https://doi.org/10.2352/ISSN.2470-1173.2020.11.HVEI-069

Quality Assessment Protocols for Omnidirectional Video Quality Evaluation

Ashutosh Singla, Stephan Fremerey, Frank Hofmeyer, Werner Robitza and Alexander Raake
Audiovisual Technology Group (AVT), Technische Universitt Ilmenau, Ilmenau, Germany
Email: [ashutosh.singla, stephan.fremerey, frank.hofmeyer, werner.robitza, alexander.raake]@tu-ilmenau.de

Abstract

In recent years, with the introduction of powerful HMDs such as Oculus Rift, HTC Vive Pro, the QoE that can be achieved with VR/360° videos has increased substantially. Unfortunately, no standardized guidelines, methodologies and protocols exist for conducting and evaluating the quality of 360° videos in tests with human test subjects. In this paper, we present a set of test protocols for the evaluation of quality of 360° videos using HMDs. To this aim, we review the state-of-the-art with respect to the assessment of 360° videos summarizes their results. Also, we summarize the methodological approaches and results taken for different subjective experiments at our lab under different contextual conditions. In the first two experiments 1a and 1b, the performance of two different subjective test methods, Double-Stimulus Impairment Scale (DSIS) and Modified Absolute Category Rating (M-ACR) was compared under different contextual conditions. In experiment 2, the performance of three different subjective test methods, DSIS, M-ACR and Absolute Category Rating (ACR) was compared this time without varying the contextual conditions. Building on the reliability and general applicability of the procedure across different tests, a methodological framework for 360° video quality assessment is presented in this paper. Besides video or media quality judgments, the procedure comprises the assessment of presence and simulator sickness, for which different methods were compared. Further, the accompanying head-rotation data can be used to analyze both content- and quality-related behavioural viewing aspects. Based on the results, the implications of different contextual settings are discussed.

Introduction

360° or omnidirectional videos offer an immersive and interactive experience to the viewers in a 360° viewing sphere while watching them with a Head-mounted Display (HMD). The resulting user experience plays an important role in making technology successful and acceptable for a wide user group. The respective "Quality of Experience" (QoE) is defined as the "...degree of delight or annoyance of the user of an application or service..." [1, 2]. One component of QoE is the perceptual media quality of the viewing session, resulting from different technical characteristics of the end-to-end chain comprising the recording, post-processing, representation, coding, retrieval, transmission, playback and display of 360° videos. Furthermore, especially for 360° video, exploration behavior, presence and simulator sickness also represent key aspects of QoE. Therefore, besides media quality, these three further constructs need to be assessed for a holistic understanding of QoE in case of omnidirectional videos. It should be noted that such findings are not only restricted to omnidirectional videos, but also hold for other interactive video applications, for example, augmented reality, VR telepresence systems, interactive omnidirectional guides, etc.

Since different constructs need to be assessed, the holistic and valid evaluation of QoE for 360° videos is a time-consuming task and requires a well-designed protocol. There are International Telecommunication Union (ITU) recommendations (ITU-T Rec. P.910 and ITU-R Rec. BT.500-13) that provide guidelines for the assessment of video quality of 2D videos on 2D displays with respective test subjects. However, no such standard document exists for omnidirectional videos. Therefore, it is necessary to develop a set of guidelines to study visual quality assessment for omnidirectional videos. To this aim, the paper investigates the following research questions: (1) What are the key aspects that need to be assessed besides media quality? (2) Which evaluation methodology and set of test protocols need to be applied for the assessment of perceived video quality of 360 videos? (3) What set of rating tasks/questionnaire types should be used to assess Simulator Sickness and Presence in the same test run together with quality?

The remainder of the paper is structured as follows: First, we discuss the different test methods and report the impact of different influencing factors on 360° video quality. In the following, aspects associated with simulator sickness such as its measurement, theories about its formation, etc. are presented. In the subsequent section, the different methods of analyzing users' viewing behaviour are reported. Conclusions are drawn at the end of the paper.

