DIRECTING ATTENTION IN 360-DEGREE VIDEO

Alia Sheikh, Andy Brown, Zillah Watson and Michael Evans

BBC Research and Development, UK

ABSTRACT

360° video and Virtual Reality are powerful techniques for giving viewers a sense of *'Being There'* [1], and are becoming increasingly popular. However, giving the viewer the freedom to look around also results in a reduced ability for filmmakers to direct the viewer's attention, a serious **impediment to successfully telling a story within a 360° environment. We have created a number of 360° clips, filmed in such a way as to demonstrate and test several unobtrusive techniques for directing a** viewer's attention **within a 360° panorama. We have evaluated these techniques in a user study in which participants viewed these clips using a head-mounted display. Qualitative and quantitative data from these tests have been analysed to evaluate the effectiveness of the different attention-directing techniques. Qualitative data was also captured to explore the** effect of the camera being addressed directly, and the viewers' responses **to action occurring at a range of distances.**

INTRODUCTION: VISUAL ATTENTION AND CINEMATOGRAPHY

360° video is a special case of virtual reality (VR) in which the audience views a sphere (or near-sphere) of video centred on a single position. 360° formats offer the filmmaker both opportunities and challenges. Unconstrained by a prescribed view, the viewer experiences a video environment in a way that **correlates more closely to real life. However, this comes at the cost of limiting the set of techniques open to the director: the use of different camera angles and the ability to cut between them, differential focus and moving camera techniques are all constrained.** In conventional TV and film, such techniques can be used by the filmmaker to take the viewer on a specific path through a narrative, ensuring the viewer's attention remains on the elements considered important to the story. In 360° presentation, however, the use of such techniques could have a negative impact on the user's experience, reducing their feeling of control and potentially inducing discomfort. **Since some of the key benefits of 360° video are a result of the viewer's control over their own gaze, the filmmaker must allow the viewer to retain that agency and direct gaze using subtler, unobtrusive techniques.**

As 360° content is rapidly evolving, directors are developing a new grammar of filmmaking. In addition to accepting a lower level of control over the audience experience, the basic methods for directing attention are **starting to be being explored, for example by using movement, sound and lighting cues. We seek to understand how effective some of these techniques are through more rigorous audience testing.**

RESEARCH QUESTIONS

This paper describes the development and presentation to viewers of some specially produced 360° video material, created to allow us to probe specific directorial mechanisms for directing visual attention in 360° footage, as well as some closely related questions about the subjective experience of this kind of video. Our key research questions are:

1. What attracts attention, what refocuses attention, and what techniques can a filmmaker use to direct the attention of a viewer?
2. How does the distance at which action occurs impact the experience of the viewer?

Are presence, immersion and enjoyment affected by characters in the content addressing the camera directly? We filmed a number of one-take single-shot setups with actors. Each setup was designed to test a specific attention directing technique, or answer a particular research question.

Clips A_1, A_2, A_3 and A_4, were designed to explore directing attention, and were filmed indoors in a large gymnasium. Each clip begins with a clear element of interest, and then uses different methods to try to direct the viewer's attention to a new element of interest introduced later in the shot. Each clip starts with two actors, clearly in view, having a conversation; this was the only action in the scene early on, and lasted at least 45 seconds before any other cues were introduced. Thus, we could be reasonably confident that the viewers would have the opportunity to familiarise themselves with the environment, and that their attention would be drawn to (and ideally retained by) the conversation. Another actor was introduced into an empty portion of the scene (behind the viewer if they were looking at the first two actors), and each clip used a different combination of visual and, in some cases, audio cues to direct the viewer's attention to the newcomer.

Clips D_{2m}, D_{3m}, D_{4m} depicted two actors practising a stagefight. As we wanted to assess the impact of distance on the viewer's comfort-level with the scene, we asked the actors to repeat this sequence at a distance of 2m, 3m and 4m from the camera. Previous studies show that people's comfort at different levels of interpersonal distance is highly context-driven [4], so in this test we filmed an activity that we anticipated people would react consistently to. Testing the reactions of viewers to the presence of people who are being active, but not threatening, allowed us to test for a more physical, instinctive response, as opposed to a more considered one.

