzCHI '20), December 2–4, 2020, Sydney, NSW, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3441000.3441063

1 INTRODUCTION

In recent years, 360-degree videos have become a popular way for people to experience Virtual Reality (VR) since both the VR headsets and consumer-level 360-degree cameras are widely available [4]. Different from conventional video, viewers watch 360-degree videos with VR headsets, using free head rotation to decide where to look. This intrinsic property of 360-degree video hinders satisfactory user experience and leads to a "narrative paradox" [1]. It describes a conflict between the authorial control, which is a pre-determined nature of a narrative, and the freedom of interaction and participation of a user in an immersive environment, such as VR [41]. It is also the cause of missing important elements on the user's end, known as the "Fear of Missing Out (FOMO)" problem [29], yielding weak narrative comprehension and low emotional engagement.

To improve the user experience, several attempts have been made by modifying the filmmaking grammar of viewpoints, shots, and scene cuts [30, 38]. However, filmmaking grammars are mostly defined by the camera frame and camera movement, which are not available in VR content. Viewers reported frequent loss of tracking of the key object in the scene, leading to failure of story comprehension [35]. Others tried importing navigation aids from video games, such as lighting changes [6], arrows [33], and scene shifting [19, 26], to direct viewer attention towards Regions of Interest (ROIs) without taking away free agency of exploration. But researchers also noticed that these navigation aids required post-processing or changes to the video content, potentially breaking the viewer's feeling of presence, and so were less preferred for VR use [26].

In this research, we propose Action Units (AU) as a new method of viewer attention guidance, in the form of a set of directorial cues used by 360-degree video content creators when telling immersive stories. The AU aims to address the narrative paradox and FOMO issues, and also remove the requirement of post-processing work usually posed by other attention-guidance techniques. The AU makes use of the social cues from the storyteller as guidance cues, including head, arm, and eye movements during the narrative, instead of artificial add-ons. They are depicted in Figure 1 a), c) and e). Before conducting a formal study, we asked four Subject-Matter Experts (SMEs) (a botanist, a polar photographer, a glaciologist and a volcanologist) to use our AU approach in their 360-degree documentary video production. Based on our findings from the pilot study, we then conducted a formal user study to compare the effects of AU with two commonly-used guidance techniques, which are "Pointing Arrow" (PA) and "Angular Shift" (AS), in terms of their effects on engagement, enjoyment, memory, viewer behavior for searching and attention, and cybersickness. We discovered that the AU is an effective attention guidance technique for 360-degree video storytelling.

2 RELATED WORK

We reviewed relevant work on 360-degree video production, 360-degree video first-person experience design, and attention guidance techniques for 360-degree videos.

2.1 Film on a Flat Screen vs. 360-degree Video

Immersive storytelling using 360-degree videos, also known as Cinematic Virtual Reality (CVR), is an experience where the viewers immerse themselves in a panorama setting with the content, instead of staring at a rectangular screen in front of them [30]. In the early stage, researchers started looking for the proper methods of storytelling using CVR by comparing it with the traditional films.

In films for flat screens, directors rely on a series of cinematic techniques to guide the audience's attention, invoke curiosity and suspension of disbelief, deliver expressive content, and in total, effectively tell a story [11, 12]. Among the main techniques, the "mise-en-scene" and "cinematography" are two that occur in the production phase [7]. Moving from a 2D flat video to a 360-degree video, the first obvious change is the Point of View (POV). When watching a 360-degree video, the viewer sits in the center of the scene, instead of looking at a rectangular flat screen. Syrett et al. [38] stated that with this new POV the viewer becomes the narrator since she can choose what to see and what to understand. The change of the camera system also contributes to this POV shift. In the "updated cinematography language and edits" for 360-degree video production from Chang [9], they pointed out that a 360-degree camera has a fixed focal length (no zooming) and several wide-angle lenses to cover the entire scene (no panning), thus commonly used cinematic terms such as "Pull-in" to focus onto an object or "Pan" to reveal an object outside the frame space are no longer available. Mateer [30] also introduced the "Theory of Transportation" aimed to assist those CVR content creators to be more certain of incorporating effective storytelling within this new medium. However, due to the change of audience POV when watching, the specialty of 360-degree camera features, and the disappearance of the "screen frame", many of the grammar elements, such as various shots, moving the camera, or time-compression editing, are not applicable in CVR content [13]. The change of the user's role in the narrative and the intrinsic characteristics of the 360-degree video itself were also not thoughtfully considered in those works.