Video Quality Evaluation

Measuring and validating QoE of omnidirectional videos using different Head-mounted Displays (HMDs) has become increasingly important. Relevant system influencing factors for the QoE of omnidirectional videos are, for example, compression algorithms, projection schemes, display devices along with their rendering technologies, resolution, bit-rates, etc. Hence, assessing the impact of these factors on the QoE of omnidirectional videos with appropriate methods is essential. In this paper, mainly subjective video quality will be addressed. Several studies have been reported in the state-of-the-art dedicated to the assessment of subjective video quality [3–15]. Table 1 summarizes several studies conducted for the assessment of visual quality of 360° videos.

In [3], Tran et al. investigated the impact of different resolutions and bit-rates on the visual quality of 360° videos with Samsung Gear VR using the Absolute Category Rating (ACR) method, using the classical 5-point scale. Experimental results showed that 4K and QHD resolutions can provide better per-

Table 1: Summary of several test methods for the assessment of video quality for 360° Videos.

Name	Method	Resolution	Frame rate	Encoder	HMD Name	No. of Subjects	Duration of Video
Tran et al. [3]	ACR	4K, QHD, FHD, HD	30	H.264/AVC	Gear VR (S6) Google Cardboard (S5)	36	30 s
Duan et al. [4]	ACR	4K, 2K, 1K	30, 15, 5	–	HTC Vive	13	15 s
Singla et al. [5]	ACR	4K, FHD	25–30	–	HTC Vive & Oculus Rift	28	60 s
Lin et al. [6]	SSCQS	4K–8K	24–30	HEVC	HTC Vive	221	10–23 s
Lopes et al. [7]	ACR & ACR-HR	R1, FHD, 4K, 8K	7.5, 10, 15 30, 60	HEVC	Oculus Rift	37	10 s
Mahmoudpour et al. [8]	ACR	4K	30	HEVC	Oculus Rift	20	12 s
Covaci et al. [9]	ACR	HD, FHD, 2.5K, 4K	25–60	H.264	Samsung Gear VR	48	60 s
Hofmeyer et al. [10]	ACR, PC	4K	25, 30, 50, 60, 90	H.265	HTC Vive Pro	12	20 s
Zhang et al. [11]	ACR-HR	3600×1800	30, 60	HEVC	HTC Vive	30	10 s
Zhang et al. [12]	SSCQS, SAMVIQ SAMPVIQ	4K	30	H.264, H.265 VP9	HTC Vive	23	10 s
Singla et al. [13]	M–ACR	4K, FHD	30	H.265	Oculus Rift	30	10 s
Singla et al. [14]	DSIS	4K, FHD	30	H.265	Oculus Rift	27	10 s
Singla et al. [15]	ACR, M–ACR DSIS	4K, 6K 8K	30	H.265	HTC Vive Pro	87	10 s

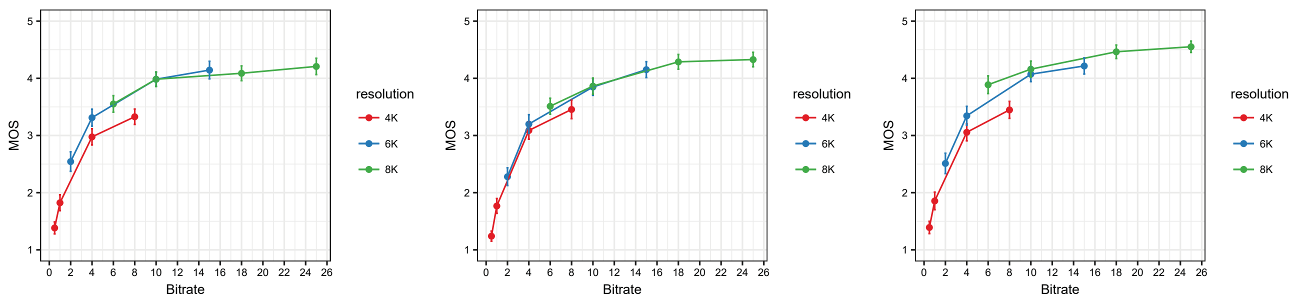
ceived quality than FHD and HD. Furthermore, they found that Samsung Gear VR (S6) provides a better perceived quality than Google Cardboard (S5) irrespective of resolution and video sequence. In [4], Duan et al. investigated the impact of framerate, resolution and bit rate using the ACR test method, establishing a dataset that contains 10 raw videos in 4096×2048 shot using an Insta360. Experimental results showed that higher resolution (4096×2048) and higher framerate (30 fps) provide better visual quality than lower resolutions (2048×1024 and 1024×512) and framerates (5 and 15 fps). Furthermore, the perceived visual quality at 70000 KBit/s is slightly higher than at 10000 KBit/s. In [5], Singla et al. assessed and compared the audiovisual quality in HTC Vive and Oculus Rift using the ACR test method. Experimental results indicated that 4K provides better-perceived quality than FHD. Additionally, the results imply that HTC Vive provides better integral quality than Oculus Rift, although both the HMDs have the same technical specifications. The difference in perceived quality could be due to the rendering algorithm or type of lenses used. The impact of content motion on the integral quality was also investigated. The content with the highest amount of motion was shown to provide the lowest integral quality and vice-versa.

