

Director’s Cut - Analysis of Aspects of Interactive Storytelling for VR Films

Colm O Fearghail¹, Cagri Ozcinar¹, Sebastian Knorr^{1,2}, and Aljosa Smolic¹

¹ V-SENSE, School of Computer Science and Statistics, Trinity College Dublin, The University of Dublin, Ireland

{ofearghc,ozcinarc,smolica}@scss.tcd.ie

² Communication Systems Group, Technical University of Berlin, Germany
knorr@nue.tu-berlin.de

Abstract. To explore methods that are currently used by professional virtual reality (VR) filmmakers to tell their stories and guide users, we analyze how end-users view 360° video in the presence of directional cues and evaluate if they are able to follow the actual story of narrative 360° films. In this context, we first collected data from five professional VR filmmakers. The data contains eight 360° videos, the directors cut, which is the intended viewing direction of the director, plot points and directional cues used for user guidance. Then, we performed a subjective experiment with 20 test subjects viewing the videos while their head orientation was recorded. Finally, we present and discuss the experimental results and show, among others, that visual discomfort and disorientation on part of the viewer not only lessen the immersive quality of the films but also cause difficulties in the viewer gaining a full understanding of the narrative that the director wished them to view.

Keywords: 360° film · storytelling · director’s cut · virtual reality.

1 Introduction

The ability to create a visual narrative in a film has seen over a century of professional experimentation and developments. In cinema, the development of a visual narrative can be seen with the start of continuity editing, where the film is cut from one scene to another to tell the story [22, 23] effectively. This allows for the viewer to create a mental model of the scene and the position of the characters and objects in it, which allows the viewer to orientate themselves within the scene as the camera moves to different locations after cuts happen.

There are a number of methods used by directors in order to direct viewers attention [10]. The effectiveness of these can be seen in traditional cinema in the ‘tyranny of film’ effect that a Hollywood style of film-making has [15]. In 360° film-making, however, these conventions must be adapted as this new format has intrinsic features that differ from traditional cinema, such as the viewer being free to explore the entire scene. This also causes difficulties in cutting from scene to scene as the director cannot be sure of where a person may be looking in the 360° film, *i.e.*, it puts an increased importance on the ability of the director to guide attention.

Some of the factors being used to guide the viewers’ attention can be categorized into the directional cues: sound, environment and motion/action. Motion or action can either be present in the scene or due to the motion of the camera itself. Actors can be used to direct attention by the viewer matching their eyeline or by directly addressing the camera.

Given the rapid pace of development, it is crucial that filmmakers in the medium understand how the use of the techniques that they are using in a 360° format affects the viewers ability to follow and enjoy a narrative. Hence, the motivation of this paper is to study visual attention of users in the presence of directional cues within professionally produced 360° films. In this context, we first collected data from five professional virtual reality (VR) filmmakers. The data contains eight 360° videos, the director’s cut, which is the intended viewing direction of the director, plot points and directional cues used for user guidance (see Section 3.1 for details). Then, we performed an extensive experiment with 20 test subjects viewing the videos while their head orientation (*i.e.*, the viewing direction) was recorded. During and after the experiment, the participants were asked to answer general and video related questions (see Section 3.2 for details).

Finally, we present and discuss the experimental results in Section 4 by comparing the director’s cut with the users’ viewing direction and by evaluating the users’ answers to the questionnaires. Our findings show, among others, that adapting directional cues from traditional filmmaking seems to work well to attract users’ attention but the potential for visual discomfort must be considered alongside managing the orientation of the viewer to ensure an immersive experience. The entire dataset is publicly available with [12]³, where a new scan-path similarity metric and its visualization is presented.

2 Related Work

Four techniques that have traditionally formed the ‘tools’ that filmmakers rely on to tell their stories are cinematography, mise-en-scene, sound, and editing [24]. The expansion of these tools into VR, however, requires each to be re-evaluated as the viewer is free to look in any direction of the 360° film without the direct control of the filmmaker.

One of the most central ideas to the notion that continuity-led film grammars [3] are also applicable to cinematic VR is the ability of the director to predict and indirectly control the user’s viewport [16]. Serrano *et al.* [21] investigated continuity editing in VR video in the context of segmentation theory [13]. Their findings include that continuity of action across cuts by aligning the regions of interest between them is best suited to fast-paced action while misaligning these regions of interest or action discontinuity between cuts leads to more exploratory behavior from the viewer. In addition, a survey was carried out in [11] which aimed to measure the effect of cut frequency on viewers disorientation and their ability to follow a story. Their findings suggested that if the point of interest remains consistent across cuts, a high frequency does not increase disorientation or affect the ability to follow the story.

