



*Citation for published version:*

Farmer, H, Bevan, C, Green, D, Rose, M, Cater, K & Stanton Fraser, D 2020, 'Did you see what I saw?: Comparing attentional synchrony during 360° video viewing in head mounted display and tablets', *Journal of Experimental Psychology: Applied*, vol. 0, no. 0, pp. 1-15. <https://doi.org/10.1037/xap0000332>

*DOI:*

[10.1037/xap0000332](https://doi.org/10.1037/xap0000332)

*Publication date:*

2020

*Document Version*

Peer reviewed version

[Link to publication](#)

©American Psychological Association, [2020]. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: <http://dx.doi.org/10.1037/xap0000332>

**University of Bath**

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **Did you see what I saw?: Comparing**  
2 **attentional synchrony during 360° video**  
3 **viewing in head mounted display and**  
4 **tablets**

5 Farmer, H<sup>1,2,3\*</sup>, Bevan, C.<sup>4</sup>, Green, D.<sup>5</sup>, Rose, M.<sup>5</sup>, Cater, K.<sup>4</sup>, Stanton Fraser, D.<sup>1</sup>

6 <sup>1</sup> University of Bath, Bath, UK

7 <sup>2</sup> University of Greenwich, London, UK

8 <sup>3</sup> University College London, London, UK

9 <sup>4</sup> University of Bristol, Bristol, UK

10 <sup>5</sup> University of the West of England, Bristol, UK

11 \*Corresponding Author: h.farmer@gre.ac.uk, School of Human Sciences, University of  
12 Greenwich, London, SE10 9LS, UK

13 **Word Count:** 6711

14 **Acknowledgment:** This work was supported by the EPSRC Virtual Realities - Immersive  
15 Documentary Encounters project - EP/P025595/1

16

1 **Abstract**

2 Advances in head mounted displays (HMDs) have increased the interest in cinematic virtual  
3 reality as an artform. However, the freedom of a viewer in 360 video presents challenges in  
4 ensuring that audiences do not inadvertently miss important events and locations. We  
5 examined whether the high level of immersion provided by HMDs encourages participants to  
6 synchronise their attention during viewing. Sixty-four participants watched the 360°  
7 documentary “Clouds Over Sidra” using either an HMD or via a flat screen tablet display.  
8 We used inter-subject correlation (ISC) analysis to measure attentional synchrony over the  
9 course of the video and to examine whether spatial and temporal factors led to different  
10 amounts of correlation both within and between groups. We found significantly greater ISC  
11 for the HMD compared to the tablet group. This effect was greatest for scenes with a  
12 unidirectional focus and at the start of scenes. We discuss our results in terms of the visual  
13 properties and the motor affordances of HMDs vs tablets. Our results show the value of  
14 HMDs in increasing attentional synchrony and may provide producers of 360° content insight  
15 in how to encourage or discourage synchronisation of viewing direction.

16

17

18

19

## 1 **Public Significance Statement**

2 The rapid uptake of VR by producers of nonfiction reflects a fascination with the potential of  
3 360 media for viewer engagement. While VR allows viewers a novel freedom to find their  
4 own visual pathway through a given scene, at the same time producers have specific  
5 information that they need to convey. The power to elicit synchronization of viewers'  
6 attention is often held up as a key indicator of a content creator's ability to lead an audience  
7 through a narrative. In this paper we have shown that 360° video can elicit a degree of  
8 synchronisation between viewers and that this effect is increased when the video is viewed  
9 via a HMD when compared to being viewed on a flat screen tablet.

10

11

12

13

14

15

16

17

18

19

## 1 **Introduction**

2 The rapid advances in virtual reality (VR) technology in the last few years has led to  
3 increased interest in VR's potential as a means of delivering narrative media. Since late 2015,  
4 the on-going release of affordable, high specification head mounted display (HMD) systems  
5 including Oculus's Rift (2016) and Quest (2019), Samsung's GearVR (2015) and HTC's  
6 Vive (2016) has driven a vast increase in VR content. Release of English language nonfiction  
7 VR pieces alone increased from under 40 pieces in 2015 to over 180 in 2017 (Bevan &  
8 Green, 2017), and recent projections suggest that the VR and AR market will be worth \$170  
9 billion by 2022 (consultancy.uk, 2018).

10 Despite this enthusiastic uptake, the unique nature of VR presents significant artistic and  
11 technical challenges for content production and delivery. Within nonfiction production, 360  
12 video is widely used. In this content production model, footage from a series of cameras set  
13 in a circular array are stitched together to form one panoramic scene. This scene is then  
14 projected onto a spherical surface and then usually presented via an HMD where directional  
15 data is used to determine what section of the video ('viewport') to display. Producers  
16 transitioning from traditional filmmaking in 2D into the world of virtual narrative have often  
17 been drawn to the use of 360° video due to similar skill requirements to traditional film  
18 making, high visual fidelity, platform adaptability and relatively low cost. A recent analysis  
19 of nonfiction VR (Bevan et al., 2019) suggests that around three quarters of all English  
20 language nonfiction VR content made between 2012 and 2018 were either solely or  
21 predominately composed of 360° video.

22 However, the omnidirectional nature of 360° video inevitably reduces the amount of control  
23 media producers have over the direction of viewers' attention, leading for calls for the

1 development of “a new screen grammar” (Dooley, 2017) and greater understanding of “the  
2 geometry of story-telling” (Pope et al., 2017). In addition to these issues around content  
3 production, greater knowledge of where viewers are likely to direct their gaze could also help  
4 to optimise the storage, transmission and rendering of 360° videos (Ozcinar & Smolic, 2018).  
5 In this context, an increased understanding of how people visually navigate through 360°  
6 videos has the potential to yield important insights for both content producers and tech  
7 developers working in VR.

8 One source of insight in designing compelling VR scenes is to examine previous research on  
9 viewing attention in the perception of dynamic scenes. Research in this area has used eye  
10 tracking technology to identify the factors that contribute to gaze fixation, taken as a proxy  
11 for attention. Using this approach, researchers have demonstrated that a variety of factors can  
12 modulate the pattern of viewer gaze. These include low level salience (Carmi & Itti, 2006;  
13 Itti, 2005), the motion of objects or people within the scene (Mital et al., 2011), task  
14 requirements (Hutson et al., 2017) amount of social content (Birmingham et al., 2009;  
15 Coutrot & Guyader, 2014; Rubo & Gamer, 2018), temporal order (Wang et al., 2012) and  
16 emotional valance (Rubo & Gamer, 2018; Subramanian et al., 2014).

17 Perhaps the most effective measure for assessing the success of content producers in driving  
18 viewing behaviour in 360° video is the extent to which different viewers attend to the same  
19 area of a scene. Such attentional synchrony (AS) plays a key role in successful story telling  
20 by allowing directors and editors to ensure that viewers are attending to the part of the visual  
21 scene that is most critical to the narrative at each particular time (Smith, 2013; Smith et al.,  
22 2012). Previous research has shown the role of several factors in modulating AS when  
23 watching traditional 2D films. Mital et al. (2011) showed that motion was a key factor in  
24 determining how much different individuals gazes clustered around similar points, a finding

1 supported by the fact that AS is higher when viewing dynamic as opposed to static scenes  
2 (Smith & Mital, 2013). AS has also been shown to be greater for free as opposed to task  
3 driven viewing (Smith & Mital, 2013) and to decrease during repeated viewing of a scene  
4 (Breathnach, 2016) suggesting that AS may be driven by the spontaneous extraction of novel  
5 scene features. Other studies have demonstrated that top down factors like viewer context  
6 (Loschky et al., 2015) can also influence AS, and AS has also been shown to have  
7 developmental and evolutionary components with higher synchrony in adults compared to  
8 children (Franchak et al., 2016; Kirkorian & Anderson, 2018) and in humans compared to  
9 monkeys (Shepherd et al., 2010).

10 One consistent and particularly relevant finding is that AS is stronger for tightly cut videos  
11 than for naturalistic scenes (Dorr et al., 2010; Hasson, Furman, et al., 2008). Temporal order  
12 also affects AS with greater AS when viewing structured compared to randomly edited  
13 sequences (Kirkorian & Anderson, 2018). These findings demonstrate the importance of  
14 techniques such as continuity editing, and close ups in directing the viewers gaze towards  
15 salient narrative events. The lack of many of these techniques in 360° video presents a  
16 challenge for content producers in ensuring that audience members perceive the story as  
17 intended, rather than fixating on areas of the scene without meaningful content.

18 A variety of approaches have been used to measure AS including: (a) bivariate contour  
19 ellipse area (Kirkorian & Anderson, 2018), which assesses the size of ellipse needed to  
20 capture all participants' gaze coordinates; (b) normalized scan path saliency (Dorr et al.,  
21 2010), which measures the correspondence between saliency maps and ground truth,  
22 computed as the average normalized saliency at fixated locations; (c) Gaussian mixture  
23 modelling (Mital et al., 2011; Smith & Mital, 2013), which probabilistically represents the  
24 clustering of eye movements across subjects; and (d) inter-subject correlation (ISC;

1 Burleson-Lesser et al., 2017; Franchak et al., 2016; Shepherd et al., 2010), in which each  
2 participant's motion path is separately correlated with that of every other participant before  
3 being averaged together to create a mean ISC for that participant. In the current study we  
4 chose to use ISC as this approach allows for the calculation of each individual's average AS  
5 with all other group members. This enables us to examine individual difference in AS while  
6 the other approaches can only characterise AS at the whole group level (Franchak et al.,  
7 2016).

8 ISC has been previously used to examine synchrony in neural responses to naturalistic film  
9 (Adolphs et al., 2016; Hasson et al., 2004; Hasson, Landesman, et al., 2008; Herbec et al.,  
10 2015). Hasson and colleagues (2008) employed functional magnetic resonance imaging  
11 (fMRI) and eye tracking and found that viewing tightly edited films such as Sergio Leone's  
12 "The Good, The Bad, and The Ugly" led to greater ISC for both viewer's eye movements and  
13 their neural response in brain areas associated with vision and attention than unedited  
14 footage. This suggests that ISC is a strong marker of how much control a director has over  
15 audience attention. Increased ISC in neural responses has been linked to improved memory  
16 encoding of events (Hasson, Furman, et al., 2008) and the amount of ISC in small groups of  
17 viewers has been shown to reliably predict the preferences of thousands (Dmochowski et al.,  
18 2014). Further work has expanded the use of ISC outside of fMRI data to show  
19 synchronisation of response in EEG (Poulsen et al., 2017), MEG (Lankinen et al., 2014) and  
20 physiological (Bracken et al., 2014; Golland et al., 2014) measures. ISC of eye movements  
21 has also been found to correlate with gaze salience (Franchak et al., 2016) indicating that it  
22 may act as a reliable indicator of visual salience.