In [6], Li et al. compared the three different projection methods Equirectangular projection (ERP), reshaped cubemap projection (RCMP) and truncated square pyramid projection (TSP), using a continuous quality scale with HTC Vive. Results suggest that TSP provides better subjective video quality. In [7], Lopes et al. evaluated the impact of different resolutions (R1 (960×480), FHD, 4K, 8K) and framerates (7.5, 10, 15, 30 and 60) on the visual quality of several omnidirectional videos using ACR and ACR-HR (ACR with Hidden-reference Removal) with an Oculus Rift. Experimental results indicate that 4K and 8K resolution provided almost the same perceived quality at 30 fps. Also, results showed the saturation in visual quality beyond 30 fps at 4K resolution. It is noted that the lack of differences between 4K and 8K resolution may be a result of either the source contents and their respective quality or of the limited resolution and resulting screen-door effect for the Oculus Rift used [16].

In [8], Mahmoudpour et al. studied the impact of judder and compression artefacts on the subjective visual quality of omnidirectional videos in Oculus Rift using a Single Stimulus Continuous Quality Scale (SSCQS) method. A dataset[1] was created that contains 12 omnidirectional videos shot with a GoPro Fusion 360° camera and three computer-generated VR videos. The test results show that at higher quality levels, the quality drop due to judder is more significant than that due to compression, at least for the considered parameter ranges. In [9], Covaci et al. investigated the impact of mulsemedia (multisensory media) on 360° videos. Experimental results showed that users perceived better quality for 360° mulsemedia than 360° multimedia irrespective of encoding quality. In [10], Hofmeyer et al. investigated several combinations of 360° players and framerates. They found that 90 fps provides significantly better quality than other framerates (25, 30, 50, 60). Further, results showed that *Whirligig* and *Virtual Desktop* both offer a smooth playback for 360° videos without flickering at 30 and 90 fps. They also investigated the impact of various motion-interpolation (MI) algorithms on QoE, when interpolating (temporally upscaling) sequences from 30 to 90 fps. The original framerate was 30 fps. It was concluded that MI, in general, is a suitable tool for increasing the QoE of 360° videos. While FFmpeg MCI should be preferred over butterflow for videos containing fast and sudden movements, for slow- and medium-motion videos, butterflow is more suitable.

In [11], Zhang et al. proposed a protocol for the subjective assessment of omnidirectional videos for coding applications. Considering the projection and resolution of the applied HMD, video sequences should be re-sampled to the optimal resolution 3600×1800 before coding for HTC Vive to alleviate the interference of HMD sampling. In [12], Zhang et al. compared the performance of three subjective test methods: SSCQS, subjective assessment of multimedia video quality (SAMVIQ) and subjective assessment of multimedia panoramic video quality (SAMPVIQ). Experimental results show that SAMPVIQ has better Pearson's linear correlation coefficient (PLCC) with the Video quality met-