³ <https://v-sense.scss.tcd.ie/?p=2477>

Video	Content Description	Resolution	FPS	Duration
<i>Help</i> [9]	Science fiction film: alien destroys buildings and objects; slow moving camera.	3840×2160	30	1m
<i>360Partnership</i>	Documentary: urban Indian quarters and schools; camera mostly static with long shots.	3840×1080	30	6m17s
<i>Cineworld</i>	Commercial: dark interior with forced viewer attention by use of graphic arrows on screen; moving camera.	2560×1280	30	1m
<i>DB</i>	Commercial: bright lit interior and exterior scenes; slow paced moving camera.	4096×1024	30	3m58s
<i>Jaunt</i>	Commercial: scene of a parties interior. Actor addresses camera. Slow moving camera.	2304×1152	60	2m52s
<i>Smart</i>	Commercial: camera point of view inside moving car; fast movement outside of car.	2880×1440	60	2m7s
<i>Luther</i>	Tourism: various German interior and exterior sites; high amount of cuts; camera mostly static.	4096×2048	30	4m25s
<i>Vaude</i>	Commercial: scenic mountain exteriors and factory floor interior; slow moving camera.	4096×2048	30	2m25s
<i>War</i>	Education: exterior trenches in World War 1; mostly static camera.	4096×1152	25	3m25s

Table 1: Description of the dataset. The *Help* video is the training video.

To direct the viewer in a 360° narrative short, Nielsen *et al.* [17] investigated two methods; one where the orientation of the virtual body was faced in the region of interest, the other where the viewers’ attention was guided by the use of implicit diegetic guidance, in this case, a firefly. They found that the viewers preferred the firefly method of guiding attention and that forcing the viewer’s attention by orientating the virtual body increased visual discomfort. A similar approach to non-narrative 360° videos can be found in [14]. Blur was also evaluated as a method to direct the viewer within a virtual environment in [7] and a 360° video in [4].

Padmaneban *et al.* [19] introduced a motion sickness predictor for stereoscopic 360° video based on a machine learning approach. Their findings show that conflict in motion and not the presence of motion itself cause sickness, if users were allowed to freely move their heads with the virtual scene. Finally, Pavel *et al.* [20] developed a 360° video player with two features; a viewport-orientated technique and active reorientation. Viewport-orientated techniques reorient the shot at each cut so that an essential content lies in the viewer’s field of view. Active reorientation is performed by the viewer pressing a button to reorient the shot to the important content immediately. Finally, an analytics tool was developed for 360° video in [2] that allows to select areas in the scenes that were key to the story.

3 Methodology

For our studies, we used a dataset of eight monocular 360° videos for testing. The dataset has a wide range of content types including documentary, advertisement, tourism, and education. Each 360° video is in the equirectangular format with various resolutions and frame rates. Table 1 describes the characteristics of the 360° videos used in this work.

3.1 Collection of data from professional filmmakers

To collect relevant and useful information about the intended viewing direction of the filmmakers, the used directional cues and essential plot points for the given set of 360° videos, we first let five filmmakers manually create a scan-path, the so-called *director’s cut*, which represents the intended viewing direction, by setting position markers in the equirectangular format of their own videos. The setting of the position markers was done with The Foundry’s professional compositing software *Nuke*⁴ using the *Tracker* node. More details about the process can be found in [12].

⁴ <https://www.foundry.com/products/nuke>

Together with the director’s cut, the filmmakers were asked to provide additional information about plot points and directional cues used to attract attention of the viewers. In particular, the filmmakers were asked to provide the level of importance for the story (“plot point”, “essential plot point”, “not relevant”) and the intended viewing behavior (“maintain attention”, “free exploration”, “not relevant”) within certain frame ranges. Besides this, the following directional cues were requested:

1. Sound (“character/object”, “other sound cues”)
2. Environment (“brightness/contrast/color”, “visual effects elements”, “other environment cues”)
3. Motion/action (“camera motion”, “character/object motion”, “other motion cues”)

3.2 Collection of user data

Apparatus and test subjects. To collect users’ scan-paths and answers from the prepared questionnaires for a given set of 360° videos, the publicly available test-bed in [5, 18] is modified to allow video playback, continuously recorded participants’ head orientation with the current time-stamp and video name.

In parallel with the video, the audio data was sent to the integrated headphone of an head mounted display (HMD), which was the Oculus Rift consumer version in this work.

Subjective experiments were conducted with 20 participants (16 males and four females). Participants were aged between 22 to 46 with an average of 30.8 years. 50% of the participants had a medium familiarity with visual attention studies; 35% and 15% of the participants had no and high familiarity with visual attention studies respectively. Furthermore, eight participants wore glasses during the experiment, and all of the participants were screened and reported normal or corrected-to-normal visual acuity.

Questionnaires. In addition, we prepared a general questionnaire for the entire experiment to evaluate the subjective experience of the test subjects and a questionnaire for each test video to collect additional information for each test subject in order to trace back potential anomalies for the statistical evaluation of the scan-paths vs. directors’ cuts. The general questionnaires $\{Q_1^g \dots Q_7^g\}$ and the video related questionnaire $\{Q_1^v \dots Q_{15}^v\}$ are listed in Tables 2 and 3, respectively.

