2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 979-8-3503-4839-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/VRW58643.2023.00227

Eye-tracked Evaluation of Subtitles in Immersive VR 360° Video

Marta Brescia-Zapata,[*] Universitat Autònoma de Barcelona

Krzysztof Krejtz SWPS University

Andrew T. Duchowski Clemson University

Christopher J. Hughes University of Salford

Pilar Orero Universitat Autònoma de Barcelona

Abstract

The paper presents an analysis of visual attention to subtitles within immersive media. We implement a 360° video rendering with eye movements recorded in Virtual Reality. Position and color of immersive subtitles are compared in terms of perceived task load and cognitive processing of the content. Results show that head-locked subtitles afford more focal visual inspection of the scene and presumably better comprehension. This type of in-depth analysis would not be possible without the eye movement analyses.

Index Terms: Applied computing—Arts and humanities—Psychology—Media arts

1 Introduction

Immersive technology such as Augmented or Virtual Reality (AR/VR), or collectively eXtended Reality (XR), are key technologies for the next generation of human-computer interaction [3]. 360° videos—also known as immersive or VR360 videos—are an effective way of offering immersion in VR, thanks in part to the proliferation of Head-Mounted Displays (HMDs) and omnidirectional cameras [4]. Subtitles are critical for multilingual distribution of media content [9] and for accessibility [1]. Standardized practices have been adopted largely in the context of 2D non-immersive media [10] but evaluation of subtitles in VR/AR is scarce [7] and evaluation of user gaze practically non-existent [6]. To establish novel subtitling standards in XR, user testing is required.

The present contribution is advancing on controlled experiments within a recently developed framework for evaluating subtitled 360° videos in VR [2] using triangulation of metrics, psychophysiological process metrics (eye movements), performance metrics (scene comprehension), and self-reports (task-load and preferences). We present the method and results of the user study to test position (head-locked vs. fixed) and color (monochrome vs. color) of subtitles in 360° videos. In current poster in focus on effects of position. The main hypothesis stated that head-locked subtitles foster better content comprehension manifested by lower task load and lower demand on cognitive processing.

Features of 360º subtitling. Guidelines and standards for 2D non-immersive subtitles such as ISO/IEC/ITU 20071-23:2018 include recommendations regarding not only language issues, but also formatting e.g., synchronization, font size, type, face, and letter cases. Media consumption behavior in 360° is no longer linear, the user has the freedom to decide where to watch and for how long. Therefore new challenges for subtitling emerge: the position of the subtitles within the 360° space, and the identification of the sound source. To evaluate the readability of subtitles in 360° these two features must be tested using a capable framework.

2 Empirical Comparison of Subtitles

The present study tested cognitive consequences of different forms of subtitles in 360° videos for viewers. The live web testing framework [5] was ported to Unity 3D to display 360° video and to capture data from the built-in eye tracker. A new system architecture emerged, as depicted by the schematic in Figure 1(a). The system architecture was developed to utilize the HTC Vive Pro Eye, which contains a Tobii eye tracker built in to the display. The application uses two Unity assets: one optimized for recording and the other for playback. The linchpin of the architecture is a Data Manager, which stores data, handles file management, and generates output data in a variety of formats as required.

Hypotheses. Head-locked subtitles were hypothesized to foster better content comprehension manifested by lower task load and lower demand on cognitive processing. Reduced cognitive demand of head-locked subtitles was expected to facilitate more focal attention to the scene in comparison to fixed subtitles.

Study design. To test the hypotheses, the eye tracking experiment followed a 2×2 mixed design where subtitle position was a between-subjects factor and subtitle color a within-subjects factor. Subtitle position varied at two levels: fixed position (relative to the scene) or head-locked (moving with the participant). Subtitle color varied at two levels: monochrome or colored text, where color was associated with each speaking character. The two different videos with the same subtitles positioning and different color were shown to each participant in counterbalanced order (see Figure 1(b)).

Participants. Twenty-four volunteers (17f, aged $M = 33.9$ $SD = 11.18$) participated in the study. All had above average reading skills and in most were digital media savvy. They were also in favor of subtitles declaring to always turn subtitles while watching media content.