[1] `http://data.etrovub.be/qualitydb/judder-vqa`



(a) ACR (b) M-ACR (c) DSIS

Figure 1. *MOS with corresponding CIs for different test methods [15].*

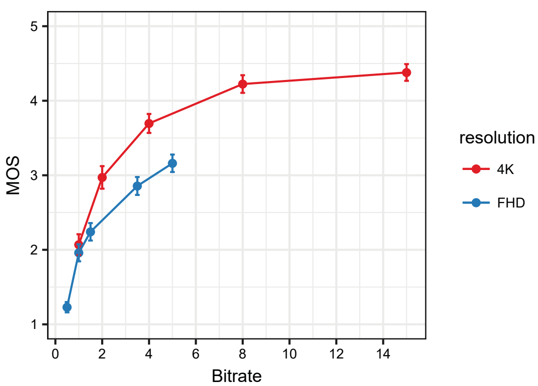


Figure 2. *MOS with corresponding CIs for M-ACR test method [13].*

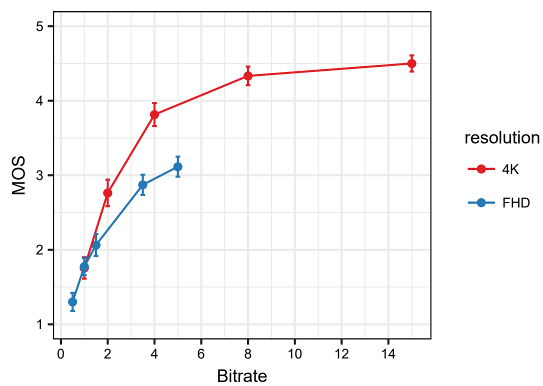


Figure 3. *MOS with corresponding CIs for DSIS test method [14].*

ric (VQM) [17] used for comparison. However, the authors have not validated which test method is statistically more reliable. Also, it is not clear if these three test methods are compared under the same contextual conditions.

Most of the studies reported above have used the ACR test method to evaluate the quality of 360° videos. The advantages of using the ACR method are that the Processed Video Sequence (PVS) is displayed only once targeting more absolute rather than relative rating. Further, with less repetitions, exploration will be more natural in spite of being in a video quality test.

Impact of Contextual Conditions

A Possible impact due to contextual settings on 360° video QoE assessment were revealed from two different sets of subjective experiments conducted at our lab. In test 1 [13, 14], we compared the performance of the DSIS and M–ACR[2] test methods under different contextual conditions using Oculus Rift. This way, even for 360° videos, the typical usage of short, 8 to 10 s long source videos can be accommodated, leaving enough time for adaptation to the omnidirectional scene and its quality judgment. The M–ACR test was conducted in winter and the DSIS test in summer. Figures 2 and 3 show the MOS averaged over all video sequences for different bit-rates for FHD and 4K resolution. It is clear from the results that 4K resolution provides better perceived quality than FHD resolution, irrespective of the test method, except for 1 Mbit/s when using DSIS.

To find out which of the test method is statistically more reliable, statistical reliability was calculated based on [18]. For M–ACR and DSIS, we computed the MCI *Mean Confidence Interval*, *MOS Range* and calculated MCI_{norm} using the method shown in [18]. Table 2 implies that M-ACR is slightly better than DSIS.

In experiment 2, we compared the performance of the DSIS, M–ACR and ACR test methods using the HTC Vive Pro. Here, all tests were run in the same season and hence under the same climatic (external) conditions. Figure 1 shows the MOS averaged over all video sequences for different bit-rates for 4K, 6K and 8K resolutions. It is clear from the results that 6K resolution provides better-perceived quality than 4K irrespective of the test method. The difference between the perceived quality provided by 6K and 8K is not clearly visible with any of the considered test methods. Statistical reliability was calculated based on [18] for all three test methods. In contrast to the results from experiment 1, Table 3 shows that DSIS provides higher reliability and resolving power between test conditions than M–ACR and ACR.

Table 2: MCI, MOS Range and MCI_{norm} for DSIS and M-ACR [14]

	DSIS	M-ACR
MCI	0.136	0.118
MOS Range	3.198	3.149
MCI_{norm}	0.042	0.037

Simulator Sickness

While viewing 360° videos on an HMD, users may experience symptoms of simulator sickness (or cybersickness) such

[2]M–ACR is an adapted version of the classical single-stimulus ACR-type approach, repeating the playout of the test stimulus twice before the actual quality rating.

Table 3: MCI, MOS Range and MCI_{norm} for ACR, M-ACR and DSIS [15].

