Detection of Strong and Weak Moments in Cinematic Virtual Reality Narration with the Use of 3D Eye Tracking

Pawel Kobylinski

Laboratory of Interactive Technologies, National Information Processing Institute

Warsaw, Poland

Grzegorz Pochwatko

Virtual Reality and Psychophysiology Lab, Institute of Psychology, Polish Academy of Sciences

Warsaw, Poland

pawel.kobylinski@opi.org.pl

gpochwatko@psych.pan.pl

Pawel Kobylinski, PhD, scientist, psychologist, data analyst

Head of the Virtual Reality Research Team

Assistant Professor

@ Laboratory of Interactive Technologies

@ National Information Processing Institute

Grzegorz Pochwatko, PhD, scientist, psychologist

Head of Virtual Reality and Psychophysiology Lab

Assistant Professor

Deputy Director for General Affairs

@ Institute of Psychology

@ Polish Academy of Sciences

WARSAW

POLAND

Introduction

(1) Cinematic Virtual Reality (CVR, 360-degree videos, watched using VR headsets) is a medium growing in popularity among both filmmakers and researchers.

(2) At this stage, some creators try to transfer old and tested narrative methods from traditional movies, while others have realized that they need to develop a new film language.

(3) Since the viewpoint has been moved to the center of the movie set, participants, not directors, are in charge of the visual focus. This means the viewers are very probable to miss an important element of a story.

(4) Therefore, one has to rely more on light, spatial sound and arrangement of the scenery while building good narration. Setting actors around the camera and acting itself must be also different and thought over well.

Introduction

(5) In order to ensure an adequate pace of development, tools are needed to conduct systematic, reliable and objective research on narration in CVR.

(6) The authors for the first time fully report results of the initial empirical test of their recently developed Scaled Aggregated Visual Attention Convergence Index (sVRCa).

(7) The quantitative index utilizes 3D Eye Tracking (3D ET) data recorded during a CVR experience. Theoretical basis of the sVRCa, among other variants of the Visual Attention Convergence Index (VRC), has been described by the authors in detail earlier. Kobylinski & Pochwatko 2019

Relation to Other Work

(1) In contrast to other attempts, the authors of the paper report a concise, timelined, near-continuous value, based solely on the measured positions of gaze fixations and computed without the need for prior computation of saliency maps or saliency optimization (neither theoretically- nor empirically-driven).

e.g. Gutiérrez-Cillán et al 2018; Upenik & Ebrahimi 2019; Sitzmann et al 2018

(2) The authors do not propose another measurement of video quality intended to improve its computational properties, neither by means of predicting nor mathematical optimization of visual attention. Moreover, the proposed method is designed to act differently than methods based on entropy measures.

e.g. Sitzmann et al 2018

(3) Low values of the utilized sVRCa index do not necessarily relate to visual attention scattered randomly around the 360-degree scene (an unrealistic scenario in the case of narrated videos). Instead, the method may detect moments in which the values of index remain low, despite the order present in the visual attention pattern when different viewers look at distinct objects located at the opposite sites of the 360-degree video (a realistic scenario).

Scaled Aggregated Visual Attention Convergence Index

(1) Values of the sVRCa index tell us if several people looked at the same or rather different virtual areas of the 360-degree scene during a chosen, short time interval.

(2) The index is based on Euclidean distances in 3D space and aggregates information about gaze fixations from a group of CVR experience participants.

(3) The values are scaled to the range between 0 and 1, which is convenient for between-experiment comparisons.

(4) The index formula takes the assumption that the index takes approximately zero value when there are only two points of viewers' focus located at a maximum possible distance from each other, at the opposite sides of the virtual scene.

Scaled Aggregated Visual Attention Convergence Index

$$sVRCa \approx 1 - \frac{\sqrt{2}}{n} \frac{\sqrt{\sum_{i,j=1}^{n} D_{ij}^{2}}}{2r} \in [0,1] \qquad (1)$$

(5) D is the n × n distance matrix calculated from 3D positions of n detected gaze fixations (corrected for the headset positions in virtual space).