Unlike the absolute authorial control in traditional filmmaking, researchers also used various cues to direct user attention to ROIs, while maintaining user's freedom of exploring the scene. Elements from video games including virtual arrows [8, 16, 18, 33], signs, markers [42], and audio [31, 32] were imported into the scene as the cues to direct viewer attention to the expected ROIs. However drawbacks have also been observed. Nielsen et al. [31] pointed out when a cue was applied, such as an arrow, it presented in the scene as a non-diegetic object. It led to a reduction of presence since the viewer was constantly looking at an item that was not contextually related to the environment. Lin et al. [26] used scene rotation as the cue to avoid non-diegetic intrusion. They rotated the rendered scene in real time to directly bring the viewer to the orientation of the ROI. However the level of cybersickness spiked when the scene rotated. Besides, production-wise those cues are all applied in the studio after the content is captured. Using these approaches therefore requires a large investment of time, resources, and skills in sophisticated post-production software.

2.2 Theatre and Attention Guidance in 360-degree Videos

Different from cinema or video games, theatre is not pre-recorded, but rather a real time performance, which runs a play without breaking chronological sequence [2]. Thus methods from filmmaking, such as cuts and camera movements, are no longer available. Play writers and directors in theatres instead implement new techniques to draw audiences' attention to the key elements [12]. In common

Figure 2: Examples of the shot types for 360-degree videos developed by Lindeman [27]. They are a) "Look There Shot", a hand and head gesture where the presenter points and looks out at a distant point of interest, then looks back at the camera; b) "Come Along Shot", A camera movement that seems to beckon the viewer to move along with the presenter at a measured pace; c) "Look Here Shot", A camera movement that brings the camera towards a very close, immovable ROI. All those images are screenshots of 360-degree videos from the viewer's POV.

practices, the audience's attention is usually controlled with lighting, stage setup, sound cues, and exaggerated movements [28]. On the other hand, the actor's body language, such as the opened body position, the long-lasting orientation towards the audiences, and more exaggerated body movements and gestures, become the key component [5].

Lighting control was first migrated from theatre stages into virtual scenes as an attention guidance technique, such as [3, 6, 14], but they still require post processing work. Several others have looked into combining attention guidance into the production stage, rather than post-processing. One of the early trials was the experiment conducted by Brown et al. [36], in which they scripted the actions of human actors in front of a 360-degree camera, asking them to use certain actions and sounds to make the viewer aware of specific events happening in the scene. Lindeman [27] also worked on providing a "shots reference" for 360-degree video capturing, by looking into both body languages used in stage performance and the traditional filmmaking grammar. He borrowed the concept of "lexicon of shot types" from the filmmaking industry and developed a series of "shot types" for 360-degree video production by combining several common social communication gestures, and camera movements together. In Figure 2, several examples of "shot types" are shown.

3 ACTION UNITS

The work of [27, 36] indicated that the application of social communication cues as attention guidance in 360-degree videos can be effective and at the same time avoid post processing. Cues like those have also been intensively utilized theatre stages [28]. In this paper, we propose the Action Units (AU) as a directorial method and an attention guidance technique a storyteller can adopt for on-site use. Compared to the other guidance techniques which require post-processing work, the AU can be implemented at the same time when the 360-degree videos are being captured.

We started the development of AU using 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 [23]. Tomasello et al. [40] stated that for mutual understanding, a direct speech was necessary as the communicative context. Accompanying speech, non-verbal cues including facial features and gestures were frequently utilized to convey extra information like intentions and changes of foci [10, 17]. By comparing films with real life conversations, Kappelhoff and Müller [20] 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 come in 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, the viewer enters a temporary egocentric position where being addressed by the storyteller in a (virtually) face-to-face manner, we believe that social cues can trigger the same effects as in real life communications, and work as attention guidance cues. Ravenet et al. [34] also stated that people use recurring patterns, e.g. gestures, to map conceptual metaphors from an entity to another. It 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.

We then created the AU from three frequently used non-verbal units of social cues [15]. Among them, arm and head movements are used to emphasize objects, eye contact is used for establishing social acknowledgment and to increase the level of co-presence [21]. Therefore, we composed three AUs for 360-degree video production based on these non-verbal cues (Examples are shown in Figure 4). They are:

- **AU01**: Conversation - the storyteller keeps eye contact with the viewer;
- **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 then choose which AU to use according to the content and the narrative she wants to deliver. The AU serves as call-to-action points, so when key elements appear, the storyteller will be assured that her viewers will follow and will have an engaging viewing experience. When one is wearing a VR headset and watching a 360-degree video of an 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 AU 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.

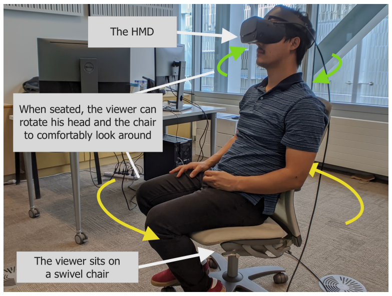

Figure 3: Swivel-chair VR: A viewer wearing a HMD, watching a 360-degree video while sitting in a swivel chair. When seated, he can turn his head and the chair to look around comfortably.

Figure 4: Examples of experts using AUs (marked with yellow rectangles) from virtual field trips recorded as 360-degree videos. 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.