Test procedure. Subjective tests were performed as *task-free* viewing sessions, *i.e.*, each participant was asked to look naturally at each presented 360° video while seated in a freely rotatable chair. Each session, which lasted approximately 30 minutes, was split into a training and a test session. During the training session, one minute of the *Help* [9] 360° video was played to ensure a sense of familiarity with the viewing setup. Then, during the test session, the test videos were randomly displayed while the individual viewport trajectories (*i.e.*, the center location of the viewport) were recorded for each participant.

After each presented video, we inserted a short questionnaire period where the test subjects were asked to answer the questions in Table 3, while a mid-gray screen was displayed. Before playing the next 360° video, we reset the HMD sensor to return to the initial position. Finally, after all videos had been presented, the test subjects had to answer the general questions Q_1^g to Q_7^g as outlined in Table 2.

4 Analysis and Discussion

4.1 Comparison of scan-paths

In order to measure the similarity between the scan-paths, i.e., the director’s cut and the head orientations of the users, we calculated the angles between both for each frame of the video sequences.

Figure 1 shows both, the scan-paths together with the viewport area and the plot points. With respect to the latter, only five of the eight videos included plot points which are highlighted in red. The user’s scan-path is here the average across all test subjects and thus only gives an indication of the average viewing direction.

4.2 Evaluation of general questionnaire

The general questions and the number of participants’ answers to the point-scale questions, Q_1^g and Q_2^g , are listed in Table 2. With respect to Q_1^g , only two test subjects felt sick during the experiment. The rest either did not feel sick (thirteen participants) or were not sure if they felt sick (five participants). The majority of the participants felt medium (twelve participants) or highly (seven participants) engaged/immersed with the 360° content.

Question	Answer	Answer (# of participants)
Q_1^g : Did you feel sick?	A_1^g : “no” / “maybe” / “yes”	(13, 5, 2)
Q_2^g : How engaged/immersed do you feel with the 360 content in general?	A_2^g : “low” / “medium” / “high”	(1, 12, 7)
Q_3^g : Did any issues occur when wearing the HMDs?	A_3^g : free answer	
Q_4^g : Was there anything you disliked?	A_4^g : free answer	
Q_5^g : Which video did you like most?	A_5^g : free answer	
Q_6^g : What was most effective in attracting your attention for the entire dataset?	A_6^g : free answer	
Q_7^g : Any further comments you would like to share?	A_7^g : free answer	

Table 2: General questionnaire.

Furthermore, for question Q_3^g , “*Did any issues occur when wearing the HMD?*”, ten of the participants commented on the problem of low-resolution playback of 360° video as an essential issue. We observed that the content resolution has a significant impact on the quality of the immersive experience for VR. A similar result was also previously reported in the MPEG survey for VR [6]. The effect of motion also has a significant impact on the viewing experience. As observed in Q_4^g , five participants complained about the motion in the *Smart* video. Four participants (the highest number) liked the *Vaude* and *DB* videos the most, as observed by the answer to question Q_5^g . With respect to question Q_6^g , most of