(6) In the case of CVR, the 360-degree video is displayed on the inner surface of a virtual sphere. r denotes the radius of the sphere. Kobylinski & Pochwatko 2019

(7) The full procedure requires computation of the index values for subsequent short time intervals and ordering the values into time series covering the whole time span of a 360-degree video.

Scaled Aggregated Visual Attention Convergence Index

(8) The time intervals should be long enough to catch enough fixations to enable calculation of the sVRCa values and short enough to approximate the precision of continuous measurement. Half-second intervals met the assumptions in the reported test.

(9) The interpretation of the sVRCa values is relative to the intentions of a CVR maker.

(10) If the CVR designer had indented to focus people on a specific area or object at a given moment and the sVRCa peaked at a high level at that moment, it means success (provided the viewers did not focus on something else, which should be verified with the use of the same 3D ET data).

(11) If the creator had wanted the viewers to explore the scene at a given moment, high levels of the visual attention convergence at that moment mean failure and low levels mean success.

Empirical Test

(1) An educational video, aimed at a younger audience, contained a number of short scenes explaining professional work on a traditional 2D film set. The film lasted about 17 minutes.

(2) To operate the experiment (displaying the CVR video, recording 3D ET and headset position data), the VIZARD 6 ENTERPRISE was used.laying the CVR video, recording 3D ET and headset position data), the VIZARD 6 ENTERPRISE was used.

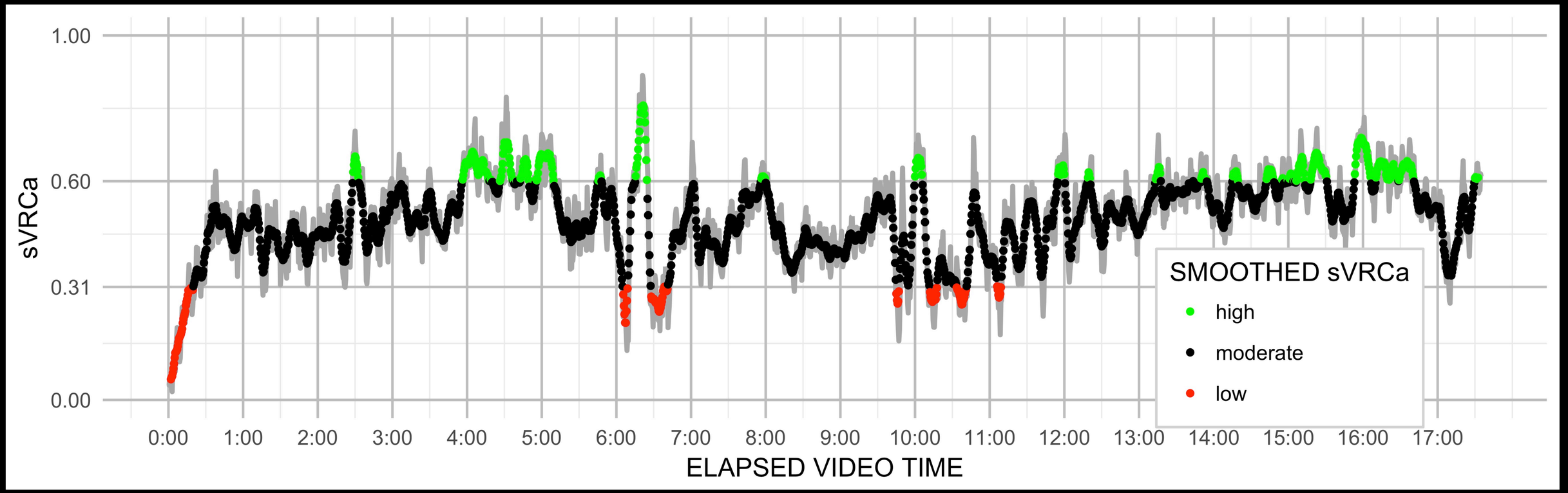
(3) HMD HTC VIVE has been equipped with a dedicated SMI eye tracker. 3D ET data were recorded synchronously with information about the location of the headset in 3D virtual space.