4 METHODS

4.1 Swivel-chair VR

With the AU composed, we first explored the preferred user scenario for 360-degree videos, especially the ones where in the content storytellers are giving a presentation about a certain topic or object. In traditional filmmaking, the directors assume a passive viewer sitting in a chair looking straight ahead [7]. Similarly, we put forward the idea of Swivel-chair VR 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. 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, which can lead 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 has to be managed though, e.g. with the help of an operator.

4.2 The Pilot Study - AU Field Tests

In a pilot study, we asked four SMEs to apply the AU when capturing their scientific activities as virtual field trips using 360-degree cameras. We first trained the experts with the AU by explaining how to use them and what scenarios were suitable for each AU. We also helped them compose their shooting plans according to the locations and intended content. As with traditional filmmaking, storytellers need be able to convey their ideas about what they expect for a given shot or shot sequence. In those field trips the AU was used as anchor points for the experts to formulate shooting plans. This allowed us to collaboratively get a better understanding of the video itself.

We extracted multiple snapshots from the 360-degree videos produced by the SMEs and highlighted where an AU was used, to demonstrate how it served the purpose of bringing effective storytelling to the 360-degree video. The snapshots are shown in Figure 4. 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 video).

In one case, 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 4 show the camera was stationary among a group of people, mimicking the viewer just being one of them. The expert used both AU02 and AU03 when he was introducing both the mountains at a far distance and the crevices under the snow nearby. It helped the viewer to locate and identify the ROIs that were being mentioned and to keep up with the narrative. In another case, 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 4. In the first two snapshots, the expert applied both AU02 and AU03, for the lava features on the mountain at a far distance, and rocks nearby. The AU worked as visual guides to aid the viewers to shift 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 viewer the feeling they were standing by the stream with him and listening to the talk at the same time.

In the interviews with SMEs they indicated that the AU notation was easy to implement and requires no extra burden of technical skills. We also brought the videos to their classrooms and asked their students to watch those virtual field trip videos with the Swivel-chair VR setup. The students anecdotally commented that they were having 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 AU in actual productions, from the perspective of SMEs as content creators, as well as its positive effects on attention guidance and reducing the FOMO issue.

4.3 Experimental Design

Based on the finding from the pilot study, we decided to move further to discover if the AU could be as effective as the commonly used post-processing methods or even outperform them, from the perspective of viewers' experience.

We conducted a within-subjects experiment to compare AU performance with two other attention-guidance techniques, "Pointing Arrow" (PA) and "Angular Shift" (AS), as well as the baseline without any guidance (BK) [19, 33]. Based on the series of 360-degree videos of a campus tour as our experiment environment, we implemented different guidance styles on the videos as the independent variable (type of attention guidance) with three levels (AU, PA, AS). Each participant experienced all the types of guidance. In the study, we measured their effects on user behaviors (the time took to locate an ROI and the time of attention on that ROI), recall rate, and subjective feelings (levels of engagement and enjoyment). We explored whether the AU would out-perform PA and AS to 1) increase levels of engagement and enjoyment, 2) affect the recall rates, 3) reduce the search time and increase the attention time on an ROI, and 4) reduce cybersickness, when applied to 360-degree videos.

4.4 Hypotheses

We proposed four hypotheses based on comparing AU with guidance techniques used by other researchers in their experiments. In the experiment conducted by Nielsen et al. [31], they pointed out that the synthesized non-diegetic cue reduced the viewer's perception of presence. Since the AU was embedded within the content during the production stage and was performed by the storytellers themselves, they were considered diegetic, thus would impede less on the viewer's feeling of presence, so as the level of engagement and enjoyment. Regarding the recall rates, Li et al. [24] observed 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 comparing to the use of conventional learning materials. Thus we propose the following two hypotheses:

- **H1**: A viewer will feel a higher level of engagement and enjoyment when watching a 360-degree video shot using AU, compared to PA and AS.
- **H2**: Compared to BK, A viewer's recall rate will be lower when AU, PA or AS is used in the video..

When testing a variety of guidance techniques with 360-degree videos, both Speicher et al. [37] and Lin et al. [26] stated that methods like the AS will introduce cybersickness and disorientation. In [37] the researchers also pointed out the accuracy of locating an expected item in the scene was higher when using a human actor as a guidance cue than using an added object. Those indicated that AS will deteriorate the viewer's performance on searching and locating as it will introduce cybersickness. Also, the PA used an add-on as the cue, while AU used human actors. Thus we also propose another two hypotheses as following:

- **H3**: AU will reduce the Time-to-Search for ROIs and will extend the time a viewer stays on ROIs, compared to PA and AS.
- **H4**: AU will reduce a viewer's level of cybersickness when watching a 360-degree video, compared to PA and AS.