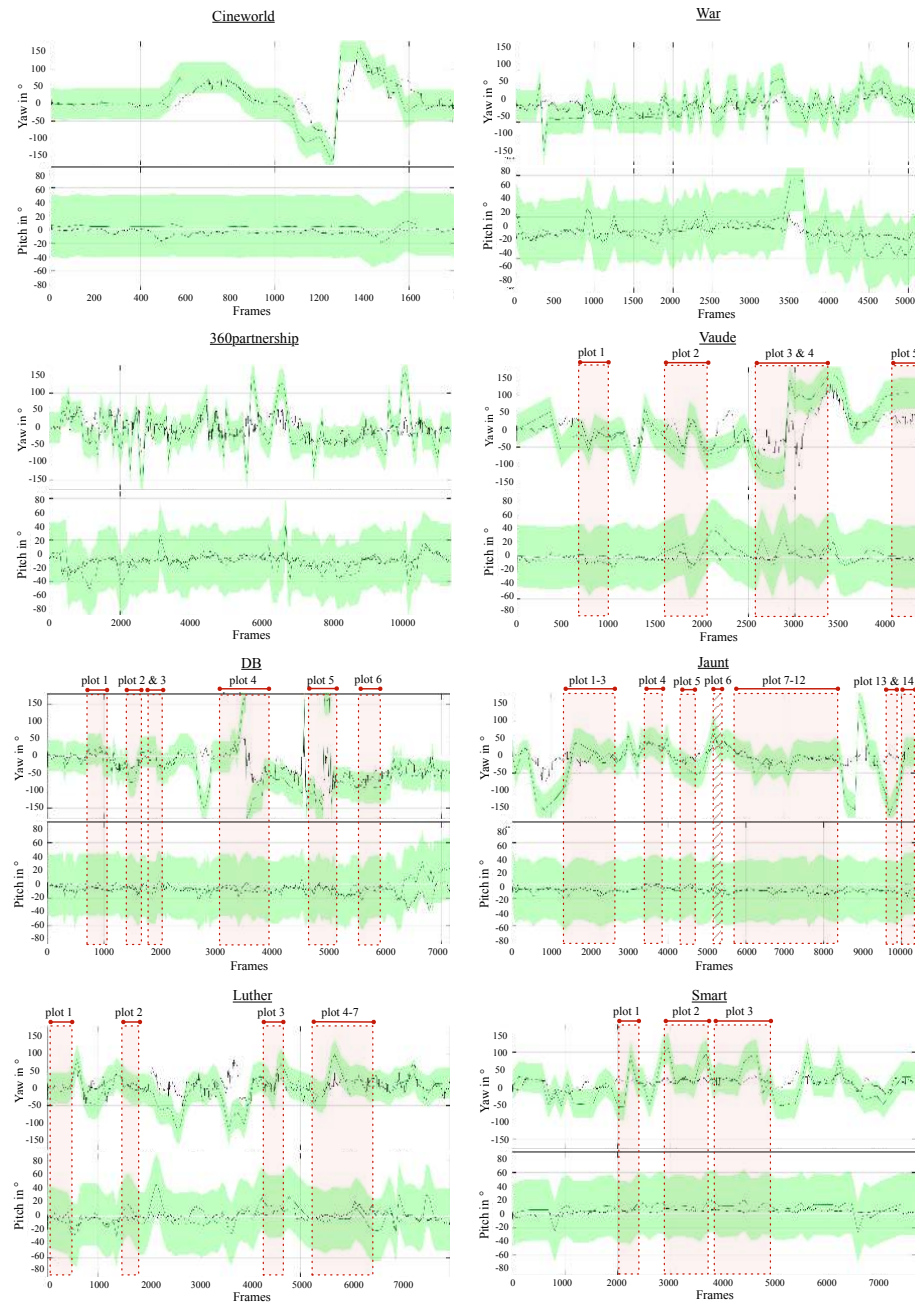


Fig. 1: Similarity measures: Director’s cut (dark green) with viewport area (light green), average user’s scan-path (black) and plot point areas (red).

the participants mentioned that the appearance of actors (six participants), au-

Question	Answer	Video	Answer (# of participants)
Q_1^v : Did you feel any discomfort?	A_1^v : “no”/“maybe”/“yes”	all videos	see Table 4
Q_2^v : Did you feel immersed in the environment/engaged with the video?	A_2^v : “no”/“maybe”/“yes”	all videos	see Table 4
Q_3^v : Did you feel any disorientation?	A_3^v : “no”/“maybe”/“yes”	all videos	see Table 4
Q_4^v : What was most effective in attracting your attention?	A_4^v : free answer	all videos	see text
Q_5^v : Any further comments you would like to share?	A_5^v : free answer	all videos	see text
Q_6^v : Which of the people passing by do you remember?	A_6^v : free answer	<i>Smart</i>	see text
Q_7^v : Did anything in particular stand out?	A_7^v : free answer	<i>Vaude</i>	see text
Q_8^v : Did you find the arrows helpful?	A_8^v : “no”/“maybe”/“yes”	<i>cineworld</i>	(10, 1, 9)
Q_9^v : Did you pay attention to the overlays?	A_9^v : “no”/“maybe”/“yes”	<i>Jaunt</i>	(0, 3, 17)
Q_{10}^v : Did you feel like you picked up the main information of the story?	A_{10}^v : “no”/“maybe”/“yes”	<i>DB</i>	(2, 6, 12)
Q_{11}^v : Did it give you a good sense of life in the trenches?	A_{11}^v : “no”/“maybe”/“yes”	<i>War</i>	(9, 2, 9)
Q_{12}^v : Did you see someone dying?	A_{12}^v : “no”/“maybe”/“yes”	<i>War</i>	(3, 3, 14)
Q_{13}^v : Did you feel a good sense of children’s life in India?	A_{13}^v : “no”/“maybe”/“yes”	<i>360partnership</i>	(1, 2, 17)
Q_{14}^v : Did you notice the overlays in the factory?	A_{14}^v : “no”/“maybe”/“yes”	<i>Vaude</i>	(5, 5, 10)
Q_{15}^v : Did you find the Luther character helpful?	A_{15}^v : “no”/“maybe”/“yes”	<i>Luther</i>	(5, 5, 10)

Table 3: Video related questions and number of participant answers (“no”, “maybe”, “yes”) to the point-scale answered questions.

dio (four participants), and overlays (four participants) were the most effective in attracting participant attention for the entire dataset. Finally, none of the participants commented on question Q_7^v .