23 To date only two studies (Bender, 2018; Sitzmann et al., 2018) have examined viewer  
24 synchrony in 360° video. Bender (2018) carried out a qualitative assessment of heat maps



1 based on the head direction of participants who watched a narrative 360° video and used a  
2 qualitative approach to classify the maps according to the concentration of attention within  
3 the scene. Bender found that most scenes displayed strong AS which appeared to be driven  
4 by the salient events such as a character speaking. Sitzmann and colleagues (2018) took a  
5 more quantitative approach tracking the head and eye movements of participants wearing an  
6 HMD while viewing 22 static 360° scenes. They calculated the receiver operating  
7 characteristic curve of each participant's fixations compared to the fixations of all other  
8 participants and found a high level of AS comparable to previous findings from 2D scenes.  
9 However, both studies had limitations in terms of measuring AS to dynamic scenes. Bender's  
10 (2018) analysis was purely qualitative and did not statistically test the amount of synchrony  
11 across the entire video, while the use of static scenes by Sitzmann and colleagues (2018)  
12 limits how far their results can be extrapolated to the dynamic scenes found in most 360°  
13 videos.

14 One factor that might alter the amount of AS when watching 360° video is the device on  
15 which the video is displayed. While the development of commercial HMDs has been a key  
16 driver of 360° video content production, one of the appeals of 360° video compared to other  
17 forms of virtual reality is that it can easily be adapted for non-HMD formats including phones  
18 and tablets, thus broadening the audience for 360° video beyond those with access to HMDs  
19 (Zoric et al., 2013).

20 Three previous studies have contrasted viewing of 360° videos on 2D displays and HMDs.  
21 Sitzmann and colleagues (2018) found no difference in the similarity of fixation locations  
22 when participants watched still 360° scenes using a 2D desktop display when compared to  
23 using HMD. The other two studies did not collect gaze data but did examine viewer  
24 responses to different display conditions. MacQuarrie and Steed (2017) compared viewing a

1 full 360° video in an HMD with viewing a non-360° version of the video on either a normal  
2 desktop display or a display-plus-peripheral projections. The authors used a variety of  
3 measures including whether attention could be more easily guided towards important items in  
4 the 2D screen conditions compared to HMD but found no significant difference in their  
5 measure of subsequent recall of cued events suggesting an equal amount of attentional  
6 control between the conditions. Finally Passmore and colleagues (2016) used qualitative  
7 interviews to assess differences between watching 360° video in an HMD or on a  
8 smartphone, or watching a 2D version on a desktop screen. They then coded these interviews  
9 and found that overall attention was high in all three conditions. However, these studies all  
10 had limitations that narrow the conclusions that can be drawn from them. Both MacQuarrie  
11 and Steed and Passmore et al. relied on user reports rather than directly comparing AS, while  
12 Sitzmann and colleagues did not examine dynamic scenes in which HMD users might be  
13 more influenced by motion cues. It is therefore unclear whether AS in 360° video differs  
14 between display devices.

15 The current study examined whether the type of device used to deliver the video and specific  
16 features within the 360° video affect synchronisation between viewers. Due to the difficulties  
17 in deploying eye-tracking devices within an HMD, and given the finding of Sitzmann and  
18 colleagues (2018) of a strong coupling between gaze and head movements when viewing  
19 360° scenes, we did not directly track gaze but rather used directional data from the  
20 HMD/Tablet with the central point in the scene as a proxy for gaze. We calculated the within  
21 group ISC for participants' movements in the pitch and yaw axes as they watched a 360°  
22 video using either an HMD or a tablet in motion-tracked 'magic window' mode. We also  
23 investigated how both the spatial and temporal aspects of viewed scenes affected ISC and  
24 how this interacted with the type of device used.

## 1 **Method**

### 2 **Participants**

3 Sixty-four participants (mean age  $\pm$  SD: 25.82  $\pm$  11.31 years; 25 males) gave their written  
4 informed consent to participate and were paid for their participation. Thirty-two participants  
5 were assigned to the HMD and Tablet groups respectively. Due to the lack of previous  
6 research examining ISC between different devices, sample size was calculated using  
7 G\*Power (Faul et al., 2007) based on having the power to detect a small ( $d = 0.25$ )  
8 interaction between viewing device and the temporal/spatial quality of scenes. All  
9 participants were screened for previous history of epilepsy or brain injury and for family  
10 history of epilepsy prior to taking part in the study. All participants had normal or corrected  
11 to normal vision without the need for glasses. Ethical approval for the study was given by the  
12 University of Bath's Department of Psychology Research Ethics Committee.

### 13 **Design**

14 The experiment had a between subjects design with one independent variable which was the  
15 device used to watch the 360° documentary (HMD vs Tablet). During the study we measured  
16 pitch and yaw rotational data, captured from the on-board orientation sensors of the HMD  
17 and Tablet computer and used those values to derive the great circle distance of each  
18 coordinate pair from a reference point of  $[0^\circ, 0^\circ]$ .

### 19 **Materials**

20 A tablet rather than a mobile phone or cursor controlled desktop setup was used as the control  
21 condition as this allowed for the presentation of 360 video in a flat screen format that most

1 closely matched the affordances of the HMD (i.e. requiring whole body movement to rotate  
2 in yaw and with mechanistic restrictions on the amount of possible rotation in pitch rather  
3 than simply depending on cursor movement).

4 In order to habituate participants to the affordances of the viewing device they were using, all  
5 participants first viewed an introductory 2:20 minute cut from the 360° video ‘Nature  
6 Meditation’ (Eco VR, 2017) which shows scenes of nature. For the main task participants  
7 watched the 8:35 minute 360° documentary ‘Clouds Over Sidra’ (VRSE.works, 2015).  
8 ‘Clouds over Sidra’ provides a tour of a refugee camp with Sidra, a young girl from Syria, as  
9 a guide. Sidra shows her family, her makeshift classroom, as well as other parts of the camp,  
10 and talks about her life in the camp and hopes for the future. The documentary offers 360-  
11 degree immersion in the settings featured in the documentary and directed sound  
12 corresponding to the view. Participants in the HMD condition viewed a higher resolution  
13 (3840px x 1920px) version of the film, presented in the monoscopic equirectangular  
14 projection format with no stereoscopic separation. Due to the limits of available screen  
15 resolution, participants in the Tablet condition viewed a slightly lower resolution (2732px x  
16 1366px) version of the film. However, as the HMD splits and duplicates visual content to  
17 present a full image to each eye separately the perceived image quality of the film was – in  
18 practice - roughly equivalent across the HMD and tablet conditions.

19 The hardware setup for the HMD condition consisted of one laptop PC station (Intel Core i7-  
20 7700HQ 2.80 GHz CPU, 16 GB of RAM, Nvidia GeForce GTX 1070 Graphics card). The  
21 HMD was an Oculus Rift (consumer model), with a field of view of 110° nominal, a  
22 resolution of 1080×1200 pixels per eye, and a refresh rate of 90Hz. In the tablet condition  
23 participants viewed the documentary on a Samsung Galaxy Tab S 10.5 (Model SM-T80016,  
24 Exynos 5 Octa 5420 1.9 GHz CPU, 3GB of RAM, Mali-T628 MP6 Graphics). This tablet has

1 a screen resolution of 2560px x1600px, a field of view of approximately 85° and a refresh  
2 rate of 60Hz. Audio in the Tablet condition was delivered via Sony MDR-ZX330BT  
3 Bluetooth wireless headphones.

4 In both setups, participants viewed the 360° videos using a custom-built 360° video player  
5 application that was created by the researchers using the Unity software development  
6 package (version 2017.3.1f1). Unity is a cross-platform development environment, allowing  
7 the same codebase to be shared between the HMD and the Tablet. Playback of 360° video  
8 content was facilitated using the in-built video player component provided by Unity, and was  
9 rendered using the Skybox Panoramic Shader method detailed in Margerie (2018).

10 In addition to our head tracking measures we also measured engagement with the  
11 documentary using a seven item scale that Schutte and Stilinović (2017) adapted from Wiebe,  
12 Lamb, Hardy, & Sharek (2014) which included statements such as “The time just slipped  
13 away” and “I lost track of the world around me”. Responses were given on a Likert scale  
14 from 1 to 5 and the response across questions were averaged together to create a mean  
15 engagement score.

## 16 **Procedure**

17 Participants were randomly assigned to either the HMD or Tablet condition and verbally  
18 briefed about the study. Participants in the HMD condition were given additional information  
19 about how to adjust the headset straps and inter-ocular distance for viewing comfort while  
20 those in the Tablet condition were shown how the motion-tracked “magic window” set up of  
21 the tablet worked. All participants viewed the 360° videos while seated on a swivel chair that  
22 allowed easy rotation in the yaw dimension. For participants in the HMD condition the  
23 viewpoint in the yaw dimension could be changed by moving either their head or their whole

1 body to either side. The viewpoint in the pitch dimension could be changed by tilting their  
2 head up or down. For participants in the tablet condition the viewpoint in the yaw dimension  
3 could be changed by rotating either the tablet itself, or their whole body while holding the  
4 tablet, from side to side in the horizontal plane. The viewpoint in the pitch dimension could  
5 be changed by rotating the tablet, held in their hands, up or down.

6 The study was conducted in the CREATE lab at the University of Bath, a large room which  
7 contained a number of desks containing computers, a small sink area in the corner nearest  
8 participants and a set of floor to ceiling windows offering a view of the Bath campus. While  
9 using the device participants were concealed from the experimenter by a screen so that they  
10 did not feel too self-conscious to visually explore the environment. After watching the short  
11 introductory video, direction tracking was turned on and participants watched “Clouds Over  
12 Sidra”. Having watched the piece participants completed a number of questionnaires  
13 including the measure of engagement. Finally, participants were thoroughly debriefed and  
14 given a detailed sheet explaining the purpose of the experiment.

## 15 **Data Analysis**

### 16 *Pre-processing*

17 For every rendered frame of playback, the current pitch and yaw orientation of the main  
18 scene camera was captured as a Euler angle and logged in a text file. This resulted in a  
19 sampling rate of 90Hz for the HMD, and a slightly reduced rate of 60Hz for the less powerful  
20 Tablet computer. Each frame was further labelled with the current time elapsed (msec) and  
21 the frame number. Orientation values were captured from the viewport camera using the  
22 transform class provided by Unity (Camera.main.transform.eulerAngles.x and  
23 Camera.main.transform.eulerAngles.y).