4.5 Measurements

In this study, we measured the levels of engagement and enjoyment by the Questionnaire of Engagement Enjoyment and Immersion (E2IQ) [25]. We divided the E2IQ into two parts (E1 and E2) where the 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 engaged/enjoyable at all, 1 = very engaging/enjoyable). The memory effects were evaluated by recall tasks that asked subjects to identify objects which were both visible and mentioned by the tour guide (but not described in details) in the videos. We measured the ease of search with the Time-to-Search (TTS) counted from when the tour guide explicitly mentioned the ROI to the point when the viewer located the ROI, i.e. it crossed the 29.2 degrees threshold (half of the single eye FOV of the Head Mounted Display (HMD), which regard to be the peripheral limit [39]) from the outside of her field of view (FOV). The level of attention was measured with the Time-on-Target (TOT) counted from when a participant located the ROI until it left her FOV (ROI crossed the 29.2 degrees threshold and 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) [22],

4.6 Apparatus

In this study, the video playback and attention guidance techniques were all implemented in Unity3D 2018.3.11f1. We used a computer running 64-bit Windows 10 Professional with a 3.2GHz i7 processor and a GeForce RTX 2080 graphics card, 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.

4.7 Material

Before the actual experiment, we captured eight 360-degree video clips of a campus tour with an Insta360 ONE X 360-degree camera [2], at the resolution of 5.7k (5760 x 2880). The recordings were conducted 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. We captured two takes 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 later augmented with PA and AS in a post-processing step and also used as the Baseline (BK). The camera was mounted on a tripod thus the viewpoint was fixed in each of the video clips. The distance of the actor to the camera was kept the same (2m) at each location so the viewer would feel 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

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

between the recordings. All the videos were trimmed to a length of approximately 90 seconds. The AU clips were ready for the experiment after this adjustment since they already had the cues embedded in the content. 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 rotated constantly to point itself towards the ROI being described by the tour guide at the moment, as shown in Figure 1 b), d) and f). AS was also added to the Blanks using Unity3D. The scenes were automatically rotated to shift the ROI to the front of the viewer when her head rotated towards the ROI past a certain threshold. We used the methods proposed by Tanaka et al. [39] as a reference when implementing this method. The rotation speed was set to 15 degrees/s to avoid rapid movement. The threshold was set to 50 degrees from the center of FOV, which is on its edge, for the HMD we used in our experiment. The shift took place when the ROI is outside of this threshold, and stopped once the ROI is within that threshold. The mechanism is also illustrated in Figure 5.

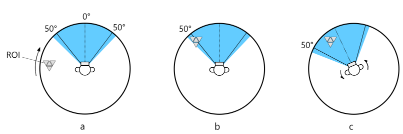

Figure 5: An example of the AS method, illustrated in a top view of the viewer wearing a HMD and the scene around her. The head orientation is marked as a dashed line, the FOV is highlighted with the blue shade. a) The viewer is looking towards the front of the scene. Now the ROI is located to her left, out of her FOV and outside the 50 degrees threshold. So AS is taking place, shifting the ROI towards the front of the viewer; b) The AS has reached the threshold, then it stops. The ROI is now within the FOV; c) When ROI is within the FOV, although the viewer rotates her head, AS will not be activated, the scene is not rotating, unless the ROI again falls out of the threshold.

4.8 Participants

We recruited 24 participants (14 females) 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, eighteen of them reported had experienced VR , but limited to only a few times a year. Among those with limited experience, eleven of them 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.

4.9 Procedures

We compared four conditions (AU, PA, AS, and BK) in this experiment by applying them to the captured 360-degree videos and having subjects watch them with VR headsets. The experiment was approved by the ethics committee of our University.

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 VR headset, and watched an intro video to get familiar with the technology and this form of medium, to remove the effects of novelty and anxiety, before starting watching the actual content. The intro video had neither a tour guide nor any attention guidance technique attached.

After the intro video, each participant watched four content video clips of the "campus tour". To keep the integrity of the story, every participant watched content video clips 1 to 4 in the same order, but with four conditions (AU, PA, AS, and a "no guidance" baseline) randomly assigned and counterbalanced with a Latin Square. At the end of each clip, we asked the participant to remove the headset, and immediately move on to complete a viewing experience questionnaire to measure cybersickness, engagement, enjoyment, then the recall tasks, while the experience and memory were still fresh. Then we gave them a one minute break before moving on to the next clip. Each participant watched a total of 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.

Table 1: The mean values of the results of Cybersickness, Engagement, Enjoyment, Recall rate, TTS, and TOT, 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**

5 RESULTS

We analyzed the data from both the participants' reported questionnaires and the logged time values from the Unity Engine by using one-way ANOVA tests ($\alpha < 0.05$). Table 1 shows the mean values of SSQ scores, E2IQ scores for Engagement (E1) and Enjoyment (E2), the recall rates, TTS, and TOT, for each condition. The most outstanding values of each measurement are highlighted in the table. The overall results are also summarized and plotted in Figure 6.