In contrast to the other clips, the one used to test presence (P) featured actors explicitly acknowledging the presence of the viewer, with both actors appealing to the viewer to support their side of an argument. This clip is very similar to a clip that was used to acclimatise participants to 360° video, in which the same actors have an argument in the same location, but do not acknowledge the viewer. This allows this direct style of viewer engagement to be investigated.

RESEARCH METHODOLOGY

Participants were recruited through an external agency and a local college. There were 26 participants in total, all of whom took part in the 'directing attention' part of the study. 17 of the participants also viewed clips designed to understand more subjective aspects of the experience, and their impressions were captured using a questionnaire and a semi-structured interview. In each case, a participant's session lasted under an hour. Video clips were between 20 and 180s in duration and consisted of monoscopic video and stereo

audio. Participants viewed the content on a head-mounted display (Oculus Rift) whilst software continuously logged their head orientation within the scene; audio was delivered through a pair of Beyerdynamic DT 770 Pro headphones. Participants were standing while viewing the clips, to match (approximately) the camera height during filming, whilst a researcher supervised for physical safety.

Participants viewed two initial clips for acclimatisation purposes. For virtually all participants this was their first experience of 360° video on a head-mounted display, and certainly under controlled conditions. An initial acclimatisation piece shot in a busy street during the Edinburgh Fringe Festival familiarised them with the style of presentation and their ability to change their orientation to look all around them. A second acclimatisation clip allowed participants to become familiar with the style, actors and location used in a number of the subsequent clips.

Figure 1: A a participant viewing the Royal Mile acclimatisation scene in the lab study.

After acclimatisation, participants were shown one of the directing attention clips (A_1, A_2, A_3 and A_4), followed by the presence clip (P) and two of the three distance clips (D_{2m}, D_{3m}, D_{4m}). In all cases, head orientation was logged and participants asked for general feedback; for some clips, participants were also asked to rate their levels of enjoyment and sense of immersion, and their ability to follow the action, using a 5-point Likert-style scale.

DIRECTING ATTENTION RESULTS

Each of the clips involved two main characters, who were having a conversation on a bench, the target, who appeared on the opposite side of the viewer to the main characters, and a bystander. The cues used in each clip to direct attention towards the target are given in Table 1, while Figure 2 illustrates the scene.

Clip	Summary	Cue 1	Cue 2	Cue 3
A_1	Motion across main characters	Bystander walks to target		
A_2	Motion across main characters with gestural cue	Bystander walks to target, waving		
A_3	Motion across main characters with audio and gestural cues	Target shouts “Alia”	Bystander responds with wave and “Hi”	Bystander walks to target
A_4	Motion of a main character following gestural and audio cues	Main character looks at target	Main characters talk about target	Main character walks to target

Table 1: Cues used to direct attention in clips A_1-A_4.

Figure 2: The scene for the directing attention clips A_1-A_4. The main characters are seated in the centre, the target is on the far left; the bystander has just left her starting position below the clock.

Figure 3 gives an overview of the results for each clip. These plots show the percentage of people who had seen the target over the time since the first cue. Comparing clips A_1, A_2 and A_3, it can be seen that whilst motion cues alone have some effect, the addition of audio and gestural cues increases the effectiveness with which we can direct attention.