1 For the tablet condition, the 360° video was rendered to a full screen monoscopic viewport,  
2 with the orientation of the virtual camera synchronised to the device's internal gyroscope.  
3 This method of viewing 360° video is commonly referred to as “Magic Window”, and  
4 provides an alternative method of viewing 360° video content (while retaining the freedom to  
5 move the viewport) in the absence of a Virtual Reality HMD.

6 By default, the inbuilt gyroscope of the tablet is oriented to assume that the tablet is placed  
7 flat on a table. To place the device in “magic window” mode, the baseline pitch of the camera  
8 required adjustment prior to playback to allow the participant to hold the device at eye level.  
9 To achieve this, participants were provided with a button to calibrate the viewport of the  
10 tablet by adjusting the pitch of the camera in 5-degree increments until an artificial horizon  
11 was centred in the middle of the screen. This adjustment was then stored and subsequently  
12 applied as an offset to each playback frame. For the HMD, no calibration was required, and  
13 no additional pre-processing was applied to the orientation data.

14 In both the HMD and tablet versions, the sampling rate of the viewport comfortably exceeded  
15 the frame rate of the video. Prior to analysis, tracking data was down-sampled to a rate of  
16 10Hz to remove redundant and duplicated data points. Since our main interest was how  
17 participants synchronised their head movements while watching naturalistic scenes, we then  
18 further cut the time series to remove the opening and closing credits, leaving a time series  
19 with a total time of 7:16.2 minutes, starting at 9.1 seconds into the original video and ending  
20 at 7:25.5 minutes into the video.

## 21 *Descriptive Statistics*

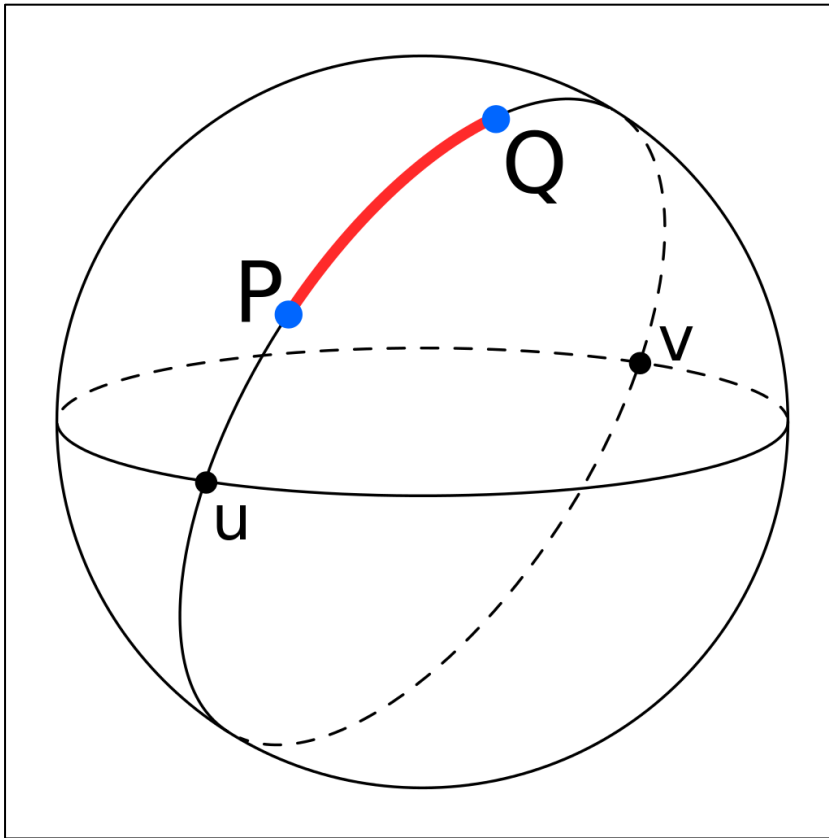
22 Due to the directional nature of our statistics we present our descriptive statistical data in the  
23 format recommended by Cremers and Klugkist (2018) using the R package “circular”

1 (Agostinelli & Lund, 2017). We give values in each condition and dimension for four  
2 variables. First, the mean direction ( $\bar{\theta}$ ) which is found by converting each circular datapoint  
3 into a vector composed of the sine and cos of that datapoints value in degrees, adding all the  
4 resulting vectors head to toe, and finding the direction of a new vector connecting the toe of  
5 the first vector to the head of the last vector. Second, the mean resultant length ( $\bar{R}$ ) which is  
6 found by taking the length of the new vector described above, meaning that a lower value  
7 represents greater dispersion in datapoints. Third, the circular variance ( $V_m$ ) which is the  
8 opposite of the mean resultant length and finally the circular standard deviation ( $v$ ) where  
9 higher values indicate greater dispersion. To test differences between the distribution of mean  
10 pitch and yaw values between our device groups we report the outcomes of Watson-Wheeler  
11 tests.

12 We used the pitch and yaw data outputted from our player to calculate the great-circle  
13 distance (GCD) between each coordinate pair and the starting point of  $[0^\circ, 0^\circ]$  using the  
14 Haversine method. This represents the shortest distance between two points on the surface of  
15 a sphere, measured along the surface (as opposed to a straight line through the sphere's  
16 interior, see Figure 1). To do this we first converted our yaw data into longitude by re-  
17 centring it so that values above  $180^\circ$  had  $360^\circ$  subtracted from them (meaning that the range  
18 of value now went from  $-180^\circ$  to  $180^\circ$  with 0 retaining its original value. Next the pitch data  
19 was converted into latitude by making pitch values below  $90^\circ$  negative, subtracting  $180^\circ$   
20 from values between  $90^\circ$  and  $270^\circ$  and subtracting  $360^\circ$  from values above  $270^\circ$  and then  
21 making that value positive. Following this we used the `distHaversine` function from the R  
22 package "geosphere" (Hijmans et al., 2019) to calculate GCD of each pair of coordinates in  
23 each participants time series relative to  $[0^\circ, 0^\circ]$  using an arbitrarily set radius value of one.  
24 This approach allowed us to focus on the amount of ISC between participants from a set  
25 reference point rather than the relative change in direction at each point of their time series.



- 1 We report the results of independent sample t-tests and Levene's test for homogeneity of
- 2 variance which examined differences between the mean GCD of our two device groups.



3 *Figure 1. A diagram illustrating great-circle distance (drawn in red) between two points on a*  
4 *sphere, P and Q. Two antipodal points, u and v, are also shown (reprinted from Wikipedia*  
5 *(2020)).*

### 6 ***Analysis of Overall ISC***

- 7 In order to calculate ISC, we first split our participants between the HMD and Tablet group.
- 8 Next, we calculated the Pearson product moment correlation coefficient ( $\rho$ ) of each
- 9 participant's GCD data with the GCD data of every other participant in their device
- 10 condition. To avoid underestimation due to the skewed sampling distribution of the
- 11 correlation coefficient (Silver & Dunlap, 1987) we then converted these correlations into
- 12 Fisher-transformed, z-normalised coefficients using the following formula:

$$z' = 0.5 \log \left( \frac{1 + \rho}{1 - \rho} \right)$$

2 The z-normalised coefficients of each participant's data with that of all other participants  
 3 were then averaged and the resulting mean z-transformed coefficients were transformed back  
 4 into  $\rho$  to gain the mean correlation coefficient of each participant's tracking data with all  
 5 other participants in their group, i.e. their ISC. We then compared whether the ISC for each  
 6 group differed from 0 using one sampled t-tests and whether ISC differed between the groups  
 7 using Welch's Two Sample t-tests. For descriptive purposes we also carried out an analysis  
 8 of ISC for the whole period split across 10 second bins (see Figure 4B).

### 9 *Analysis of Each Scene's ISC*

10 In order to examine whether different types of scene showed greater correlation between  
 11 participants in either group or between participant groups, we split our GCD data into time  
 12 series based on the start and end point of each scene and then calculated each participant's  
 13 ISC relative to their group. We then qualitatively categorised the scenes according to how far  
 14 the scene's features directed the viewer's attention to a central point. Three categories were  
 15 used: 1) Unidirectional: scenes in which there is one central focus that could be viewed  
 16 within one 110° window. In all cases this was either an individual person or group of people;  
 17 2) Intermediate: scenes in which there was more than one location containing people or  
 18 objects of interest, but where at least 110° of the frame contained no objects of interest. 3)  
 19 Panoramic: Scenes in which there were people or objects of interest spread across the full  
 20 360° panorama, or where there were no clear objects of interest. Figure 2 shows two  
 21 examples of each of the scene categories, while Table 1 gives a brief description of each  
 22 scene, the category in which it was placed and its start and end times. To test whether the  
 23 amount of ISC differed across these different scene types (and if the type of device used

1 influenced that difference), we first calculated each participant's mean fisher-transformed, z-  
2 scored ISC for each scene using the method described above. We then averaged the scenes in  
3 each category and transformed the z-scored means back into Pearson's  $\rho$ . We then ran an  
4 ANOVA comparing levels of ISC for each scene type.

### 5 *Analysis of ISC Within Scenes*

6 We were also interested in examining how ISC changed according to time period within each  
7 scene. To investigate this, we first calculated each participant's mean fisher-transformed, z-  
8 scored GCD ISC over 10 seconds after the beginning, around the mid-point and before the  
9 end of each scene. We then averaged together the ISCs across scenes for each time period  
10 and then transformed the z-scored means back into Pearson's  $\rho$ . We then ran an ANOVA  
11 comparing levels of ISC for each time period.

### 12 *Analysis of Relationship between ISC and Engagement*

13 To examine the effect of Device on engagement we ran a Welch's Two Sample t-test. To  
14 examine the relationship between ISC and engagement we carried out a linear regression with  
15 GCD ISC as the respective predictor and mean engagement score as the outcome variable.

16

17

18

19

20

- 1 Table 1. Scene numbers and description along with categories used in the scene analysis and  
 2 start and end times used to define each scene time series.