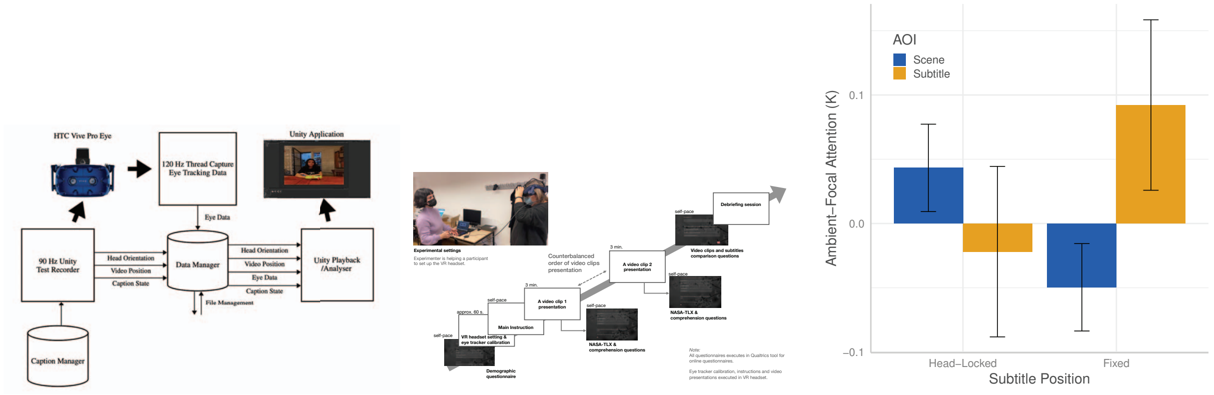
Experimental Procedure. After agreeing and signing a consent form, participants were given a demographic questionnaire which included questions on their usage and attitudes towards digital media, VR, and subtitles in media content. Next, they were familiarized with the VR headset, and then the built-in eye tracker was calibrated (see Figure 1(b) (top-left inset)). When comfortable with the VR headset, participants were presented two stimuli videos with different subtitle color (monochrome vs. color) both with subtitles presented in head-locked or fixed position depending on the experimental condition. The order of video presentation with different subtitle color and position and between-subjects condition assignment was counterbalanced. The main task for the participants was to watch two videos to get familiarized with their content and plot. Two custom recorded 360° videos served as stimuli for the experiment. The first was of a family, speaking in Arabic, discussing their vacation plans. The second was of a group of researchers introducing themselves, each speaking in their language (Spanish, Korean, Catalan, Portuguese, and English). After viewing videos participants filled out the following questionnaires: a NASA Task Load Index (NASA-TLX), subtitle readability, and video content comprehension (see Figure 1(b)).

3 Results

Comparison of subtitle display modes is broken down into analysis of perceived difficulty (task load) and cognitive processing as

[*] marta.brescia@uab.cat

979-8-3503-4839-2/23/$31.00 ©2023 IEEE
DOI 10.1109/VRW58643.2023.00227



Note:
All questionnaires executes in Qualtrics tool for online questionnaires.

Eye tracker calibration, instructions and video presentations executed in VR headset.

(a) VR system architecture

(b) Experimental procedure

(c) Ambient/focal visual attention

Figure 1: Eye-Tracking VR testing system (a) architecture (from Brescia-Zapata et al. [2]) (b) use in experimental procedure timeline and example experimental settings (top-left inset), (c) resulting in differences of ambient/focal attention over scene and subtitle Areas Of Interest.