(4) 92 school children, 36 girls (Mage=14.1) and 56 boys, (Mage=14.2) participated in the study with the consent of a parent or a guardian. An ethical committee approved the procedure.

(5) The participants watched the 360-degree video separately, one by one, in controlled laboratory conditions. Only two people were present in the laboratory room: a participant and a trained experimenter.

Empirical Test

(6) Figure 1 illustrates the changes in the smoothed and non-filtered sVRCa values (computed for half-second intervals) over the time span of the entire CVR educational video. Such visualization enables looking at the dynamics of the visual attention convergence results from the bird's eye view and grasping the general attentional pattern shaped by the properties of the narration employed in the immersive video.



Empirical Test

(7) Table I presents chosen few examples of automatically detected high and low peaks in sVRCa time series within the detected fragments of the immersive CVR educational video.

(8) High peaks represent high levels of the visual attention convergence between the CVR participants at a given moment of the video. Low peaks represent low levels of the visual attention convergence.

(9) Colors denote values chosen for example ex post qualitative interpretation (Figures 2-4, next slides).

Video fragment begins at:	Video fragment ends at:	Median *sVRCa*	Peak *sVRCa*	Peak type	Peak at:
06:16	06:25	0.75	0.89	**max**	**06:21**
15:18	15:31	0.64	0.75	**max**	**15:23**
04:43	04:50	0.63	0.72	**max**	**04:47**
10:34	10:41	0.27	0.22	**min**	**10:40**
06:28	06:41	0.28	0.19	**min**	**06:35**
06:06	06:09	0.22	0.14	**min**	**06:09**

Empirical Test

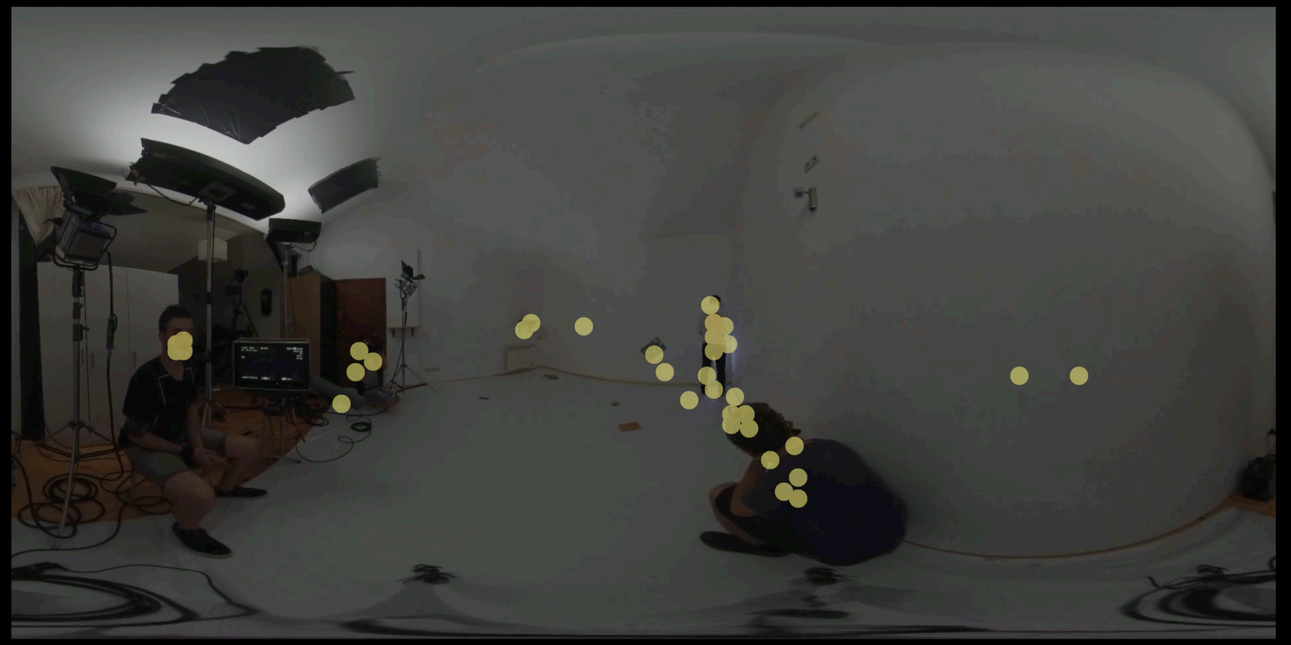
(10) The 4:47 frame (Figure 2, green in Table 1) can be interpreted as a strong moment. A high value of the sVRCa (0.72) corresponds to participants' attention focused on actors in the depicted set and on an additional screen positioned above the set.