<b>Scene</b>	<b>Description</b>	<b>Category</b>	<b>Start (m:s)</b>	<b>End (m:s)</b>	<b>Length (m:s)</b>
<b>1</b>	View of desert	Panoramic	0:09.1	0:33	00:23.9
<b>2</b>	Sidra in bedroom	Unidirectional	0:33.1	0:50.9	00:17.8
<b>3</b>	Sidra's family in house	Bidirectional	0:51	1:10	00:19.2
<b>4</b>	Journey to school	Panoramic	1:10.3	1:38.8	00:28.5
<b>5</b>	Classroom	Intermediate	1:38.9	2:08.6	00:29.7
<b>6</b>	Bakery	Intermediate	2:08.9	2:44	00:35.2
<b>7</b>	Children not in school	Unidirectional	2:44.1	3:04.8	00:20.7
<b>8</b>	Computer Room	Panoramic	3:04.9	3:48	00:43.1
<b>9</b>	Gym	Panoramic	3:48.1	4:13.5	00:25.4
<b>10</b>	Wrestling	Intermediate	4:13.6	4:44.3	00:30.7
<b>11</b>	Football Pitch	Intermediate	4:44.4	5:33.1	00:48.7
<b>12</b>	Family dinner	Unidirectional	5:33.2	6:16.1	00:42.9
<b>13</b>	Crowd of children	Panoramic	6:16.2	6:41	00:24.8
<b>14</b>	Sidra's bedroom	Unidirectional	6:41.1	6:55.6	00:14.5
<b>15</b>	View of Camp	Panoramic	6:55.7	7:25.3	07:25.3

3



1 *Figure 2. Examples of unwrapped 360° video from each of the three categories used in the*  
 2 *scene ISC analysis.*

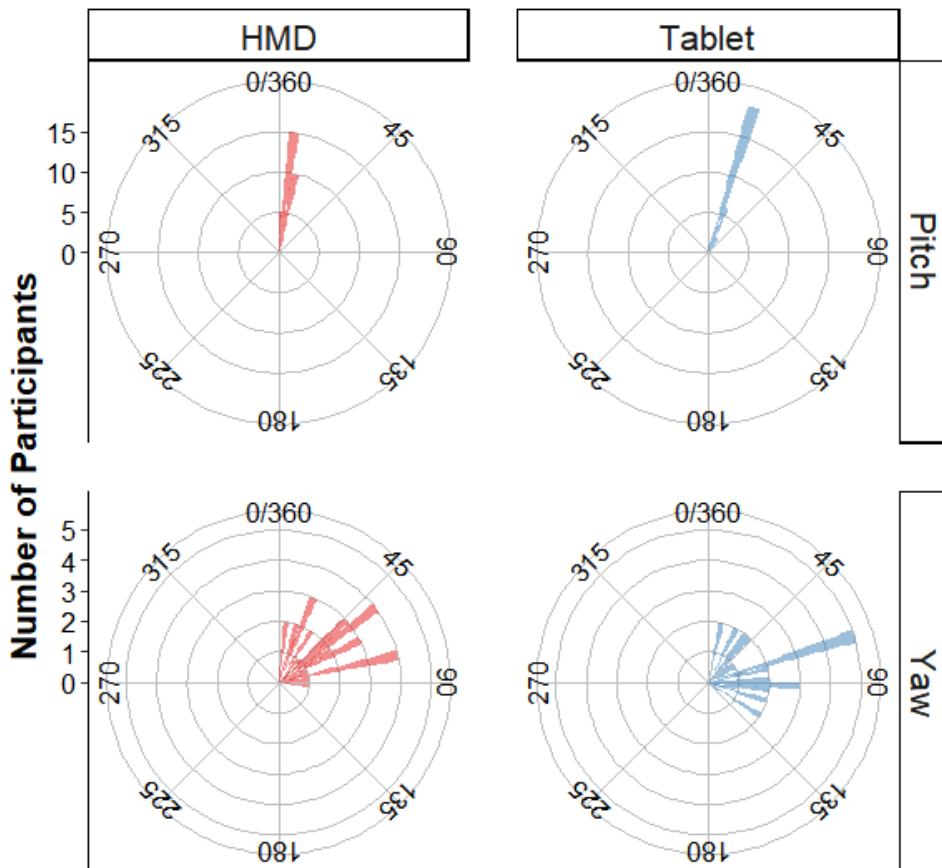
### 3 **Results**

#### 4 **Descriptive Statistics**

5 Table 2 shows the descriptive statistics for each group. Due to the directional nature of our  
 6 statistics we present our data in the format recommended by Cremers and Klugkist (2018).  
 7 As can be seen the mean directions ( $\bar{\theta}$ ) for both pitch and yaw were further from 0 in the  
 8 Tablet group compared to the HMD group, suggesting that the Tablet group deviated further  
 9 from the starting position than the HMD group (see Table 1 and Figure 3). In addition, the  
 10 Tablet group show a smaller mean resultant length ( $\bar{R}$ ) for both yaw and pitch suggesting  
 11 greater spread in the mean direction in the Tablet group than in the HMD group. Both groups  
 12 also showed a smaller  $\bar{R}$  on the yaw compared to pitch axis.

13 To further investigate these differences, we carried out separate Watson-Wheeler tests on the  
 14 data from each axis in order to assess whether the distribution of  $\bar{\theta}$  between the device groups  
 15 indicated that these group could be considered to be drawn from the same population. The

1 test for the pitch axis showed a significant difference in  $\bar{\theta}$  distribution between the HMD and  
 2 Tablet groups  $W(2) = 35.83, p < .001, \eta^2 = 0.42$ . The test for the yaw axis also showed a  
 3 significant difference in  $\bar{\theta}$  distribution between the HMD and Tablet groups  $W(2) = 6.37, p =$   
 4  $.041, \eta^2 = 0.15$ .



5 *Figure 3. Plots showing the distribution of participants mean viewing direction ( $\theta$ ) across the*  
 6 *whole viewing session for the pitch (top) and yaw (bottom) axes for the HMD (left) and*  
 7 *Tablet (right) groups.  $\theta$  values were organised in bins every 5°, longer lines from the centre*  
 8 *represent a larger number of participants in that bin.*

9 To investigate whether the two device groups differed in their average GCD a Welch two  
 10 sample t-test comparing the HMD and Tablet groups' mean distance across the session. This  
 11 revealed no significant difference between the HMD (Mean = 1.28, SD = 0.23) and tablet

1 (Mean = 1.39, SD = 0.30) groups,  $t(58.03) = -1.75$ ,  $p = .086$ ,  $\eta^2 = 0.08$ . Levene's test for  
 2 homogeneity of variance indicated that there was no significant difference in variance  
 3 between the two groups ( $F = 0.27$ ,  $p = .607$ ).

4 Table 2. Descriptive statistics for viewing angle data with mean direction ( $\bar{\theta}$ ), mean resultant  
 5 length ( $\bar{R}$ ), circular variance ( $V_m$ ) and circular standard deviation ( $v$ ) for each device.

Axis	Device	$\bar{\theta}$	$\bar{R}$	$V_m$	$v$
Pitch	HMD	9.22°	.997	.003	1.84°
	Tablet	19.29°	.996	.004	1.47°
Yaw	HMD	51.57°	.919	.081	2.21°
	Tablet	66.48°	.828	.172	2.03°

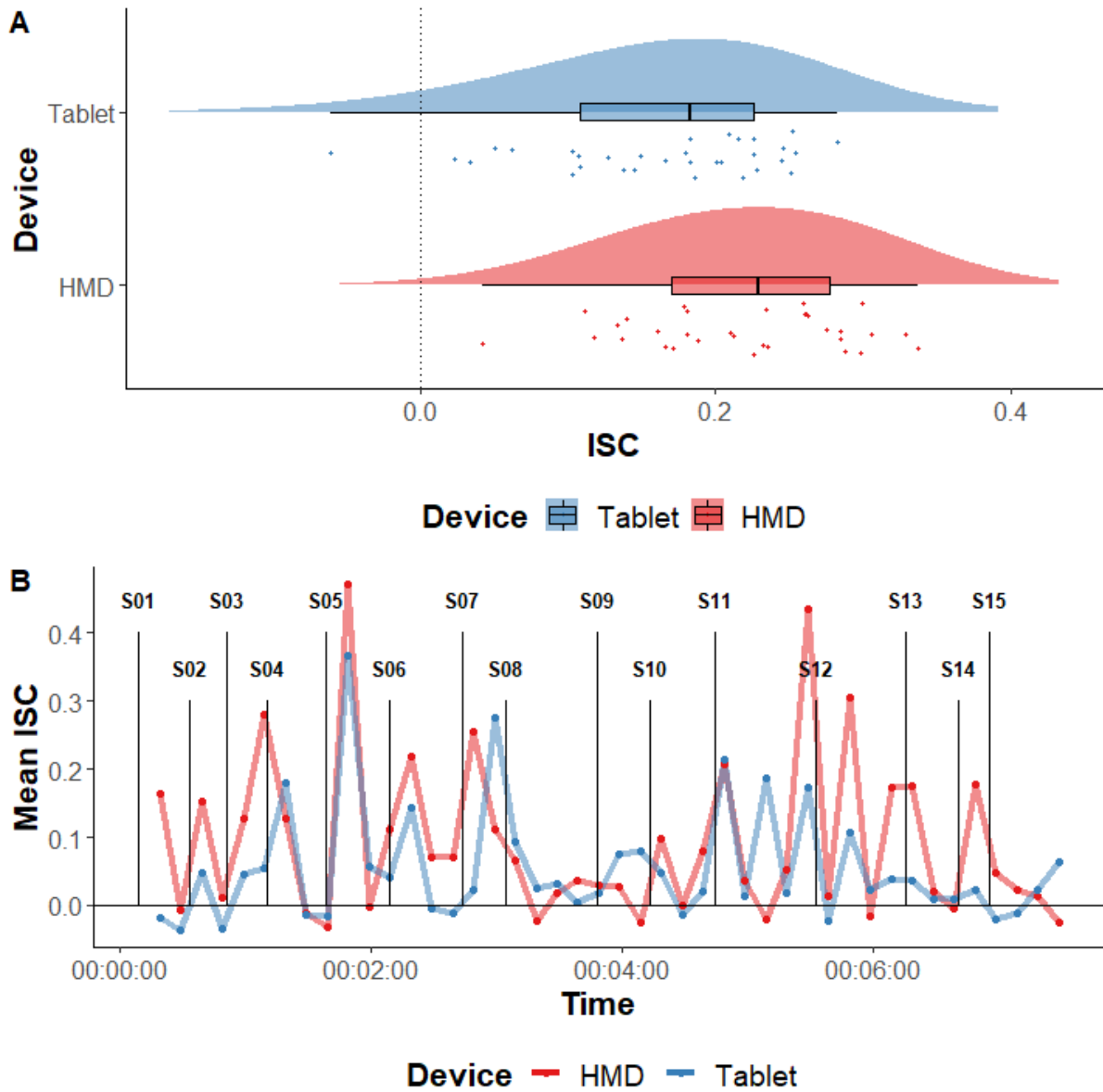
## 6 **ISC Analysis**

### 7 *Inter-subject Correlation Across the 360° video*

8 To test whether our groups showed significant ISC, we first compared whether the mean  $\rho$   
 9 for GCD in each group differed from 0 using one sample t-tests. As Table 3 shows ISC for  
 10 both devices was significantly greater than 0.