4.3 Individual evaluation of videos and video related questionnaires

The video related questionnaire with its 15 questions is presented in Table 3, where questions Q_1^v to Q_5^v are questions which were asked for all videos, while Q_6^v to Q_{15}^v are video specific questions. Questions Q_1^v to Q_3^v and Q_8^v to Q_{15}^v are questions using a 3-point-scale with possible answers “no”, “maybe”, “yes”, respectively. The number of answers of the 20 test subjects and eight test videos for the questions Q_8^v to Q_{15}^v and Q_1^v to Q_3^v are reported in Tables 3 and 4, respectively. In the following, we evaluate the findings first for each video separately.

Video	Q_1^v	Q_2^v	Q_3^v
<i>360Partnership</i>	(16, 2, 2)	(2, 4, 14)	(19, 1, 0)
<i>Cineworld</i>	(12, 2, 6)	(13, 2, 5)	(9, 1, 10)
<i>DB</i>	(18, 1, 1)	(6, 1, 13)	(15, 3, 2)
<i>Jaunt</i>	(17, 0, 3)	(4, 2, 14)	(16, 2, 2)
<i>Smart</i>	(9, 5, 6)	(5, 1, 14)	(15, 1, 4)
<i>Luther</i>	(16, 1, 3)	(2, 5, 13)	(17, 1, 2)
<i>Vaude</i>	(15, 1, 4)	(2, 7, 11)	(14, 2, 4)
<i>War</i>	(13, 4, 3)	(3, 5, 12)	(12, 4, 4)

Table 4: Answers (“no”, “maybe”, “yes”) to the point-scale answered questions (in terms of no. of participants) for all test subjects.

360partnership. This video, shot in a documentary style, hoped to give the viewer a good sense of the environment and conditions of the children that the program helps with live in India. The director provided a scan-path as to how they would prefer the viewer to watch the video but did not consider any part particular to be essential enough to the videos understanding to be considered

a plot point. This is reflected in the strong variation of the yaw in Figure 1, which is an indication for exploratory behavior of the users. However, information was relayed through the use of audio commentary, so there were no plot points displayed visually within the scene. This video was also found to be the least disorientating of the videos, with only one participant answering ‘maybe’ (Q_3^v); the shots were long enough that the viewer could take their time in exploring the environment at a relaxed pace. From the received answers of the video questionnaires, when asked: “*what was most effective in attracting your attention?*” (Q_4^v), five participants found that text overlays were the most effective, the movement and direction of people was the second most common response with four participants. All participants except one felt that it gave them a good idea as to the challenges that the children shown in the video are required to face in daily life (Q_{13}^v), and four participants answered ‘maybe’ when asked if they found the video to be engaging (Q_2^v) while two participants answered ‘no’.

Cineworld. This video took the style of a first-person shooter that would be more commonly seen in a video game such as Doom [1] and applied it to a cinema interior. The area of interest was ringed by a circle which was very clearly illustrated the area where the director intended the viewer to look. This was further emphasized by the use of two large arrows to either side that pointed directly towards it. At certain times in the video, this circle turned at a sharp 90° . Here, the filmmaker intends that the viewer would turn likewise in the same direction. For instance, the first time that this occurs in the video is at frame 515. As it can be seen in Fig. 1, degrees of the yaw angle increase with this sharp turn.

The use of this as a mechanism caused discomfort, with twelve participants answering ‘yes’ to (Q_1^v), and disorientation, with ten participants answering ‘yes’ to (Q_3^v). The effect of the confusion experienced by the viewer made *Cineworld* also score lowest for engagement with only five participants feeling engaged by the video and two participants answering ‘maybe’ (Q_2^v).

Nine of the participants said that they found the arrows helpful (Q_8^v) in knowing where to look. From the Fig. 1 it is clear that there was a delay in the viewers orientating themselves in the direction the arrows indicated. Two responses made to (Q_5^v) help to explain this behavior, one being that the movement of the arrows was uncomfortable and the other that the arrows were too forceful in commanding attention.

DB. This video had six plot points and was a commercial in which the viewer could see the use of technology in transforming modern banking. The presence of the viewer was used in different ways in various scenes. At the start the viewer is directly addressed by the family’s matriarch. For the rest of the video the viewer has more of an observatory role. In later scenes the viewer is addressed directly again.

At plot number four, for example, the director used a number of graphics, in this case, furniture appearing in a room, behind two characters as they walked around a room at frame number 3,300. The response of the viewers can be seen

in Fig. 1. The mean shows that viewers followed it but most did not make a full bodily turn in the chair but rather followed it until the point that they could rotate their neck across to the other side in order to pick up the action.

DB had the lowest score on discomfort (Q_1^v), with 18 participants answering 'no'. In attracting attention (Q_4^v), the movement and placement of actors were the most effective for with six participants followed by graphics and overlays that were imposed into the scene with five participants mentioning them as the most successful in leading their attention (Q_5^v). Finally, the voice-over dialogue was present in only the left ear and was mentioned by four participants (Q_5^v).