indicated by the $\mathcal{K}$ coefficient [8] over Areas Of Interest (AOIs) representing scene and subtitle regions. Analysing subjective measures of task load we used a series of 2×2 mixed design analyses of variance (ANOVA) with two factors in terms of subtitle position (head-locked vs. fixed) and subtitle color (color vs. monochrome). Eye movement measures were analysed with a 3-way ANOVA with mixed design extended by an AOI (scene vs. subtitles) factor. All statistically significant effects were followed by pairwise comparisons with HSD Tukey correction when needed. Results of ANOVA with the NASA-TLX general score as the dependent variable revealed a statistically significant main effect of subtitle position, $F(1,21) = 6.39, p < 0.05,\ \eta^2 = 0.089$. Fixed subtitles induced higher task load ($M = 5.28, SE = 0.28$) than head-locked subtitles ($M = 4.24, SE = 0.30$). Moreover, a detailed series of ANOVAs of the same design for each subscale of the NASA-TLX revealed that fixed subtitles indicated stat. significantly greater frustration ($F(1,15) = 7.89, p < 0.02,\ \eta^2 = 0.150$), effort ($F(1,17) = 6.27, p < 0.05,\ \eta^2 = 0.072$), temporal demand ($F(1,20) = 7.84, p < 0.02,\ \eta^2 = 0.125$), and mental demand (with marginal significance). In line with the hypothesis, ANOVA of ambient/focal $\mathcal{K}$ coefficient revealed only one statistically significant effect, the interaction of AOI and subtitle position, $F(1,20) = 5.13, p < 0.05,\ \eta^2 = 0.050$ (see Figure 1(c)). Pairwise comparisons showed that scenes with head-locked subtitles were viewed with more focal attention than with fixed subtitles although the difference in $\mathcal{K}$ is marginally significant ($t(20) = 1.93, p = 0.07$).

4 Conclusion

The aim of the present study was to evaluate the readability of different forms of subtitles in 360° environments focusing on subtitle position. Results show that head-locked subtitles are easier and/or faster to process and afford more focal visual inspection of the scene, leading to increased performance in terms of content comprehension. This type of in-depth analysis of subtitle reading would not be possible without detailed examination of eye movements over subtitles and scene in 360° videos. This work is thus a landmark study of visual processing of subtitles that will hopefully lead to improved accessibility of immersive media for different groups of users and video content.

5 Acknowledgments

This work is supported in part by Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya, under the SGR funding scheme (ref. code 2017SGR113). European Union projects TRACTION (under Grant Agreement 870610), MEDIAVERSE, LEAD-ME COST Action (CA19142), the U.S. National Science Foundation (IIS-1748380, and Polish MENiSW Regional Initiative of Excellence, (012-RID-2018/19).

References

[1] C. Agrawal and R. L. Peiris. I See What You'Re Saying: A Literature Review of Eye Tracking Research in Communication of Deaf or Hard of Hearing Users. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '21. Association for Computing Machinery, New York, NY, 2021. doi: 10.1145/3441852.3471209

[2] M. Brescia-Zapata, K. Krejtz, P. Orero, A. T. Duchowski, and C. J. Hughes. VR 360° subtitles: Designing a test suite with eye-tracking technology. *Journal of Audiovisual Translation*, 5(2):233–258, 2022. doi: 10.47476/jat.v5i2.2022.184

[3] S.-C. Chen. Multimedia in Virtual Reality and Augmented Reality. *IEEE Multimedia*, 28(2), 2021.

[4] R. Du and A. Varshney. Saliency Computation for Virtual Cinematography in 360° Videos. *IEEE Computer Graphics*, 41(4), 2021.

[5] C. J. Hughes, M. Brescia-Zapata, and P. Orero. Evaluating subtitle readability in media immersive environments. In *DSAI 2020 proceedings*. Association for Computing Machinery (ACM), October 2020.

[6] D. Jain, B. Chinh, L. Findlater, R. Kushalnagar, and J. Froehlich. Exploring Augmented Reality Approaches to Real-Time Captioning: A Preliminary Autoethnographic Study. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*, DIS '18 Companion, pp. 7—11. Association for Computing Machinery, New York, NY, 2018. doi: 10.1145/3197391.3205404

[7] E. M. Klose, N. A. Mack, J. Hegenberg, and L. Schmidt. Text Presentation for Augmented Reality Applications in Dual-Task Situations. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 636–644, 2019. doi: 10.1109/VR.2019.8797992

[8] K. Krejtz, A. T. Duchowski, I. Krejtz, A. Szarkowska, and A. Kopacz. Discerning Ambient/Focal Attention with Coefficient $\mathcal{K}$. *Transactions on Applied Perception*, 13(3), 2016.

[9] K. Kurzhals, F. Göbel, K. Angerbauer, M. Sedlmair, and M. Raubal. A View on the Viewer: Gaze-Adaptive Captions for Videos. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–12. Association for Computing Machinery, New York, NY, 2020. doi: 10.1145/3313831.3376266

[10] A. Matamala and P. Orero. Standardising accessibility: Transferring knowledge to society. *Journal of Audiovisual Translation*, 1:139–154, 2018.