11 To discover if ISC significantly differed between devices across the whole study, we ran a  
 12 Welch two sample t-test comparing the ISC of the HMD and Tablet groups. This revealed a  
 13 significant difference between groups,  $t(61.13) = 2.89$ ,  $p = .005$ ,  $\eta^2 = 0.17$ , (see Figure 4).

# Attentional synchrony during 360° video viewing



1 *Figure 4. A) Mean ISC across the whole viewing session across device. Clouds represent*  
2 *distribution, raindrops represent individual participants. B) Results of an analysis of ISC in*  
3 *10 second bins across the entire time period. Points represent mean ISC for the preceding 10*  
4 *seconds. Markers represent scene starts.*

5

6



1 Table 3. Mean ISC and standard deviation (SD) along with the t statistic (t) and degrees of  
 2 freedom (df) for one sample t-tests comparing against 0 for each group.

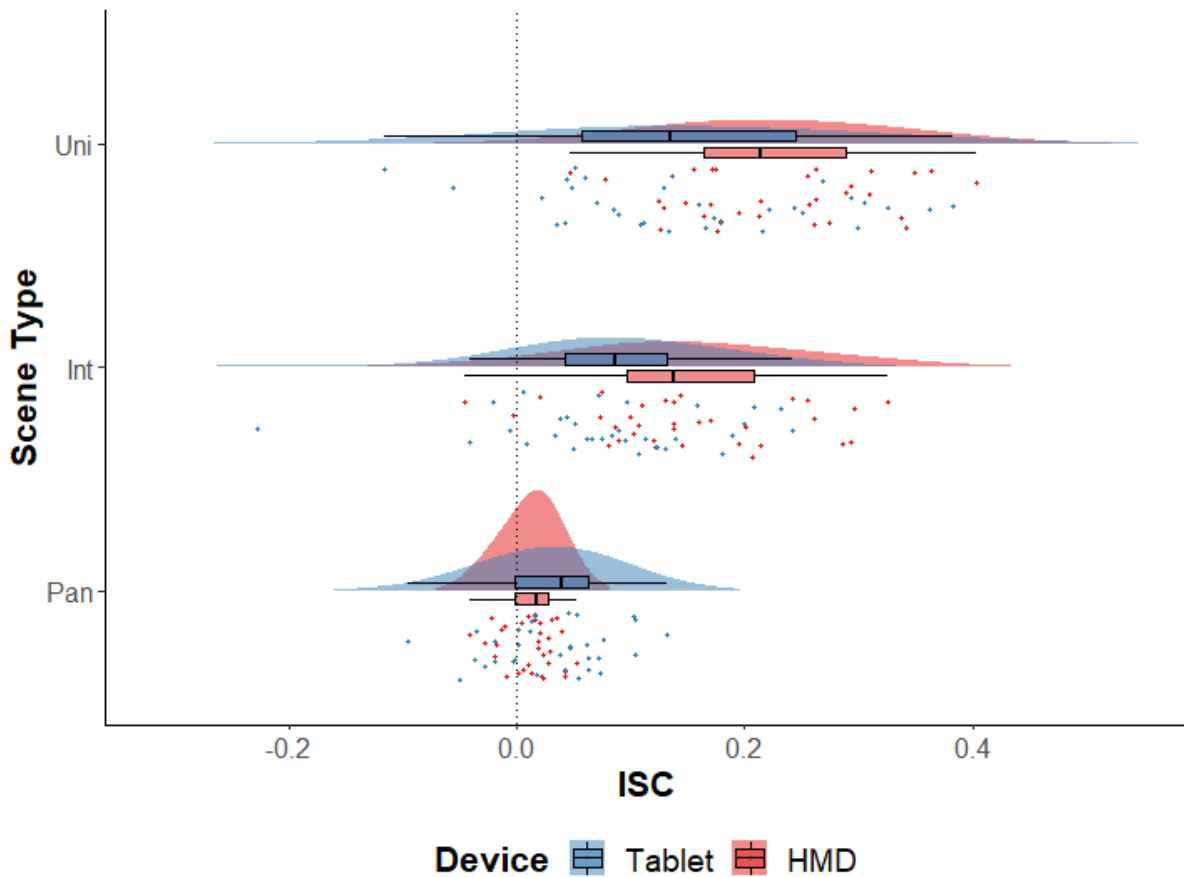
Device	Mean ISC	SD	t	df
HMD	0.219	.071	17.34	31
Tablet	0.163	.081	9.88	31

### 3 Inter-subject Correlation by Scene Type

4 To investigate how directional cues within scenes affected ISC, we carried out a 2x3  
 5 ANOVAs with device (HMD vs Tablet) as the between subject factor and scene category  
 6 (Unidirectional vs Intermediate vs Panoramic) as the within subjects factor (see Table 4 and  
 7 Figure 5).

8 Table 4. Means and standard deviations of ISC across scene types, device and measure. Same  
 9 letter superscripts represent a significant difference between devices within scene type and  
 10 measure in post hoc t-tests of estimated marginal means, † indicates  $p < .01$ .

Device	Scene Type	Mean ISC	SD
HMD	Unidirectional	.225 <sup>†a</sup>	.088
	Intermediate	.152 <sup>†b</sup>	.089
	Panoramic	.012	.023
Tablet	Unidirectional	.151 <sup>†a</sup>	.012
	Intermediate	.085 <sup>†b</sup>	.092
	Panoramic	.030	.051



1 *Figure 5. Mean ISC across scene type and device. Clouds represent distribution, raindrops*  
 2 *represent individual participants. Uni equals Unidirectional, Int equals intermediate and Pan*  
 3 *equals Panoramic.*

4 We found a significant effect of device,  $F(1,62) = 9.15, p = .004, \eta_p^2 = 0.13$ , which was  
 5 driven by higher ISC in the HMD group compared to the Tablet group. There was also a  
 6 significant effect of scene type,  $F(1.82,113.04) = 74.09, p < .001, \eta_p^2 = 0.54$  which was  
 7 driven by significantly higher ISC in Unidirectional scenes compared to Intermediate scenes,  
 8  $t(189) = -4.49, p < .001$ , and Panoramic scenes,  $t(189) = -10.70, p < .001$  and significantly  
 9 higher ISC in Intermediate scenes compared to Panoramic scenes,  $t(124) = -6.25, p < .001$ .  
 10 There was also a significant interaction between device and scene type,  $F(1.62,100.42) =$   
 11  $6.78, p = .002, \eta_p^2 = 0.10$ . Holm corrected post hoc tests revealed that this interaction was  
 12 driven by significantly greater ISC in Unidirectional scenes, for the HMD group compared to

1 the Tablet group,  $t(186) = 3.53, p = .001$ , and significantly greater ISC in Intermediate scenes  
 2 for the HMD group compared to the Tablet group,  $t(186) = 3.19, p = .002$ . No significant  
 3 difference was found in Panoramic scenes between HMD and Tablet groups,  $t(186) = -0.83, p$   
 4  $= .405$ .

5 ***Inter-subject Correlation by Time Period Within Scene***

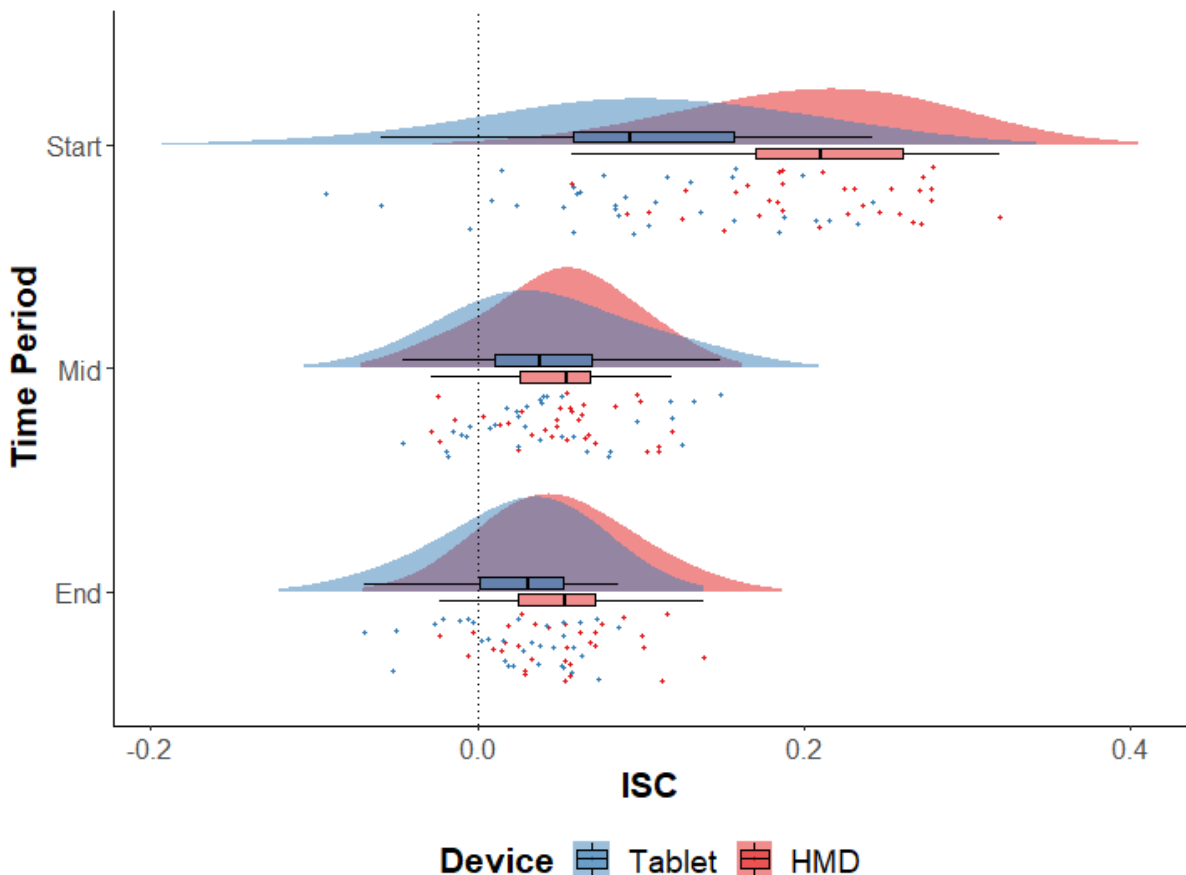
6 To investigate how the time period within scenes affected ISC, we carried out a 2x3 ANOVA  
 7 with device (HMD vs Tablet) as the between subject factor and time period (Beginning vs  
 8 Middle vs End) as the within subjects factor (see Table 5 and Figure 6).

