Did You See What I Saw?: Comparing User Synchrony When Watching 360° Video In HMD Vs Flat Screen

Harry Farmer*, University of Bath, Bath, UK
Chris Bevan, University of Bristol, Bristol, UK
David P. Green, University of the West of England Bristol, UK
Mandy Rose, University of the West of England Bristol, UK
Kirsten Cater, University of Bristol, Bristol, UK
Danaë Stanton-Fraser, University of Bath, Bath, UK

Abstract

This study examined whether the high level of immersion provided by HMDs encourages participants to synchronise their attention during viewing. 39 participants watched the 360° documentary "Clouds Over Sidra" using either a HMD or via a flat screen tablet display. We found that the HMD group showed significantly greater overall ISC did the tablet group and that this effect was strongest during transition between scenes.

Keywords: 360° video, synchrony, inter-subject correlation analysis.

Index Terms: Topic Area #1 [Technologies & Applications] Topic Area #3 [Interaction]

1 Introduction

The rapid advances in virtual reality (VR) technology in the last few years have led to increased interest in VR's potential as a means of delivering narrative media. Since late 2015 the release of affordable, high specification head mounted display (HMD) systems has driven a vast increase in VR content.

At present the most common current form of VR content is 360° video in which footage from a series of camera's set in a circular array are stitched together to form one panoramic scene. However, the omnidirectional nature of 360° video inevitably reduces the amount of control media producers have over the direction of viewer's attention leading for calls for the development of "a new screen grammar" [1]. Greater knowledge of how viewers are likely to look in VR could also help to optimise the storage, transmission and rendering of 360° videos [4]. In this context an increased understanding of how people visually navigate through 360° videos has the potential to yield important insights for both content producers and tech developers working in VR.

2 Method

2.1 Participants

Thirty-nine participants (mean age ± SD: 28.5 ± 13.2 years; 15 males) gave their written informed consent and were paid for their participation. Nineteen of our participants were assigned to the HMD group and 20 to the Tablet group. Ethical approval for the study was given by the University of Bath, Department of Psychology Research Ethics Committee.

2.2 Design

The experiment had a between subjects design with one independent variable which was the device used to watch the 360° video content (HMD vs Tablet). During the study we measured pitch and yaw rotational data, captured from the onboard orientation sensors of the HMD and Tablet.

2.3 Materials

To habituate participants to the affordances of the viewing device they were using all participants first viewed an introductory 2:20 minute video which shows scenes of nature. Following this tracking data was recorded from participants as they watched the 8:35 minute 360° documentary 'Clouds Over Sidra' [7].

In both setups, participant's viewed the 360° videos using a custom-built 360° video player application that was created by the researchers using the Unity software development package (version 2017.3.1f1). Playback of 360° video content was facilitated using the in-built video player component provided by Unity [2].

2.4 Procedure

Participants viewed the videos while seated on a swivel chair that allowed easy rotation in the yaw dimension. At the end of the experiment participants were thoroughly debriefed and given a detailed sheet explaining the purpose of the experiment.

2.5 Data Analysis

2.5.1 Pre-processing

For every rendered frame of playback, the current pitch and yaw orientation of the main scene camera was captured as an Euler angle and logged in a text file. This resulted in a sampling rate of 90Hz for the HMD, and a slightly reduced rate of 60Hz for the less powerful Tablet computer. Each frame was further labelled with the current time elapsed (msec) and the frame number. Orientation values were captured from the viewport camera using the transform class provided by Unity. In both the HMD and tablet versions, the sampling rate of the viewport comfortably exceeded the frame rate of the video. Prior to analysis, tracking data was down-sampled to a rate of 10Hz to remove redundant and duplicated data points. Since our main interest was how participants synchronised their head movements while watching naturalistic scenes, we then further cut the time series to remove the opening and closing credits leaving time series with a total time of 7:16.2 minutes, starting at 9.1 seconds into the original video and ending at 7:25.3 minutes into the video.

2.5.2 Analysis of Overall ISC

In order to calculate ISC, we used the R package "circular" to separately calculate each participants pairwise circular correlations in yaw and pitch for each of the two device groups. We then calculated the mean of each participant's pairwise fisher-transformed, z-normalised coefficients. Finally we transformed the mean z-transformed values back into r values to gain the mean correlation coefficient of each participants tracking data with all other participants in their group, i.e. their ISC.

*h.farmer@bath.ac.uk

3 RESULTS

3.1 ISC across the 360° Video

To discover if ISC significantly different between devices across the whole study we run two Welch two sample t-tests comparing the HMD and Tablet groups on the yaw and pitch axes respectively (See Figure 2). The t-test for the yaw axis revealed a significant difference in ISC between the two device groups $t(35.7)=4.44$, $p < .001$. The t-test for the pitch axis found ISC no significant difference in ISC between the two device groups.

3.2 Transition vs Scene Analysis

To test whether ISC different around transitions between scenes we extracted yaw and pitch data from 10 second windows around each scene transition. This gave us sixteen transition time series and so we also created fifteen 10 second time series based within the scenes and a final time series during the documentaries outro centred on 7:30.