5.1 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 (AS: $M = 0.16, SD = 0.455$). The ANOVA test indicated that there were 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 there were no other statistical differences found between AU and

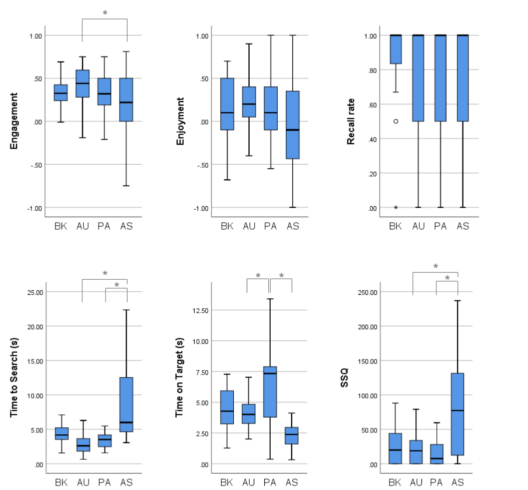

Figure 6: Boxplots summarizing the results of levels of Engagement and Enjoyment, Recall Rates, viewer performances of search and attention, and the level of discomfort of each condition, in terms of medians, interquartile ranges, minimum and maximum ratings. Top row, from left to right: Engagement, Enjoyment, and Recall rate. Bottom row: TTS, TOT, and SSQ.

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 ($M = 0.12, SD = 0.39$). But participants reported lower levels of enjoyment with AS (AS: $M = -0.039, SD = 0.53$). However, the ANOVA test did not reveal any statistically significant differences in the level of enjoyment among the four conditions.

5.2 Memory

The recall rates were all lower than the baseline ($M = 0.8825, SD = 0.24$), 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 those four conditions ($F = 0.968; p = 0.411$). We also did a Post-hoc comparison using the A 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 ($p = 0.999$).

5.3 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 (AS: $M = 8.30, SD = 5.310$). An ANOVA test ($F = 17.992, p < 0.001$) also indicated that there were 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). But we did not see a significant difference 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 ($p = 0.013$).

5.4 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 videos with AS, the level of cybersickness was increased ($M = 80.773, SD = 74.013$). The ANOVA test showed significant differences among their performance ($F = 12.344, p < 0.001$). The Post-hoc test indicated AS was out-performed significantly by other three conditions (BK-AS: $p < 0.001$, AU-AS: $p < 0.001$, PA-AS: $p < 0.001$). However, we could not detect any significant differences between the AU, PA, and BK. The results indicate that participants felt less discomfort when watching the videos with AU or PA, compared to the ones with AS.

5.5 User Preference

In post-test interviews, we asked each of the participants Q1: Which attention guidance method they thought was the easiest to use?, Q2: Which method was the most uncomfortable?, and Q3: Which in general did they prefer? The totals of each choice made by participants are shown in Figure 7. The results show 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 the general preference of the methods, 13 chose AU, 11 chose PA, and none chose AS. We also noticed that a choice for easiest 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 because the AU was diegetic and easy to understand. Those who preferred PA stated they liked how the arrow stood out from the video and was immediately recognizable. Several also pointed out the arrow felt strange because it was not part of the scene but was an add-on.

6 DISCUSSION

We discuss the results of the experiments together to highlight the main findings and implications.

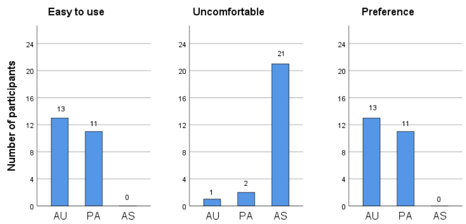

Figure 7: Bar charts showing the total count of participants on each of the preference choices. Left: Which method was the easiest to use? Middle: Which method was the most uncomfortable? Right: Which did you prefer?

6.1 Implications for Engagement, Enjoyment, and Discomfort

The study results showed that both the AU and PA increased the level of engagement and enjoyment, compared to the baseline. They also both reduced the level of cybersickness. These results support hypotheses H1. Since AU did reduced the level of cybersickness comparing to AS, but not as PA, the hypotheses H4 is only partially supported. For AU, it was implemented during the production phase by the storyteller, 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 they were being addressed in a face-to-face manner by the storyteller, and thus felt more engaged in the narrative. They were also less aware of the fact that they were wearing VR headsets and looking at simulated images.

6.2 Implication on Memory Effects

The study results indicated that the recall rates of the three conditions with cues applied were all slightly lower than the baseline. However, there was no significant difference between them. This did support H2, which predicted that attention guidance would lower the recall rates, acting partially as distraction in the recall task. 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 exact details were not included in the narrative from the storyteller, we believed that the introduction of cues did distracted the viewer from carefully inspecting the ROIs with visuals. The guiding cues, paired with the acting from the storyteller, both dragged the viewer's attention more towards the narrative (which no details were covered). Thus, viewers turned out to remember less of the visual details. However the difference between the amount of decrements was insignificant so we are unable to tell if AU was less distracting than PA or AU.