9 Table 5. Means and standard deviations of ISC across time periods, device and measure.  
 10 Same letter superscripts represent a significant difference between devices within time period  
 11 and measure in post hoc t-tests of estimated marginal means, ‡ indicates  $p < .001$ .

Device	Time Period	Mean ISC	SD
HMD	Start	.206 <sup>‡a</sup>	.063
	Middle	.050	.040
	End	.051	.038
Tablet	Start	.101 <sup>‡a</sup>	.081
	Middle	.043	.049
	End	.024	.039

12 We found a significant effect of device,  $F(1,62) = 21.40, p < .001, \eta_p^2 = 0.26$ , which was  
 13 driven by higher ISC in the HMD group compared to the Tablet group. There was also a  
 14 significant effect of time point,  $F(1.67,103.78) = 132.68, p < .001, \eta_p^2 = 0.682$ , which was

1 driven by significantly higher ISC at the Start of scenes compared to in the Middle,  $t(189) = -$   
 2 9.71,  $p < .001$ , or at the End,  $t(189) = -10.60$ ,  $p < .001$ , of scenes. There was no significant  
 3 difference between ISC in the Middle and End of scenes,  $t(189) = -0.84$ ,  $p = .400$ . There was  
 4 also a significant interaction between device and time point,  $F(1.67, 103.78) = 21.36$ ,  $p <$   
 5  $.001$ ,  $\eta_p^2 = 0.26$ . Holm corrected post hoc tests revealed that this interaction was driven by  
 6 significantly greater ISC at the Start of scenes for the HMD group compared to the Tablet  
 7 group,  $t(186) = 7.75$ ,  $p < .001$ . By contrast, in the Middle of scenes there was no significant  
 8 difference in ISC between HMD and Tablet groups,  $t(186) = -0.53$ ,  $p = .594$ . The difference  
 9 in ISC between HMD and Tablet groups at the End of scenes approached significance,  $t(186)$   
 10  $= 1.93$ ,  $p = .055$ .



11 *Figure 6. Mean ISC across time point and device. Clouds represent distribution, raindrops*  
 12 *represent individual participants.*

## 1 *Relationship between ISC and Engagement with the Documentary*

2 To discover if engagement scores significantly differed between devices, we ran a Welch two  
 3 sample t-test comparing the HMD and Tablet groups on engagement. This revealed that  
 4 participants reported significantly higher engagement in the HMD group (Mean = 28.25, SD  
 5 = 4.67) compared with the tablet group (Mean = 23.47, SD = 7.06),  $t(61.13) = 2.89$ ,  $p = .005$ ,  
 6  $\eta^2 = 0.17$ .

7 We also investigated the relationship between engagement and ISC by running a regression  
 8 analysis with engagement as the outcome variable and z-scored GCD as the predictor. A  
 9 significant regression equation was found,  $F(1, 62) = 5.54$ ,  $p = .022$ , with an  $R^2$  of 0.08.  
 10 Participants' predicted engagement is equal to  $25.86 + 1.85$  (ISC). Engagement increased  
 11 1.85 for each SD increase in ISC.

## 12 **Discussion**

13 In this section we first discuss our findings and offer possible interpretations before moving  
 14 on to discuss some of the limitations of the current study and possible directions for future  
 15 research in this area.

16 Our analysis of pitch and yaw data revealed significant differences in mean directions ( $\theta$ ) for  
 17 both the pitch and yaw axis with significantly greater distribution of pitch and yaw  $\theta$  in the  
 18 Tablet group compared to the HMD group. This finding suggests that participants in the  
 19 HMD group were more cohesive in their mean viewing direction while those in the Tablet  
 20 group centred their viewing around a wider variety of points. In the yaw condition, the  
 21 greater distribution of  $\theta$  in the Tablet group may be due to the fact that the tablet was wireless  
 22 - allowing full unencumbered rotational movement in 360° - whereas the HMD required a

1 cable connection to the testing laptop which may have led users to avoid excessive yaw  
2 rotation due to fear of becoming tangled in or pulling out the cable. When we converted pitch  
3 and yaw into a single measure of great circle distance however there were no significant  
4 differences between the groups in either their mean distance or the variance in mean distance.  
5 This suggests that, while participants' mean viewing direction varied more in the Tablet than  
6 the HMD group, the average amount of distance that participants' viewpoints moved from the  
7 reference point was not significantly different between devices.

8 The overall inter-subject correlation (ISC) analysis revealed that ISC across both groups was  
9 significantly different from 0 indicating that the piece did elicit a level of attentional  
10 synchrony in both groups. Further, participants exhibited higher ISC in the HMD as  
11 compared to the Tablet condition. However, it should be noted that, while ISC in both  
12 conditions were significantly above 0, even the HMD ISC values were considerably lower  
13 than those found between adult humans in previous studies that investigated ISC for standard  
14 video watching (.219 for HMD in the current study compared to .597 in Franchak et al.,  
15 2016; and .39 in Shepherd et al., 2010). This reduction in overall ISC may in part be due to  
16 the fact that, while previous studies directly measured gaze from eye tracking, the current  
17 study relied on head movement data alone. It may also, however, reflect the additional  
18 challenges that 360° video creates for coherent storytelling.

19 In addition to examining attentional synchrony (AS) over the entire time course we also  
20 carried out two additional analyses. The first of these examined whether differences in the  
21 directionality of attentional cues in each scene encouraged greater AS. This analysis found  
22 that - perhaps unsurprisingly - participants showed the highest level of AS for scenes that had  
23 a single direction of focus, with decreased AS in intermediate scenes and the lowest value for  
24 panoramic scenes that lacked a clear direction. We also found an interaction between device

1 and scene, in which participants watching using an HMD showed higher AS than those  
2 watching using a Tablet for unidirectional and intermediate scenes but not for panoramic  
3 scenes. These findings suggest that HMD users are more responsive to directional cues within  
4 a scene than Tablet users, which may help to explain the overall increase in AS for those  
5 using HMDs. It is notable that in many of the directional scenes, the item of interest was a  
6 person or group of people which previous studies have shown act as strong attentional  
7 attractors (Birmingham et al., 2008, 2009; Williams et al., 2018).

8 As well as investigating how the spatial properties of a scene affected AS we also  
9 investigated the influence of temporal properties on AS by comparing the average ISC of 10  
10 second periods at the start, middle and end of scenes. We found significantly greater AS at  
11 the beginning of scenes compared to in the middle or at the end. Again, this effect interacted  
12 with device type with a stronger effect of time point on AS in our HMD group compared to  
13 our Tablet group. The overall pattern of increased AS at the start of scenes is similar to  
14 previous findings from 2D film which suggest that AS declines over the course of a scene  
15 (Dorr et al., 2010). One explanation for this effect is evidence that, when viewing a new static  
16 or dynamic scene, gaze behaviour can be divided into two phases (Eisenberg & Zacks, 2016;  
17 Unema et al., 2005). First there is an ambient period, characterised by a reliance on input  
18 from peripheral vision and the rapid acquisition of low frequency information. This is  
19 followed by a focal period characterised by longer fixation on areas of high frequency  
20 information. It is likely that AS is higher during the more feature driven ambient period due  
21 to less room for individual differences in interests and motivations to interfere with the  
22 capture of visual attention by salient aspects of the scene.

23 There are several explanations as to why watching 360° videos on a 2D display as opposed to  
24 HMD might act to weaken AS. First, unlike HMDs, which fully immerse the viewer within a

1 visual scene, 2D displays do not fully block out the external environment, which in the  
2 current case contained a significant number of different objects and a wide view of part of the  
3 Bath University campus. These additional visual features that were present for participants in  
4 the Tablet condition may have made them more likely to have been influenced by visual  
5 information not within the 360° video (Neumann & Moffitt, 2018) than were participants in  
6 the HMD condition.

7 Second, the full visual immersion offered by HMDs means that more of the scene is seen in  
8 peripheral vision compared to a flat screen set up. Since peripheral vision is highly sensitive  
9 to motion cues (Strasburger et al., 2011) which have been shown to influence attentional  
10 control (Inoue et al., 2017), it is possible that viewers in a HMD will be more likely to  
11 respond to peripheral cues than will participants in a flat screen condition. The fact that our  
12 effects seemed to be particularly strong at the start of scenes suggests that the greater AS  
13 observed in the HMD group during these times is driven by access to saliency cues in  
14 peripheral vision that act to guide viewer's attention toward salient features of the scene. On  
15 this account, greater access to these saliency cues in the HMD group would lead participants'  
16 attention to move in a more synchronous manner during the early exploratory period within  
17 each scene (Serrano et al., 2018). Similarly increased peripheral vision might aid participants  
18 using an HMD in locating the salient areas within directional scenes.

19 Our final analysis investigated the relationship between AS and participants' rating of their  
20 engagement with the documentary. A linear regression found that ISC was a positive  
21 predictor of participants' scores on the engagement questionnaire, suggesting that participants  
22 whose viewpoints overlapped more with those of their fellow participants found the  
23 documentary more engaging. This finding is important as it indicates that our measures of AS



1 substantially affect viewers' perceptions of the 360 piece and that increasing the amount of  
2 AS found among viewers could increase viewer enjoyment.

3 In terms of implications our findings suggest several possible strategies for content producers  
4 seeking to maximise AS between viewers in VR. First, the finding of stronger AS for  
5 unidirectional compared to panoramic scenes suggests the value of avoiding visual “clutter”  
6 in non-relevant areas for scenes in which creators intend there to be only one area of interest.  
7 Second the findings of stronger AS at the beginning compared to the middle and end of  
8 scenes suggest that the content creator could benefit from adding scene transitions just prior  
9 to times when they want viewers’ attention to shift to a new location in the visual scene.

## 10 **Limitations**

11 Our study had several limitations in its design. First, we measured attentional direction based  
12 on the central point of the participant’s view rather than examining actual fixation data using  
13 eye-tracking. This means that we cannot be sure that our participants were fixating on the  
14 coordinates of the tracking data as opposed to moving their eyes to a different point within  
15 the visual scene. However, it should be noted that Sitzmann and colleagues (2018)  
16 investigated the amount of coupling between head and eye movements and found evidence of  
17 a strong correlation with head movements following eye gaze direction with an average delay  
18 of 58ms. They further demonstrated that head movement data alone predicted saliency as  
19 well gaze based predictors when viewing 360° scenes. These findings suggest that, at least  
20 within an HMD, viewing direction is a suitable proxy for eye-gaze.