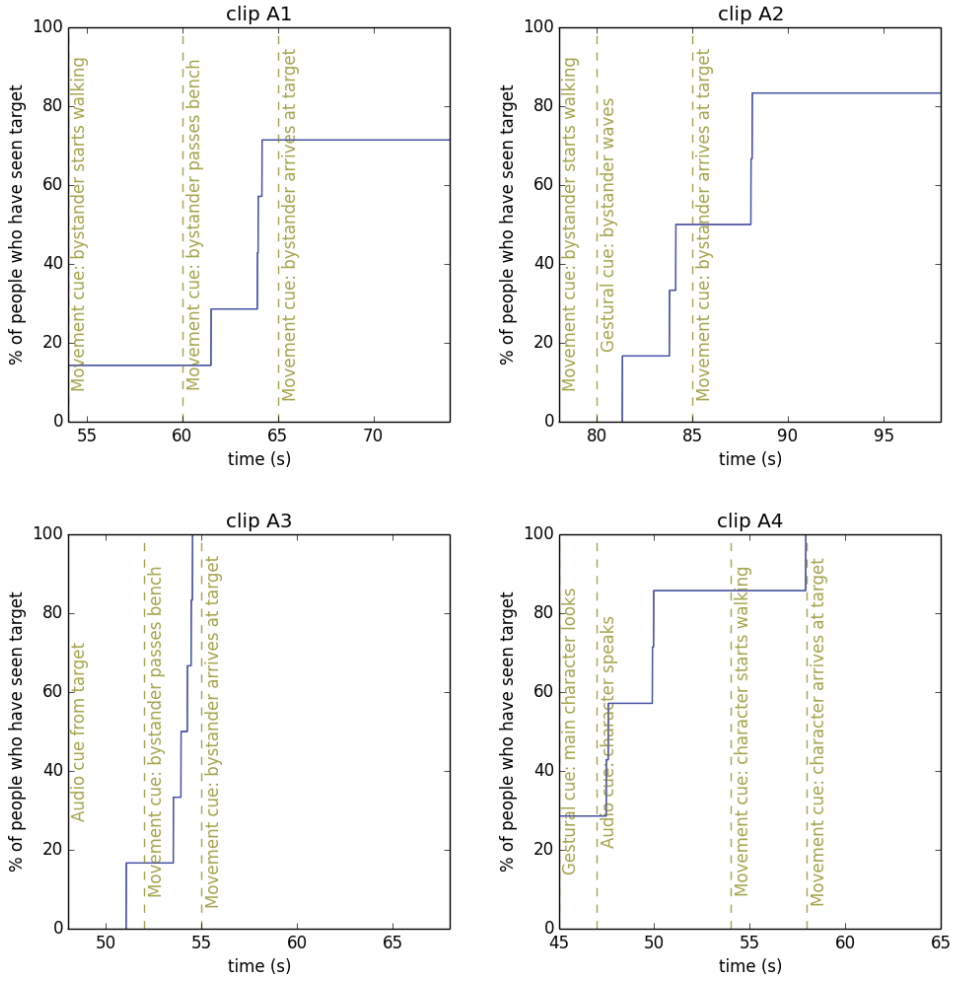

Figure 3: Plots illustrating the effectiveness of the directing attention cues. These show the cumulative percentage of participants who had seen the target at any time since the first cue (the times of cues are shown with vertical dashed lines).

The bystander walking past was, in isolation, moderately effective (clip A_1), although 2 of the 7 participants did not see the target. In both cases, the cue was seen, but not followed to the target. Supplementing this with a second visual cue – the wave from the bystander (clip A_2) – alerted the viewers that the bystander was walking towards someone. This was more effective, with only 1 participant from 7 missing the target – this person did not see the wave and did not follow the bystander's motion. Adding an audio cue from both target and bystander meant that some cue was perceived no matter where the viewer was looking, and was much more effective – all viewers saw the target, and attention shifted very quickly (within 7s).

Clip A_4 used different techniques, involving one of the protagonists making reference to, and ultimately moving towards the target. In this case, it can be seen that of the 7 viewers, 2 people were looking at the target before the first cue, 4 responded to the second cue (the mention of the target), while the last person followed the protagonist walking across to the target. It would be interesting to evaluate in more detail whether a cue from a protagonist is more effective than a cue from an inactive bystander.

SUBJECTIVE EXPERIENCE RESULTS

Subjective data were collected from participants for several of the clips. These allow us to explore preferences, and the reasons behind them. In particular, we are interested in how participants felt about watching the fight scene at two distances, which distance they preferred, and why, and how participants felt about the actors addressing the camera directly.

Distance

Participants showed a clear preference for the fight training taking place at 3m. Both the 2m and 3m distances were preferred when compared with 4m, but when 2m and 3m were compared directly, there was a clear preference for 3m (Table 2).

Participant's ratings showed that the distance did not impact their ability to follow the action, but it did affect their enjoyment of the clip, and their sense of immersion. The enjoyment ratings matched the preferences to show that 4m was too distant, but there was no clear difference in the ratings given to the 2m and 3m.

Comparison	Closer	Further
2m vs. 4m	4	1
3m vs. 4m	4	1
2m. vs 3m	1	5

Table 2: Participant preferences for each pair of distance clips. This reports the number of participants preferring either the closer or the farther of each pair.

The interviews revealed a number of reasons for this preference. When taking place at 4m, the action was felt to be too distant.