Smart. The *Smart* video made use of three plot points. The level of importance did not differ too much between the plots points highlighted in the video. This video was more about the viewer experiencing a sense of fun and excitement as they were driven through the city.

Smart also had the highest score for discomfort (Q_1^v). The reason for this score might be a sharp turn at the end of the video and the fast motion of the car. The turn starts just after frame number 6,450 and in Fig. 1 it can be denoted by the viewers leaving the directors scan-path before rejoining it again once the turn was completed just after frame number 7,000. The car itself operated as an agency for the movement and gave the viewer a familiar setting in where they could anticipate how and where to look, and the path that the car would be following, along the road. The sharp turn at the end made a full 180° and was unexpected for viewers and, as mentioned by seven participants (Q_5^v), was a reason for discomfort. Because of these reasons, viewers experienced vection or perceived self-motion which lead to the discomfort reported in the experiment. *Smart* also has, together with *360Partnership* and *Jaunt*, the highest score for immersion (Q_2^v) with fourteen participants answering 'yes', which would lead to suspect that a familiar setting or agency can increase the immersion as long as this setting or agency operates in the manner that the viewer would expect it to. The video has just one single shot and no cuts, *i.e.*, it is more natural and thus may increase the feeling of being present. The direction perceived from the principal actor and the movement of the car and the direction that it was moving in, both mentioned by six participants, were the most frequent answers as to what attracted attention (Q_4^v). The band playing music, which was in the direction of the camera motion, was the most memorable of the people that the car passed (Q_6^v) with nine participants mentioning it.

Jaunt. There were 14 plot points in the video that had a high level of importance for the viewer to follow as can be seen in Fig. 1. The director used the principal actor along with graphical overlays to attract and direct attention within the video. The video consisted on just one scene without any cuts and this could be a reason for it scoring highly for engagement (Q_2^v), with fourteen participants answering that they did feel immersed in the environment; two answered maybe and four answered that they did not.

When asked about the overlays that were used in the video (Q_9^v) 17 participants answered 'yes', the rest answered 'maybe'. The audio was the highest

answer when it came to attracting attention (Q_4^v) mentioned by seven participants followed by the direction of the principal actor, mentioned by six participants. *Jaunt* had the second lowest score on discomfort (Q_1^v), with only three participants answering that they felt discomfort.

Vaude. Among the directorial cues received, five were cues that had a high level of importance for the viewer to follow and considered to be essential occurring to the director. The most of these plot points consisted of a dialogue delivered by the principal actor as she addressed the camera directly. Given the commercial nature of the video, the narrative, in this case, was the relating of information about the product, as per plot points 1 and 2, where the principle actor talked directly to the camera. During plot point 2, the use of overlays were again used, and when asked (Q_{14}^v) five participants did not notice them, and five participants answered ‘maybe’.

The direction of the principal actor was the most frequent answer when asked what device was the most effective in attracting the attention of the viewer (Q_4^v) with eight participants mentioning it. The line up of cuts in between shots noticeable for a number of viewers, with four participants mentioning it as an additional comment (Q_5^v), they found that the area of interest was not matched correctly across a cut, it caused them to have to find it again after a cut happened.

The main causes of discomfort (Q_1^v) were the vibrations when the camera was mounted on a bicycle with four participants mentioning in response to (Q_5^v), which might have had a bearing on people not noticing the Panda figure around frame 2,600 in Fig. 1, which was in the director’s cut.

Luther. The video had an animated character, a Playmobil character that took the appearance of *Luther*, imposed on a number of shots and across a few cuts. There was a mixed reaction to the use of this character which can be seen in response to (Q_{15}^v). For this question, five of the participants found that the use of this character distracted from their ability to freely explore the environment while others found it helped to orient themselves around the area of interest, *i.e.*, the character of *Luther*. One viewer’s response was that if he had lost track of the character, he would spend time looking for him while the scene changed which disorientated him even further.

Luther also had the highest number of shots and the shortest scene length. Four participants mentioned for (Q_5^v) that there was too much information as the scenes were perceived to be changed too quickly. Only two participants found the video to disorientate them (Q_3^v), and three participants found the video to cause discomfort (Q_1^v).

War. The *War* video, which was educational, was the second highest scoring video for disorientation (Q_3^v). In this video, two allied soldiers were shown in a trench and then a firefight was displayed at night.

Taking place in a nighttime environment, the most common response to (Q_4^v) was the bright lights that were used in the film with four participants mentioning

so that start at frame number 2,970. A flare used to attract the attention of the viewer upwards while the scene cut below at frame number 3,440 was mentioned by three participants to the same question and can be seen in Fig. 1 as a large increase in the pitch of the directors cut.

The dark environment alongside the hand-held movement of the camera in the later part of the video caused discomfort for three participants (Q_4^v).