21 A second limitation of our study is that, as mentioned above, there were significant  
22 differences in the affordances of the two devices. The presence of the cable attaching the  
23 HMD to the laptop may have restrained participant movement in the yaw axis. This problem

1 could be overcome by use of the new generation of wireless headsets to test AS in future  
2 studies. A related issue is that the “magic window” set up of the tablet is not a particularly  
3 naturalistic way to watch 360° video on a tablet and may have limited viewers’ exploratory  
4 behaviour. On the other hand, had we allowed participants to explore the scene using finger  
5 movements (as most 360° video viewers for flat screens do) then there would have been even  
6 greater differences in visuo-motor affordances between the two devices than in the present  
7 setup. We would also note that the analyses showing modulation of AS by both the temporal  
8 and spatial properties of scenes were strongest in the HMD condition meaning that these  
9 findings have relevance for those working in VR content production independently of the  
10 comparison with other non-HMD formats.

11 A final limitation of this study is that it is not clear the extent to which the different degrees  
12 of ISC seen in our results translate into actionable intelligence about the coherence of scenes  
13 or the effectiveness of attentional cueing. While this question will require further research  
14 looking specifically at ISC in 360 video to answer, we note that the positive relationship we  
15 found between ISC and engagement suggests that higher AS can positively affect viewers’  
16 perceptions of 360 video. We also note that although our overall levels of ISC were lower  
17 than those found in previous studies that used ISC to study AS of gaze direction the  
18 differences in ISC found between the HMD and Tablet devices and between scene types and  
19 time periods on the HMD conditions were large suggesting that these factors are likely to  
20 play a meaningful role in modulating AS.

## 21 **Future Directions**

22 While our study offers an early step towards understanding AS in 360° video, there are  
23 several directions that future research could take. One obvious advance on the current study  
24 would be to directly collect gaze data in addition to head/tablet orientation in order to have a

1 more fine-grained measure of AS. Another potential future direction would be to manipulate  
2 the aspects of different scenes or pieces more systematically. For example, would a  
3 deliberately structured 360° video lead to stronger AS than an unstructured one? Do we see  
4 the same pattern of AS when observing landscape scenes as opposed to scenes populated with  
5 people? Another approach would be to use reverse correlation analysis (Hasson et al., 2004)  
6 in order to extract the time points that led to the greatest ISC. If this approach was combined  
7 with more quantitative measures of scene complexity and annotation it might yield additional  
8 insights into the scene features that promote AS in 360° video.

9 Another question is how much the co-presence of viewers affects AS. Golland and colleagues  
10 (Golland et al., 2015) showed that watching a 2D film with another person increased the  
11 amount of ISC in physiological measures. While HMDs are a particularly solipsistic medium,  
12 the growing interest in VR cinema (Kil, 2018) raises the question of how much co-presence  
13 with others acts to synchronise attention. There are also questions as to how AS is related to  
14 measures of neural and physiological synchrony and to viewer enjoyment of (and immersion  
15 in) the piece, all of which are ripe for further investigation.

## 16 **Conclusion**

17 The rapid uptake of VR by producers of nonfiction reflects a fascination with the potential of  
18 360 media for viewer engagement. While VR allows viewers a novel freedom to find their  
19 own visual pathway through a given scene, at the same time producers have particular  
20 information that they need to convey. The power to elicit AS is often held up as a key  
21 indicator of a content creator's ability to lead an audience through a narrative. In this paper  
22 we have shown that 360° video can elicit a degree of AS between viewers and that this effect  
23 is increased when the video is viewed via a HMD. This effect was greater for scenes with a  
24 clear direction of focus than for more panoramic scenes and greater at the onset of scenes

1 than at the middle or end. These insights may aid content producers as to how to ensure that  
2 they structure scenes such that viewers are most likely to see crucial features. Finally, we  
3 hope that showing the feasibility of this method of analysis encourages other researchers to  
4 consider applying a similar analysis to their own 360° video tracking data and help to develop  
5 an empirically informed “screen grammar” for virtual reality.

## 6 **Online Material**

7 The raw datafile and the R scripts used for the analyses reported in this study are available  
8 online at <https://osf.io/axgb5/>.

## 9 **References**

- 10 Adolphs, R., Nummenmaa, L., Todorov, A., & Haxby, J. V. (2016). Data-driven approaches  
11 in the investigation of social perception. *Philosophical Transactions of the Royal Society*  
12 *B: Biological Sciences*, 371(1693). <https://doi.org/10.1098/rstb.2015.0367>
- 13 Agostinelli, C., & Lund, U. (2017). R package “circular”: Circular Statistics (version 0.4-  
14 93) (0.4-93). <https://r-forge.r-project.org/projects/circular/>
- 15 Bender, S. (2018). Headset attentional synchrony: Tracking the gaze of viewers watching  
16 narrative virtual reality. *Media Practice and Education*.  
17 <https://doi.org/10.1080/25741136.2018.1464743>
- 18 Bevan, C., & Green, D. P. (2017). A mediography of virtual reality non-fiction: insights and  
19 future directions. *Proceedings of the 2018 ACM International Conference on Interactive*  
20 *Experiences for TV and Online Video (TVX '18)*, 161–166.  
21 <https://doi.org/10.1145/3210825.3213557>

- 1 Bevan, C., Green, D. P., Farmer, H., Rose, M., Cater, K., Stanton Fraser, D., & Brown, H.  
 2 (2019). Behind the curtain of the “ultimate empathy machine”: On the composition of  
 3 virtual reality nonfiction experiences. *Proceedings of the 2019 CHI Conference on*  
 4 *Human Factors in Computing Systems*, 509. <https://doi.org/10.1145/3290605.3300736>
- 5 Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Saliency does not account for  
 6 fixations to eyes within social scenes. *Vision Research*, 49(24), 2992–3000.  
 7 <https://doi.org/10.1016/j.visres.2009.09.014>
- 8 Birmingham, E., Bischof, W., & Kingstone, A. (2008). Gaze selection in complex social  
 9 scenes. *Visual Cognition*, 16(2–3), 341–355.  
 10 <https://doi.org/10.1080/13506280701434532>
- 11 Bracken, B. K., Alexander, V., Zak, P. J., Romero, V., & Barraza, J. A. (2014). Physiological  
 12 synchronization is associated with narrative emotionality and subsequent behavioral  
 13 response. In D. D. Schmorow & F. C. M (Eds.), *Foundations of Augmented Cognition.*  
 14 *Advancing Human Performance and Decision-Making through Adaptive Systems. AC*  
 15 *2014. Lecture Notes in Computer Science, vol 8534. Springer, Cham* (pp. 3–13).  
 16 Springer. [https://doi.org/https://doi.org/10.1007/978-3-319-07527-3\\_1](https://doi.org/https://doi.org/10.1007/978-3-319-07527-3_1)
- 17 Breathnach, D. (2016). Attentional synchrony and the effects of repetitive movie viewing.  
 18 *CEUR Workshop Proceedings, 1751*, 260–271.
- 19 Burleson-Lesser, K., Morone, F., DeGuzman, P., Parra, L. C., & Makse, H. A. (2017).  
 20 Collective behaviour in video viewing: A thermodynamic analysis of gaze position.  
 21 *PLoS ONE*, 12(1), 1–19. <https://doi.org/10.1371/journal.pone.0168995>
- 22 Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in  
 23 dynamic scenes. *Vision Research*, 46(26), 4333–4345.

- 1        <https://doi.org/10.1016/j.visres.2006.08.019>
- 2        consultancy.uk. (2018). *Virtual and augmented reality market to boom to \$170 billion by*  
3        2022.    [https://www.consultancy.uk/news/17876/virtual-and-augmented-reality-market-](https://www.consultancy.uk/news/17876/virtual-and-augmented-reality-market-to-boom-to-170-billion-by-2022)  
4        [to-boom-to-170-billion-by-2022](https://www.consultancy.uk/news/17876/virtual-and-augmented-reality-market-to-boom-to-170-billion-by-2022)
- 5        Coutrot, A., & Guyader, N. (2014). How saliency, faces, and sound influence gaze in  
6        dynamic social scenes. *Journal of Vision*, 14(8), 1–17. <https://doi.org/10.1167/14.8.5>
- 7        Cremers, J., & Klugkist, I. (2018). One direction? A tutorial for circular data using R with  
8        examples in cognitive psychology. *Frontiers in Psychology*, 9, Article 2040.  
9        <https://doi.org/10.3389/FPSYG.2018.02040>
- 10       Dmochowski, J. P., Bezdek, M. A., Abelson, R. P., Johnson, J. S., Schumacher, E. H., &  
11       Parra, L. C. (2014). Audience preferences are predicted by temporal reliability of neural  
12       processing. *Nature Communications*, 5, 1–9. <https://doi.org/10.1038/ncomms5567>
- 13       Dooley, K. (2017). Storytelling with virtual reality in 360-degrees: A new screen grammar.  
14       *Studies in Australasian Cinema*, 11(3), 161–171.  
15       <https://doi.org/10.1080/17503175.2017.1387357>
- 16       Dorr, M., Martinetz, T., Gegenfurtner, K. R., & Barth, E. (2010). Variability of eye  
17       movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10), 28–28.  
18       <https://doi.org/10.1167/10.10.28>
- 19       Eco VR. (2017). *Virtual Nature 360° - 5K Nature Meditation for Daydream*, Oculus, Gear  
20       VR. [https://www.youtube.com/watch?v=rG4jSz\\_2HDY](https://www.youtube.com/watch?v=rG4jSz_2HDY)
- 21       Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic  
22       activity. *Journal of Vision*, 16(2), 5. <https://doi.org/10.1167/16.2.5>

- 1 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical  
2 power analysis program for the social, behavioral, and biomedical sciences. *Behavioural*  
3 *Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- 4 Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze  
5 behavior in infants and adults. *Infancy*, 25(3), 289–313.  
6 <https://doi.org/110.1016/j.bbi.2017.04.008>
- 7 Golland, Y., Arzouan, Y., & Levit-Binnun, N. (2015). The mere co-presence:  
8 Synchronization of autonomic signals and emotional responses across co-present  
9 individuals not engaged in direct interaction. *PLoS ONE*, 10(5), Article e0125804.  
10 <https://doi.org/10.1371/journal.pone.0125804>
- 11 Golland, Y., Keissar, K., & Levit-Binnun, N. (2014). Studying the dynamics of autonomic  
12 activity during emotional experience. *Psychophysiology*, 51(11), 1101–1111.  
13 <https://doi.org/10.1111/psyp.12261>
- 14 Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced intersubject  
15 correlations during movie viewing correlate with successful episodic encoding. *Neuron*,  
16 57(3), 452–462. <https://doi.org/10.1016/j.neuron.2007.12.009>.Enhanced
- 17 Hasson, U., Landesman, O., Knappmeyer, B., Vallines, I., Rubin, N., & Heeger, D. J. (2008).  
18 Neurocinematics: The neuroscience of film. *Projections*, 2(1), 1–26.  
19 <https://doi.org/10.3167/proj.2008.020102> Volume
- 20 Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject  
21 synchronization of cortical activity during natural vision. *Science*, 303, 1634–1640.  
22 <https://doi.org/10.1126/science.1089506>