'It felt like you were watching something across the street' [P13, 4m]

In contrast, the same participant found the 2m clip much more realistic and immersive:

'It didn't feel like I was watching TV or anything, it felt like I was actually there' [P13, 2m].

Other people, however felt that at 2m the action was too close, in the personal space of the viewer:

[It was] 'very, very close to where I was… that wouldn't usually be happening that close to somebody' [P01, 2m]

'When they stood too close, it felt like they were more in your personal space' [P12, 2m]

The Impact of Acknowledging the Viewer Participants were prompted for their thoughts on the presence clip (P) by being asked to rate their experience using the (informally worded) question: *"How immersed did you feel; how much did it feel like you were there? 1 represents 'not at all'; 5 'very much'"*. Comparing the ratings given to the acclimatisation clip and the Presence clip, the latter receives a higher proportion of ratings of 4 or 5 (94% vs. 47%). The enjoyment ratings followed a similar pattern: for the acclimatisation clip, 56% of participants gave a rating of 4, and 12% of 5; for the Presence clip these values were 50% and 35% respectively. Other than the actors addressing the camera, these clips were similar: both were filmed in the same location, and involved the same actors in an argument. In both cases the actors spent some time moving around the camera. The major differences were that P1 was longer (3 minutes rather than 1.5 minutes), and was always shown second.

While the ratings give an indication of the impact of this technique, the qualitative feedback was much richer. The overwhelming reaction to the actors talking to and gesturing towards the camera was positive:

'After a while it felt like I was just standing talking to two friends… it felt like real life to me, not just a staged environment' [P13]

'It kind of felt like you were actually involved in the conversation… I thought it was good… it makes you feel like you're there.' [P4]

Interestingly, the sense of immersion was significantly affected by a minor gesture by one of the actors in the D_{4m} distance clip : as the fight finishes, the female actor points to where she is going next — this happened to be close to the camera, and several participants commented that she pointed at them, and that this made them feel more part of the scene. Participants also commented favourably on being looked at by passers by in the Royal Mile acclimatisation clip.

DISCUSSION

What attracts attention, what refocuses attention (how do viewers distribute their visual attention), and what techniques can we use to direct the attention of a viewer?

Four techniques for directing viewer's attention to one portion of a single shot scene were evaluated. The most effective used both audio and visual cues from the target area and the part of the scene where the viewers were assumed to be looking. The most unobtrusive technique was using a bystander to walk across the action towards the target; this was effective for 5 of the 7 participants who watched the clip. Audio cues have the advantage that no assumption is made about the viewer's focus of attention at the time of the cue. Even without fully spatialised audio, the use of sound also alerts the viewer that there is something to see; with the visual cue alone, participants sometimes followed the

cue, but not as far as the target. When both audio and visual cues were used, all participants saw the target.

How does the distance at which action occurs affect the immersion or enjoyment of the viewer?

In the context of viewing the fight training scene in the distance clips D_{2m}, D_{3m} and D_{4m}, there was a general consensus among participants that when the action occurred at 4m it felt too distant, but 2m felt unnaturally close. 3m provided a good balance of being close enough to see clearly and provide a sense of immersion, but far enough that it wasn't happening in their private space. The evidence in the literature from virtual reality [2,3] suggests that "the response to invasion of virtual personal space shows the same trend as the response to the same stimuli in a live setting" [2], so it is interesting that this experiment indicates that 360° video may have similar effects. In addition, it is known that people under threat maintain a greater personal distance [4], so the comfort felt at a given distance will vary with context (the 1-3m range in the close setting comes into the viewer's personal distance zone [5], where the practice fighting involving large body movements could feel threatening). Thus, the results found in this experiment may reflect the balance of the viewer's desire to maintain a 'safe' distance from the action with their ability to have a good view of the action, and feel part of it.

Furthermore, it is known that depth perception is distorted when viewing static camera monoscopic 360° video, with objects appearing further away than they are, an effect that scales approximately linearly with actual distance [6]. The material used in this experiment was filmed using a static camera and in an environment (a relatively empty rooftop courtyard) with few depth cues (e.g., occlusion). These combine to create an apparent 'collapse in perspective' where, as the actors move further from the camera they become harder to separate from the background, and thus feel artificially further away. Compromised depth perception is a fundamental challenge for monoscopic 360° video that needs to be considered by directors, but it can be mitigated to some extent by set design or camera motion.