6.3 Implications for 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 conclude 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 H3, that the AU will reduce TTS when applied as an attention guidance technique. We also believe that AS elongated TTS because when the AS was applied, the video content itself was still a blank scene with no visible cue. The scene rotation happened during the narrative made viewers confused, 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 difficulties to understand what to do before they realized it was part of the system design.

The results of TOT showed a different trend than TTS. We found that when PA was applied as the attention guidance, viewers stayed significantly longer on the ROIs after they were mentioned by the storyteller. We concluded 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, while the AU was only visible when the viewer was looking at the storyteller. This result does not support hypothesis H3, indicating the need for a further look into the effects of cue exposure time. TOT results also showed that participants dwelled less when AS was applied, compared to the other conditions. One clear 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 they were trying to "fight the system" and subjectively looked away to resist, as they felt the system was forcing them to look at something and taking away their free agency of looking around. This led them to look at the ROIs for only a relatively short period.

6.4 Design Reference for Content Creators

From both the pilot study and the user study, we discovered properties of AUs that were different from other attention guidance methods. On the one hand, AUs were applied directly during the capture of content, without the requirement of any post-processing work. Also, a storyteller could compose AUs into her script as a part of the content itself. She could then naturally perform it when delivering the narrative. This has huge potential and is a suitable method for creators who need to share and publish their work soon after it is captured, such as field documentary videos for science-outreach purposes. In other storytelling scenarios but not using 360-degree videos as their media, such as computer-synthesized 3D scenes, or immersive environment with real avatars or virtual characters, as long as social communication is part of its interaction, the actors, humanoid avatars and virtual characters can all embed AUs into their actions. We expect that the usage of AUs in the production stage will help presenters to ensure the players will follow the story line and have a higher level of presence in the immersive scene.

However, we also discovered a trade-off between AUs' effects on attention (i.e., the recall rate and TOT) and the effects on the levels of both engagement and enjoyment. From the previous analysis, we noticed that when applied to videos, AU increased the participants' levels of engagement and enjoyment, and directed them faster to the ROIs. But it also distracted them from paying attention to details of the ROIs. The participants recalled the details less than we expected. This was also reinforced by the feedback from participants during the post-test interview, as some preferred AU for its diegetic characteristics and some others preferred PA for its outstanding contrast and strong implications. This could be an important design reference for content creators, as one needs to make decisions by considering the effects of AU, the content she wants to deliver, and the response and results she expects to see from viewers. AUs can be useful for scenarios like scientific presentations, or other face-to-face conversations, where a natural social atmosphere is important. 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 AU as the only attention guidance may not be sufficient; adding extra cues will be recommended.

6.5 Limitations and Future Work

During the interpretation of the experimental results and reviewing the footage, we noted several limitations. First, the AU we designed was not constantly directing the viewers throughout the entire process of search and attention. Since the AU was implemented by the storyteller herself, they would only be visible when the viewer was looking at her. In other words, the AU was only effective to the viewer when it was visible in the viewport. Once the viewer turned away, such to search for the ROIs, the AU was not visible and its function of attention guidance was suspended. In the previous chapter when we looked into the results of TTS and TOT, we pointed out that the attention guidance did not become effective instantly but rather went through a built-up process over time from its appearance. Thus when we considered the total "exposure time" a viewer had under each guidance cue, one possible cause is that the built-up effectiveness of AU was closely dependent on a viewer's actual behavior when watching the video. The PA and AS, however, had lower dependencies and faster built-up processes. To reinforce the effectiveness of AU to direct viewers attention, future work will look into the integration of high fidelity spatial sound, including simulated reverb effects. Second, the video clips we used were all about 90 seconds in length. We noticed in some of the results, such as level of enjoyment (E2) and recall rates showed similarities between all three conditions, and no significant differences were found. This could be due to the relatively short time (six minutes) subjects spent watching all three videos, such that the differences between conditions were not yet revealed and detected. Third, PA and AS were both added during the post-production phase, and AU was implemented during the production phase. It is possible that implementing AU becomes an extra task added to the storyteller, other than the existing task of "present the content". The increased workload and related effects on the storyteller, when she is using AU and giving the narrative at the same time, is also worth further exploration. Forth, we assumed that when participants felt higher engagement and enjoyment, they were more immersed in the narration given by the storyteller, and were less aware of memorizing the visual details of ROIs. However, we were unable to either conclude this correlation, or to determine if the AU, as a diegetic cue, can be less distracting than the non-diegetic PA and AS, since the data we collected were limited. In future studies, we plan to adjust the design of AU with extra assisting cues that are constantly visible to the viewer to ensure it is taking effect throughout the viewing experience. We also plan to use longer video footage as study materials to ensure the participants will have long-term immersion. New filming locations and ROIs will also be carefully chosen for further study about the distraction from cues to recall tasks when watching. Lessons learned from AU in Swivel-chair VR can potentially be applied to other forms of cinematic VR, where the user's movement within the virtual environment are even less constrained. In particular, a careful integration and combination of AU and other cues to "walk" the immersed user to spaces of interest.