4.4 Overall findings and discussion

From the data collected and the responses to the questionnaire, it would appear that viewers prefer to have their attention led rather than forced. This finding was also reported in [17]. The shot lengths for the videos that scored highest for engagement were longer than those that scored more lowly, which allowed the viewer time to freely explore the environment without having to worry about the shot changing before they had time to do so. Audio and the direction of the principle actor were the two most significant factors for the attracting of attention across all the videos. Another factor that had a significant influence on the engagement of the video was the orientation of the viewer. If the viewer becomes disorientated within the scene, they also become disorientated in the narrative the director is displaying. This also causes problems for viewer immersion as they are more worried about missing the area of interest than enjoying the video. One way that can happen is a bad match of action across the cuts. The disorientation can be emphasized even further if a cut happens when the viewer is already disorientated from a previous cut. Not only should action match across scenes but other factors such as scale should also be taken into account.

Motion/Action. The motion was used in various ways by the videos. *Smart* was most evident in the use of it as a device in order to transport the viewer through the narrative. However, this was also a conflicting cue with respect to action cues to the left and right of the camera path. There was camera movement in a number of the videos, and it was received with mixed reviews in terms of effectiveness based on the manner used and the personal preference of the viewer. One answer to (Q_4^v) made on the *Smart* was that it was faster than walking speed and this was the cause of the discomfort that the participant felt. One factor that did have a very noticeable impact when camera motion was used was how stable the camera was when the motion was taking place. In general, camera motion was accepted when it was clear to the viewer along what track that the motion would be taking place.

The use of actors in other to direct the viewer within the scene was used successfully in a number of the videos. *Vaude*, *BD*, *Jaunt* and *Smart* all used the principle character in order to direct the viewer. However, the interaction between the viewer and the animated character in *Luther* differs from one in which people are used. There are many advantages in the use of a person to direct attention across the scene, which is learned behavior from childhood to focus on what other people are looking at or what they have their attention

directed towards. It also gave the viewers a clear idea as to where to look, and in general, the principle actor was easy to find within the scene.

Environment. Environmental cues including visual effects were used by a number of videos also, most noticeably by *360partnership*, *DB*, *Vaude* and *Jaunt*. *Luther* and *Vaude* had a large number of scenic locations often dominated by a landmark building such as Wartburg castle, which had the effect of attracting the viewers' attention and often let the viewers explore freely the scenery. However, in *Vaude*, the environmental cues were also conflicting cues with respect to drawing attention to the actual product. Many of the videos used graphics in various ways to better illustrate information at various points. They also served as a method to guide attention, perhaps most effectively in *DB*. In general across the videos, the use of graphics clearly showed the viewer the area of interest in the scene that they were watching.

Sound. Even though sound is known to aid visual processing in VR [8], they did not solely form plot points provided by the directors of the films. Sound cues that were provided were often used in conjunction with visual ones. *Luther* at various times gave commands to the viewer such as at frame number 570 when the voice-over said "look around you" in order to encourage exploratory behavior from the viewer and later at frame number 4,430, more directly by telling the viewer to "take a look to the right". *Vaude* used audio in the form of dialogue from the principle actor to direct attention, such as at frame number 1,582 where she directly addresses the camera from the factory floor.

5 Conclusion

While traditional directing techniques can serve to lead viewer attention in 360° film, there are a number of differences required in the conceptual approach of their use. 360° film means moving from a window onto a world to being present within one. Rather than directing the viewer to conceptualize their environment through a series of images, the task is to orientate the viewer within one. This orientation is even more crucial when a cut is present, as the viewer is required to re-orientate themselves in the space of the new scene and disorientation will lessen the quality of the immersive experience. The nature of adapting these traditional directorial cues to 360° will require a directorial approach that moves away from using a time based sequence of images into one that makes use of the spatial nature of virtual reality. Further studies on this dataset including the introduction of a new metric for scan-path comparison were carried out in our paper [12] which offers an intuitive visualization for use in a post-production environment.

Acknowledgment

This publication has emanated from research conducted with the financial support of Science Foundation Ireland (SFI) under the Grant Number 15/RP/2776.