(11) All the distractors, members of the crew, camera, equipment, etc., were hidden in low light areas (we had actually two sets here: the set of the CVR, embracing the "set" of the depicted traditional 2D movie and all the surroundings; similarly actors of the CVR included "actors" and "crew" of the depicted 2D movie).

Empirical Test

(12) Another strong moment (Figure 3) corresponds this time to a low sVRCa value (0.19).

(13) The 6:35 frame (blue in Table 1) represented a situation from the beginning of a scene. The stage was being prepared to show the role of lighting in the movies. There were many important elements around a participant. The CVR creator might have expected participants to look around the scene to figure out where the most important elements were.



Empirical Test

(14) Low sVRCa indicates weak moments as well.

(15) In the 10:40 frame (Figure 4, red in Table 1, index value: 0.22), instead of focusing subsequently on elements of lighting being described by the lector one at a time, participants focused on all of the elements of lighting and other, unrelated elements, like moving camera, members of the crew, etc.

Empirical Test

(16) Ideally, it is a CVR creator who should decide the qualitative interpretation of the quantitative measurement!

Conclusion

(1) The initial results of the test are promising. The method seems to substantially augment the process of detection of strong and weak moments in CVR narration.

(2) It delivers both the bird’s eye view on the changes in reaction to narration and detailed information allowing either automated or point-by-point analysis of specific cuts, fragments, and moments in the immersive 360-degree video.

(3) The authors propose a human-oriented measure, values of which reflect effectiveness of the process of attention directing along a narration line intended by a CVR experience creator.

(4) The method measures empirically the inter-viewer convergence in visual attention in order to give CVR creators feedback regarding whether they managed to converge the visual attention or whether they managed to dissipate it at any given moment of the video, according to their original creative intentions.

Conclusion

(5) From a purely technology-oriented perspective, the need for the qualitative interpretation of the quantitative sVRCa values might be perceived as a limitation of the proposed method.

(6) However, the authors stand on the ground that, in the case of artistic pursuits, it is an artist, not a software system (not even a scientist), who should stay free to draw final conclusions from scientific data and be responsible for all the decisions as to the changes in the narration.

(7) The introduction of methodology, such as proposed and described in the paper, seems necessary to help CVR makers in the process of development and validation of the emerging CVR narration language and means of expression. Pillai & Verma 2019

(8) The authors hope the method could serve not only as feedback for CVR creators, but also as a criterion for other visual attention measures, physiological indices of attention (e.g., Heart Rate Variability (HRV)) or declarative, quantitative, and qualitative measures.

Thank you for your (visual) attention

pawel.kobylinski@opi.org.pl

gpochwatko@psych.pan.pl

Kobylinski P., Pochwatko G. (2020) Detection of Strong and Weak Moments in Cinematic Virtual Reality Narration with the Use of 3D Eye Tracking. In: Mauri J.L., Saplacan D., Çarçani K., Ardiansyah P.O.D., Vasilache S. (eds) ACHI 2020: The Thirteenth International Conference on Advances in Computer-Human Interactions, 185-189. International Academy, Research, and Industry Association IARIA

https://www.thinkmind.org/articles/achi_2020_5_270_20136.pdf