We conducted two 2x2 mixed methods ANOVAs with device (HMD vs Tablet) as the between subject factor and transition category (Transition vs Bin) as the within subjects factor. One ANOVA investigated ISC in the yaw axis and the other in the pitch axis (see Figure 1).

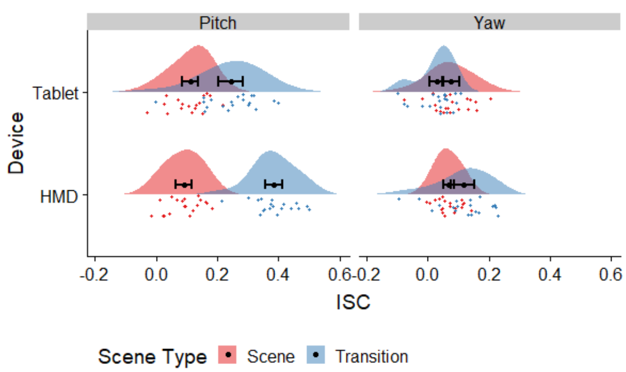

Figure 1: Overall mean ISC for the HMD and Tablet groups in the pitch and yaw axes. Raindrops represent Error bars 95% confidence intervals.

For the yaw axis one participant from the Tablet condition was found not to have moved during one of the transitions meaning that it was not possible to calculate an ISC for them in the transition condition so this data was removed from the sample. The analysis revealed a significant effect of device, $F(1,37) = 5.86$, $p = .021$, with a higher ISC for the HMD group. There was no significant effect of transition category. A significant interaction between device and scene type was found for the yaw axis, $F(1,36) = 7.99$, $p > .001$. Pairwise comparisons determined that this effect was driven by a significantly higher ISC during Transitions in the HMD group compared to the Tablet group.

The pitch axis analysis revealed significant effects of device, $F(1,37) = 12.18$, $p = .001$, and transition category, $F(1,37) = 165.85$, $p > .001$, There was also significant interaction between device and scene type in the pitch axis, $F(1,37) = 24.37$, $p > .001$. Pairwise comparisons determined that this effect was driven by a significantly higher ISC during Transition periods in the HMD group compared to the Tablet group ($EMM = 0.25$, $SE = 0.02$), $t(36) = -5.96$, $p > .001$.

4 DISCUSSION

The overall ISC analysis revealed that ISC across both group and both axes were significantly different from 0 indicating a significant amount of ISC in all four groups and that participants were had higher ISC in the yaw axis when in the HMD as compared to Tablet condition.

We also significant main effects of viewing device with greater synchrony in the HMD group and greater synchrony at transition in both the yaw and pitch axis. Finally in both axes we observed an interaction effect driven by significantly greater AS between viewing devices during transition periods but not during scene periods.

There are several explanations as to why watching 360° videos on a 2D display as opposed to HMD might act to weaken ISC. First, 2D displays do not fully block out the external environment making it more likely viewers will be influenced by distracting visual information [3]. Second, the full visual immersion offered by HMDs means that more of the scene is seen in peripheral vision. Since peripheral vision is highly sensitive to motion cues [5] it is possible that viewers in a HMD will be more likely to respond to peripheral cues than will participants in a flat screen condition. The fact that our effects seemed to be particularly strong during transition between scenes suggests that the greater AS observed in the HMD group during these times is driven by access to saliency cues in peripheral vision that act to guide viewer's attention toward salient features of the scene. On this account, greater access to these saliency cues in the HMD group would lead participants' attention to move in a more synchronous manner during the early exploratory period within each scene [6].

REFERENCES

[1] K. Dooley, "Storytelling with virtual reality in 360-degrees: A new screen grammar," Stud. Australas. Cine., vol. 11, no. 3, pp. 161–171, 2017.

[2] T. de Margerie, "How to integrate 360 video with Unity," 2018. [Online]. Available: https://blogs.unity3d.com/2017/07/27/how-to-integrate-360-video-with-unity/?_ga=2.159851645.1308616920.1543498780-421605029.1517835635#. [Accessed: 29-Nov-2018].

[3] D. Neumann and R. Moffitt, "Affective and attentional states when running in a virtual reality environment," Sports, vol. 6, no. 3, p. 71, 2018.

[4] C. Ozcinar and A. Smolic, "Visual attention in omnidirectional video for virtual reality applications," in 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), 2018, pp. 1–6.

[5] H. Strasburger, I. Rentschler, and M. Juttner, "Peripheral vision and pattern recognition: A review," J. Vis., vol. 11, no. 5, pp. 13–13, 2011.

[6] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia, "Movie editing and cognitive event segmentation in virtual reality video," ACM Trans. Graph., vol. 36, no. 4, p. 47, 2018.

[7] VRSE.works, "Clouds over Sidra," 2015. [Online]. Available: https://www.youtube.com/watch?v=mUosdCQsMkM&t=3s. [Accessed: 26-Nov-2018].