7 CONCLUSION

In this research, we proposed the AU system as a set of directorial cues. A storyteller can adopt the AU onsite to direct viewer's attention in an immersive storytelling experience. The AU aims to address the narrative paradox and FOMO issues that come with 360-degree videos and to increase the levels of engagement and enjoyment. We designed the system based on social cues people use in daily conversations, such as gestures and eye contact, eliminating the requirement of adding post-processing elements to the video.

The results from both the pilot study and the formal user study indicated that the AU can effectively guide the viewer's attention when applied. Comparing to the commonly used AS, the AU significantly increased levels of engagement and enjoyment. It also introduced lower level of cybersickness comparing to AS, but still surpassed by PA. It also reduced the time viewers needed to search for an ROI. Users also prefer AU over PA and AS for its diegetic property. We also discovered that when applying AU to 360-degree video productions, the creators need to consider AU's negative effect on attention time and memory of details. We also concluded that the AU can be applied to other forms of social-context based immersive storytelling activities. In further studies, we will look at the combination of AU with other proper cues to make them constantly useful and applicable to long-term immersion experiences.

ACKNOWLEDGMENTS

This work was supported by the New Zealand Science for Technological Innovation National Science Challenge project Ātea.

REFERENCES

[1] Ruth Aylett. 1999. Narrative in Virtual Environments - Towards Emergent Narrative. In *Proceedings of the AAAI Fall Symposium on Narrative Intelligence*. 83–86.
[2] Ruth Aylett and Sandy Louchart. 2003. Towards a narrative theory of virtual reality. *Virtual Reality* 7, 1 (2003), 2–9.
[3] Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle gaze direction. *ACM Transactions on Graphics (TOG)* 28, 4 (2009), 1–14.
[4] Heather Bellini, Wei Chen, Masaru Sugiyama, Marcus Shin, Shateel Alam, and Daiki Takayama. 2016. *Virtual & Augmented Reality: Understanding the race for the next computing platform.* Technical Report. Goldman Sachs Global Investment Research. 1–30 pages. https://doi.org/10.1017/CBO9781107415324.004 arXiv:arXiv:1011.1669v3

[5] Anne Bogart and Tina Landau. 2004. *The viewpoints book: a practical guide to viewpoints and composition*. Theatre Communications Group.

[6] Matt Boggus and Roger Crawfis. 2010. Distance field illumination: a rendering method to aid in navigation of virtual environments. In *International Symposium on Visual Computing*. Springer, 501–510.

[7] David Bordwell and Kristin Thompson. 2013. *Film art: an introduction* (10th ed.). McGraw-Hill, New York.

[8] Stefano Burigat and Luca Chittaro. 2007. Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human-Computer Studies* 65, 11 (2007), 945–958.

[9] Wooksang Chang. 2016. Virtual reality filmmaking methodology (animation producing). *TECHART: Journal of Arts and Imaging Science* 3, 3 (2016), 23–26.

[10] Barbara Clucas, John M Marzluff, David Mackovjak, and Ila Palmquist. 2013. Do A merican Crows Pay Attention to Human Gaze and Facial Expressions? *Ethology* 119, 4 (2013), 296–302.

[11] Robert Cohen and John Von Szeliski. 1983. *Theatre: brief edition*. Mayfield Publishing Company.

[12] Celso de Melo and Ana Paiva. 2005. Environment expression: Expressing emotions through cameras, lights and music. (2005), 715–722. https://doi.org/10.1007/11590361_15

[13] Jan Decock, Jan Van Looy, Lizzy Bleumers, and Philippe Bekaert. 2014. 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 (2014), 449–459. https://doi.org/10.1007/s00146-013-0487-6

[14] Magy Seif El-Nasr, Athanasios Vasilakos, Chinmay Rao, and Joseph Zupko. 2009. Dynamic intelligent lighting for directing visual attention in interactive 3-D scenes. *IEEE Transactions on Computational Intelligence and AI in Games* 1, 2 (2009), 145–153.

[15] Catherine S. Fichten, Vicki Tagalakis, Darlene Judd, John Wright, and Rhonda Amsel. 1992. Verbal and nonverbal communication cues in daily conversations and dating. *Journal of Social Psychology* 132, 6 (1992), 751–769. https://doi.org/10.1080/00224545.1992.9712105 arXiv:arXiv:1011.1669v3

[16] Patrick Flanagan, Ken I McAnally, Russell L Martin, James W Meehan, and Simon R Oldfield. 1998. Aurally and visually guided visual search in a virtual environment. *Human factors* 40, 3 (1998), 461–468.