- 1 Herbec, A., Kauppi, J. P., Jola, C., Tohka, J., & Pollick, F. E. (2015). Differences in fMRI  
 2 intersubject correlation while viewing unedited and edited videos of dance performance.  
 3 *Cortex*, 71, 341–348. <https://doi.org/10.1016/j.cortex.2015.06.026>
- 4 Hijmans, R., Williams, E., & Vennes, C. (2019). *Package ‘ geosphere ’* (1.5-10).  
 5 <https://CRAN.R-project.org/package=geosphere>.
- 6 HTC. (2016). *Vive*. <https://www.vive.com>
- 7 Hutson, J. P., Smith, T. J., Magliano, J. P., & Loschky, L. C. (2017). What is the role of the  
 8 film viewer? The effects of narrative comprehension and viewing task on gaze control in  
 9 film. *Cognitive Research: Principles and Implications*, 2, Article 46.  
 10 <https://doi.org/10.1186/s41235-017-0080-5>
- 11 Inoue, Y., Tanizawa, T., Utsumi, A., Susami, K., Kondo, T., & Takahashi, K. (2017). Visual  
 12 attention control using peripheral vision stimulation. *2017 IEEE International  
 13 Conference on Systems, Man, and Cybernetics, SMC 2017, 2017-Janua*, 1363–1368.  
 14 <https://doi.org/10.1109/SMC.2017.8122803>
- 15 Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in  
 16 dynamic scenes. *Visual Cognition*, 12(6), 1093–1123.  
 17 <https://doi.org/10.1080/13506280444000661>
- 18 Kil, S. (2018). *Virtual Reality Comes to Mainstream Cinema in South Korea*. *Variety*.  
 19 <https://variety.com/2018/biz/news/vr-virtual-reality-gina-kim-1202801584-1202801584/>
- 20 Kirkorian, H. L., & Anderson, D. R. (2018). Effect of sequential video shot  
 21 comprehensibility on attentional synchrony: A comparison of children and adults.  
 22 *Proceedings of the National Academy of Sciences*, 115(40), 9867–9874.



- 1        <https://doi.org/10.1073/pnas.1611606114>
- 2        Lankinen, K., Saari, J., Hari, R., & Koskinen, M. (2014). Intersubject consistency of cortical  
3        MEG signals during movie viewing. *NeuroImage*, *92*, 217–224.  
4        <https://doi.org/10.1016/j.neuroimage.2014.02.004>
- 5        Loschky, L. C., Larson, A. M., Magliano, J. P., & Smith, T. J. (2015). What would Jaws do?  
6        The tyranny of film and the relationship between gaze and higher-level narrative film  
7        comprehension. *PLoS ONE*, *10*(11), 1–23. <https://doi.org/10.1371/journal.pone.0142474>
- 8        MacQuarrie, A., & Steed, A. (2017). Cinematic virtual reality: Evaluating the effect of  
9        display type on the viewing experience for panoramic video. *2017 IEEE Virtual Reality*  
10        (*VR*), 45–54. <https://doi.org/10.1109/VR.2017.7892230>
- 11        Margerie, T. de. (2018). *How to integrate 360 video with Unity*.  
12        [https://blogs.unity3d.com/2017/07/27/how-to-integrate-360-video-with-](https://blogs.unity3d.com/2017/07/27/how-to-integrate-360-video-with-unity/?_ga=2.159851645.1308616920.1543498780-421605029.1517835635#)  
13        [unity/?\\_ga=2.159851645.1308616920.1543498780-421605029.1517835635#](https://blogs.unity3d.com/2017/07/27/how-to-integrate-360-video-with-unity/?_ga=2.159851645.1308616920.1543498780-421605029.1517835635#)
- 14        Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during  
15        dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*(1), 5–24.  
16        <https://doi.org/10.1007/s12559-010-9074-z>
- 17        Neumann, D., & Moffitt, R. (2018). Affective and attentional states when running in a virtual  
18        reality environment. *Sports*, *6*(3), 71. <https://doi.org/10.3390/sports6030071>
- 19        Oculus. (2016). *Oculus Rift*. <https://www.oculus.com/rift>
- 20        Oculus. (2019). *Oculus Quest*. <https://www.oculus.com/quest/>
- 21        Ozcinar, C., & Smolic, A. (2018). Visual attention in omnidirectional video for virtual reality

- 1 applications. *2018 Tenth International Conference on Quality of Multimedia Experience*  
2 (*QoMEX*), 1–6. <https://doi.org/10.1109/QoMEX.2018.8463418>
- 3 Passmore, P. J., Glancy, M., Philpot, A., Roscoe, A., Wood, A., & Fields, B. (2016). Effects  
4 of viewing condition on user experience of panoramic video. In D. Reiners, D. Iwai, &  
5 F. Steinicke (Eds.), *Proceedings of the 26th International Conference on Artificial*  
6 *Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments*  
7 (pp. 9–16). The Eurographics Association. <https://doi.org/10.2312/egve.20161428>
- 8 Pope, V. C., Dawes, R., Schweiger, F., & Sheikh, A. (2017). The geometry of storytelling:  
9 Theatrical use of space for 360-degree videos and virtual reality. *Proceedings of the*  
10 *2017 CHI Conference on Human Factors in Computing Systems - CHI '17*, 4468–4478.  
11 <https://doi.org/10.1145/3025453.3025581>
- 12 Poulsen, A. T., Kamronn, S., Dmochowski, J., Parra, L. C., & Hansen, L. K. (2017). EEG in  
13 the classroom: Synchronised neural recordings during video presentation. *Scientific*  
14 *Reports*, 7, 1–9. <https://doi.org/10.1038/srep43916>
- 15 Rubo, M., & Gamer, M. (2018). Social content and emotional valence modulate gaze  
16 fixations in dynamic scenes. *Scientific Reports*, 8(1), 1–11.  
17 <https://doi.org/10.1038/s41598-018-22127-w>
- 18 Samsung. (2015). *Gear VR*. <https://www.samsung.com/global/galaxy/gear-vr/>
- 19 Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality.  
20 *Motivation and Emotion*, 41(6), 708–712. <https://doi.org/10.1007/s11031-017-9641-7>
- 21 Serrano, A., Sitzmann, V., Ruiz-Borau, J., Wetzstein, G., Gutierrez, D., & Masia, B. (2018).  
22 Movie editing and cognitive event segmentation in virtual reality video. *ACM*

- 1        *Transactions on Graphics*, 36(4), 47. <https://doi.org/10.1145/3072959.3073668>
- 2        Shepherd, S. V, Steckenfinger, S. A., Hasson, U., & Ghazanfar, A. A. (2010). Human–  
3        monkey gaze correlations reveal convergent and divergent patterns of movie viewing.  
4        *Current Biology*, 20(7), 649–656. <https://doi.org/10.1016/j.cub.2010.02.032>. Human
- 5        Silver, N. C., & Dunlap, W. P. (1987). Averaging Correlation Coefficients: Should Fisher’s z  
6        Transformation Be Used? *Journal of Applied Psychology*, 72(1), 146–148.  
7        <https://doi.org/10.1037/0021-9010.72.1.146>
- 8        Sitzmann, V., Serrano, A., Pavel, A., Agrawala, M., Gutierrez, D., Masia, B., & Wetzstein,  
9        G. (2018). Saliency in VR: How do people explore virtual environments? *IEEE*  
10        *Transactions on Visualization and Computer Graphics*, 24(4), 1633–1642.  
11        <https://doi.org/10.1109/TVCG.2018.2793599>
- 12        Smith, T. J. (2013). Watching you watch movies: Using eye tracking to inform cognitive film  
13        theory. In A. P. Shimamura (Ed.), *Psychocinematics* (pp. 165–192). Oxford University  
14        Press. <https://doi.org/10.1093/acprof:oso/9780199862139.003.0009>
- 15        Smith, T. J., Levin, D., & Cutting, J. E. (2012). A window on reality: Perceiving edited  
16        moving images. *Current Directions in Psychological Science*, 21(2), 107–113.  
17        <https://doi.org/10.1177/0963721412437407>
- 18        Smith, T. J., & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task  
19        on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 1–24.  
20        <https://doi.org/10.1167/13.8.16>
- 21        Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern  
22        recognition: A review. *Journal of Vision*, 11(5), 13–13. <https://doi.org/10.1167/11.5.13>

- 1 Subramanian, R., Shankar, D., Sebe, N., & Melcher, D. (2014). Emotion modulates eye  
2 movement patterns and subsequent memory for the gist and details of movie scenes.  
3 *Journal of Vision*, 14(3), 31–31. <https://doi.org/10.1167/14.3.31>
- 4 Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of  
5 information processing during scene perception: The relationship between saccade  
6 amplitude and fixation duration. *Visual Cognition*, 12(3), 473–494.  
7 <https://doi.org/10.1080/13506280444000409>
- 8 VRSE.works. (2015). *Clouds over Sidra*.  
9 <https://www.youtube.com/watch?v=mUosdCQsMkM&t=3s>
- 10 Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye  
11 movement strategies during naturalistic viewing. *Journal of Vision*, 12(1), 16–16.  
12 <https://doi.org/10.1167/12.1.16>
- 13 Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video  
14 game-based environments: Investigation of the User Engagement Scale. *Computers in*  
15 *Human Behavior*, 32, 123–132. <https://doi.org/10.1016/j.chb.2013.12.001>
- 16 Wikipedia. (2020). *Great-circle distance*. Wikipedia. [https://en.wikipedia.org/wiki/Great-](https://en.wikipedia.org/wiki/Great-circle_distance)  
17 [circle\\_distance](https://en.wikipedia.org/wiki/Great-circle_distance)
- 18 Williams, E. H., Cristino, F., & Cross, E. S. (2018). Human body motion captures visual  
19 attention and elicits pupillary arousal. *Cognition*, 193(November), 104029.  
20 <https://doi.org/10.1016/j.cognition.2019.104029>
- 21 Zoric, G., Barkhuus, L., Engström, A., & Örnevall, E. (2013). Panoramic video: Design  
22 challenges and implications for content interaction. *Proceedings of the 11th European*

1 *Conference on Interactive TV and Video (EuroITV '13)*, 153–162.

2 <https://doi.org/10.1145/2465958.2465959>

3

4