References

1. Doom. [PC CD-ROM] (December 10, 1993)
2. Bala, P., Dionisio, M., Nisi, V., Nunes, N.: IVRUX: A tool for analyzing immersive narratives in virtual reality. In: Nack, F., Gordon, A.S. (eds.) *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 3–11. Springer International Publishing, Cham (2016). https://doi.org/10.1007/978-3-319-48279-8_1
3. Bolle, R., Aloimonos, Y., Fermüller, C.: Toward motion picture grammars. In: Chin, R., Pong, T.C. (eds.) *Computer Vision — ACCV'98*. pp. 283–290. Springer Berlin Heidelberg, Berlin, Heidelberg (1997)
4. Danieau, F., Guillo, A., Dore, R.: Attention guidance for immersive video content in head-mounted displays. In: *2017 IEEE Virtual Reality (VR)*. pp. 205–206. IEEE, Los Angeles, CA, USA (March 2017). <https://doi.org/10.1109/VR.2017.7892248>
5. De Abreu, A., Ozcinar, C., Smolic, A.: Look around you: Saliency maps for omnidirectional images in VR applications. In: *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*. pp. 1–6. IEEE, Erfurt, Germany (May 2017). <https://doi.org/10.1109/QoMEX.2017.7965634>
6. the MPEG Virtual Reality Ad-hoc Group: Summary of survey on virtual reality. Tech. Rep. N16542, JTC1/SC29/WG11, ISO/IEC, Chengdu, CN (Oct 2016)
7. Hillaire, S., Lecuyer, A., Cozot, R., Casiez, G.: Depth-of-field blur effects for first-person navigation in virtual environments. *IEEE Computer Graphics and Applications* **28**(6), 47–55 (Nov 2008). <https://doi.org/10.1109/MCG.2008.113>
8. Hoeg, E.R., Gerry, L.J., Thomsen, L., Nilsson, N.C., Serafin, S.: Binaural sound reduces reaction time in a virtual reality search task. In: *2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. pp. 1–4 (March 2017). <https://doi.org/10.1109/SIVE.2017.7901610>
9. Justin (Director), L.: *Help* (2015). <http://www.imdb.com/title/tt4794550/> (2015)
10. Katz, S.D.: *Film directing shot by shot: visualizing from concept to screen*. Gulf Professional Publishing (1991)
11. Kjær, T., Lillelund, C.B., Moth-Poulsen, M., Nilsson, N.C., Nordahl, R., Serafin, S.: Can you cut it?: an exploration of the effects of editing in cinematic virtual reality. In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. p. 4. ACM, Gothenburg, Sweden (Nov 2017)
12. Knorr, S., Ozcinar, C., O Fearghail, C., Smolic, A.: Director's cut - a combined dataset for visual attention analysis in cinematic VR content. In: *The 15th ACM SIGGRAPH European Conference on Visual Media Production*. London, UK (Dec 2018)
13. Kurby, C.A., Zacks, J.M.: Segmentation in the perception and memory of events. *Trends in cognitive sciences* **12**(2), 72–79 (2008)
14. Lin, Y.C., Chang, Y.J., Hu, H.N., Cheng, H.T., Huang, C.W., Sun, M.: Tell me where to look: Investigating ways for assisting focus in 360 video. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. pp. 2535–2545. ACM, Denver, Colorado, USA (May 2017)
15. Loschky, L.C., Larson, A.M., Magliano, J.P., Smith, T.J.: What would jaws do? the tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS one* **10**(11), e0142474 (2015)
16. Mateer, J.: Directing for cinematic virtual reality: how traditional film director's craft applies to immersive environments and notions of presence. *Journal of Media Practice (author-produced version)* **18**(1), 14–25 (5 2017). <https://doi.org/10.1080/14682753.2017.1305838>

17. Nielsen, L.T., Møller, M.B., Hartmeyer, S.D., Ljung, T.C.M., Nilsson, N.C., Nordahl, R., Serafin, S.: Missing The Point: An Exploration of How to Guide Users' Attention During Cinematic Virtual Reality. In: VRST '16 Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technolog. pp. 229–232. Munich, Germany (Nov 2016). <https://doi.org/10.1145/2993369.2993405>
18. Ozcinar, C., Smolic, A.: Visual attention in omnidirectional video for virtual reality applications. In: 10th International Conference on Quality of Multimedia Experience (QoMEX). Sardinia, Italy (May 2018)
19. Padmanaban, N., Ruban, T., Sitzmann, V., Norcia, A.M., Wetzstein, G.: Towards a machine-learning approach for sickness prediction in 360 stereoscopic videos. *IEEE transactions on visualization and computer graphics* **24**(4), 1594–1603 (2018)
20. Pavel, A., Hartmann, B., Agrawala, M.: Shot orientation controls for interactive cinematography with 360 video. In: Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology. pp. 289–297. ACM (2017)
21. Serrano, A., Sitzmann, V., Ruiz-Borau, J., Wetzstein, G., Gutierrez, D., Masia, B.: Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics* **36**(4), 47:1–47:12 (Jul 2017). <https://doi.org/10.1145/3072959.3073668>
22. Smith, T.J.: An attentional theory of continuity editing (2006)
23. Smith, T.J., Levin, D., Cutting, J.E.: A window on reality: Perceiving edited moving images. *Current Directions in Psychological Science* **21**(2), 107–113 (2012)
24. Vosmeer, M., Schouten, B.: Project orpheus a research study into 360° cinematic VR. In: Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video. pp. 85–90. TVX '17, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3077548.3077559>, <http://doi.acm.org/10.1145/3077548.3077559>