[17] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 309–316. https://doi.org/10.1145/365024.365121

[18] Neil Gramopadhye, Kapil Chalil Madathil, Jeffery Bertrand, and Christopher Shannon Blair. 2014. An investigation of the effectiveness of navigational aids in a virtual environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 58. SAGE Publications Sage CA: Los Angeles, CA, 1949–1953.

[19] Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. 2016. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. (2016), 1996–2000.

[20] Hermann Kappelhoff and Cornelia Müller. 2011. Embodied meaning construction: Multimodal metaphor and expressive movement in speech, gesture, and feature film. *Metaphor and the Social World* 1, 2 (2011), 121–153. https://doi.org/10.1017/CBO9781107415324.004 arXiv:arXiv:1011.1669v3

[21] Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta psychologica* 26 (1967), 22–63. https://doi.org/10.1016/0001-6918(67)90005-4

[22] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220. https://doi.org/10.1207/s15327108ijap0303_3

[23] Mark L Knapp, Judith A Hall, and Terrence G Horgan. 2013. *Nonverbal communication in human interaction*. Cengage Learning.

[24] Chen Li, Horace H. S. Ip, Yuen M Wong, and Wing S Lam. 2020. An empirical study on using virtual reality for enhancing the youth's intercultural sensitivity in Hong Kong. *Journal of Computer Assisted Learning* September 2019 (2020), 1–11. https://doi.org/10.1111/jcal.12432

[25] JJ-W Lin, Henry Been-Lirn Duh, Donald E Parker, Habib Abi-Rached, and Thomas A Furness. 2002. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. (2002), 164–171. https://doi.org/10.1109/vr.2002.996519

[26] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360 video. (2017), 2535–2545. https://doi.org/10.1145/3025453.3025757

[27] Rob Lindeman. 2018. 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* (2018).

[28] Logan Dwight. [n. d.]. These VR Film Tips Show How To Direct Audience Attention. https://uploadvr.com/vr-film-tips-guiding-attention/

[29] Andrew MacQuarrie and Anthony Steed. 2017. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*. IEEE, 45–54. https://doi.org/10.1109/VR.2017.7892230

[30] John Mateer. 2017. Directing for Cinematic Virtual Reality: how the traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice* 18, 1 (2017), 14–25. https://doi.org/10.1080/14682753.2017.1305838

[31] Lasse T Nielsen, Matias B Møller, Sune D Hartmeyer, Troels CM Ljung, Niels C Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. (2016), 229–232. https://doi.org/10.1145/2993369.2993405

[32] Tabitha C Peck, Henry Fuchs, and Mary C Whitton. 2009. Evaluation of reorientation techniques and distractors for walking in large virtual environments. *IEEE Transactions on Visualization and Computer Graphics* 15, 3 (2009), 383–394. https://doi.org/10.1109/TVCG.2008.191

[33] Andrea C. Pierno, Andrea Caria, and Umberto Castiello. 2005. 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 (2005), 728–737. https://doi.org/10.1518/hfes.46.4.728.56815

[34] Brian Ravenet, Catherine Pelachaud, Chloé Clavel, and Stacy Marsella. 2018. Automating the production of communicative gestures in embodied characters. *Frontiers in psychology* 9 (2018), 1144. https://doi.org/10.3389/fpsyg.2018.01144

[35] Sylvia Rothe, Heinrich Hußmann, and Mathias Allary. 2017. Diegetic cues for guiding the viewer in cinematic virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. 1–2. https://doi.org/10.1145/3139131.3143421

[36] Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. 2016. Directing attention in 360-degree video. (2016). https://doi.org/10.1049/ibc.2016.0029

[37] Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krúger. 2019. Exploring visual guidance in 360-degree videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. 1–12. https://doi.org/10.1145/3317697.3323350

[38] Hannah Syrett, Licia Calvi, and Marnix van Gisbergen. 2016. The oculus rift film experience: a case study on understanding films in a head mounted display. In *International Conference on Intelligent Technologies for Interactive Entertainment*. 197–208. https://doi.org/10.1007/978-3-319-49616-0_19

[39] Ryohei Tanaka, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2015. Attracting user's attention in spherical image by angular shift of virtual camera direction. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*. 61–64. https://doi.org/10.1145/2788940.2788951

[40] Michael Tomasello, Malinda Carpenter, and Ulf Liszkowski. 2007. A new look at infant pointing. *Child development* 78, 3 (2007), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

[41] Glenn Joseph Winters and Jichen Zhu. 2014. Guiding players through structural composition patterns in 3D adventure games. (2014).

[42] Jonathan Wonner, Jérôme Grosjean, Antonio Capobianco, and Dominique Bechmann. 2013. Bubble bee, an alternative to arrow for pointing out directions. In *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*. 97–100. https://doi.org/10.1145/2503713.2503753