Finally, it should be noted that it will not always be the aim of the filmmaker to keep the viewer at a comfortable distance, but understanding what this is will allow them to manipulate this variable in order to achieve a desired response.

Are presence, immersion and enjoyment affected by characters in the content addressing the camera directly?

There was a strong feeling that participants felt more immersed in the content, and enjoyed it more when they were acknowledged by characters in the scene. The technique of having the actors directly reference the camera as another character was effective, but even pointing at or making eye-contact with the camera without verbal reference had a notable effect.

CONCLUSIONS AND FURTHER WORK

These experiments have demonstrated the effectiveness of various visual and audio cues for directing the viewer's attention within a 360° scene. We found that the combination of audio and visual cues is more powerful than visual cues alone, this is mainly because audio cues are less dependent on the focus of attention at the time of the cue. While they cannot guarantee that attention will be given to the desired part of the scene, such cues can be used by filmmakers to guide the audience through a narrative. We also found that

participants reacted positively to being directly addressed or acknowledged by characters within the scene. The response of participants to fight practice occurring at different distances in 360° video matched what would be expected in real-world scenarios and virtual environments. Thus, we anticipate that camera distance can be used by directors to induce particular emotions in a way that maps real life. Further research is required to understand how distance is perceived in 360° video, given any specific projection and mapping, and also to understand how other scenarios would be experienced.

We believe that developing an understanding of the user experience of 360° video, through studies such as this, can inform filmmakers and accelerate the development of the craft. These experiments represent only a start, however. There are other techniques that are being used to direct attention, which we would like to explore further. For example: Which lighting techniques are most effective? When rendering 360° **video in a virtual environment, how can additional objects be composited on the scene to guide the viewer? Spatial audio is recognised as enabling a richer experience [7]; given that stereo was effective, how much better is fully spatialised sound for directing attention?** Other techniques using choreography are also possible; we would like to explore how storytellers versed in the theatre space approach and solve these problems – the theatre is, after all, a single set with a fixed audience viewpoint. What additional blocking techniques used in the theatre can be applied to 360° video?

Furthermore, this is early work using limited numbers of participants: it will be necessary to explore these questions further, moving beyond bespoke test material, to understand how they work in longer sequences with a more defined/driven narrative. This will include researching methods to successfully edit sequences together within a narrative, in ways that feel natural and un-prescribed. Retaining the viewer agency afforded by 360° video is crucial to the experience, so we need a suite of techniques that will allow us to move the viewer through the story without them being aware of the guiding hand of the director.

REFERENCES

1. **Cummings, J. J. & Bailenson, J. N. (In Press). How Immersive Is Enough? A meta-analysis of the effect of immersive technology on user presence. Media Psychology.**

2. **Wilcox, L.M., Allison, S.E., Elfassy, S. & Grelik, C. Personal Space in Virtual Reality. ACM Transactions on Applied Perception 3,4 (2006), 421-428. DOI: http://dx.doi.org/10.1145/1190036.1190041**

3. **Bailenson, J.N., Blascovich, J., Beall, A.C. & Loomis, J.M. Interpersonal Distance in Immersive Virtual Environments. Pers Soc Psychol Bul. 29,7 (2003), 819-833. DOI: http://dx.doi.org/10.1177/0146167203253270** Uzzell D. & Horne, N. The influence of biological sex, sexuality and gender role on interpersonal distance. *British Journal of Social Psychology* 45 (2006). 579-597.

4. **Hall, E.T. A system for the notation of proxemic behavior. American Anthropologist 65,5 (1963). 1003–1026.**

5. Kaufman, L. Sight and Mind: An Introduction to Visual Perception. *OU. Press* (1974).

6. Skalski, P., & Whitbred, R. Image versus Sound: A Comparison of Formal Feature Effects on Presence and Video Game Enjoyment. *Psychology Journal*, 8(1) (2010), 67-84.

ACKNOWLEDGEMENTS

The authors would like to thank Mike Armstrong, Angela McArthur and Vanessa Pope for their assistance with background research, Lianne Kerlin, Mia Roscoe and Maxine Glancy for helping to run the study, and the participants for their time and feedback.