Test Method	ACR	M-ACR	DSIS
MCI	0.138	0.1392	0.1405
MOS Range	2.824	3.088	3.16
MCI_{norm}	0.0489	0.0450	0.0444

as nausea, dizziness, vertigo, sweating, etc. [19]. Simulator sickness may occur during and after exposure to a virtual environment (VE). Hence, for a holistic picture of omnidirectional video QoE, simulator sickness should be monitored during respective tests.

For the self-evaluation of simulator sickness, the Simulator Sickness Questionnaire (SSQ) developed by Kennedy et al. in 1993 [19] may be used. As an alternative, physiological measures may be used 20, which are out of scope for the present paper. This questionnaire consists of 16 scales for symptoms such as headache, nausea, fatigue, sweating, etc. The users rate the presence of each symptom using a four-point scale (0 = not at all, 1 = slightly, 2 = moderately, and 3 = severely). Some modifications to SSQ were proposed by Kim et al. [21] as VRSQ (Virtual Reality Sickness Questionnaire). They analyzed the SSQ profiles for VR devices and observed that Nausea scores were the lowest among the three categories (nausea, oculomotor, disorientation). Hence, they eliminated the nausea component from SSQ, reducing the initially 16 scales to 9. Another questionnaire, MASQ (Motion Sickness Assessment Questionnaire) is proposed by Gianaros et al. [22], where 16 symptoms/descriptors are categorized into 4 categories (Gastrointestinal, Central, Peripheral, and Sopite-related), with each descriptor rated on a 9-point scale.

There are rapid self-report methods for the assessment of the well-being of users' applying a single scale. There are many single-scale tests reported in the literature such as Fast Motion Sickness Scale (FMS) by Keshavarz et al. [23], Misery Scale Index (MISC) by Wertheim et al. [24] and Short SSQ by Tran et al. [3]. Additionally, the motion sickness susceptibility questionnaire (MSSQ) [25] could be used in the experiment to know about the susceptibility of users to motion sickness.

Given the number of Long and Short Questionnaires, it can be difficult for the researcher to know which will be the most suitable for assessing simulator sickness in viewers of 360° video. Singla et al [26] compared a short SSQ [3] approach with the widely used full SSQ [19]. Experimental results indicated that the "Short SSQ" cannot replace the SSQ to evaluate the impact of factors such as resolution, bitrate, framerate, etc. on simulator sickness. However, the Short SSQ can effectively replace the SSQ to distinguish video contents based on whether they cause high or low simulator sickness. The impact of test duration and breaks on simulator sickness was investigated by Singla et al. in [15]. Experimental results show that with an increase in time, simulator sickness also increases, but breaks in-between help to reduce the effect of simulator sickness.

Presence

Omnidirectional videos allow users to interactively explore the scene, which gives them a perception of being present in the virtual environment (VE). To measure presence in a VE, the Presence Questionnaire (PQ) [27] has been widely used. Similarly to the case of the SSQ, also short versions of the approach have been applied. For example, the Short Presence Questionnaire (Short PQ) as described in [3] can be used. The PQ (version 2) consists of 32 questions which are further classified into four different categories: control, sensory, distraction and realism factors. Subjects rate each factor on a 7-point scale, with different descriptive labels for some factors. For the Short PQ, subjects are asked to answer one question on a 5-point scale (1: absolutely no sense of presence and 5: a true sense of presence as in a real environment).

Singla et al. [26] compared the modified PQ (version 2)[3] and short PQ for the case of omnidirectional video QoE testing. Experimental results show that the Short PQ could replace PQ only to evaluate the extent of presence provided by a specific video sequence. However, the Short PQ cannot replace the PQ for a detailed analysis such as to evaluate the impact of factors such as resolution, bitrate, framerate, etc.

Exploration Behaviour

When consuming videos using traditional 2D displays, the user is restricted to watch only a certain viewport of the scene. However, in omnidirectional videos, viewers can look wherever they want. The users become the director of the movie and are free to watch any view. With this freedom, a lot of researchers are interested to find out at which specific parts of the video users are looking at any given moment in time. In other words, they are interested in investigating the exploration behavior of the users. There are publicly available tools that can be used to conduct subjective tests using HMDs and collect the corresponding head-rotation data such as AVTrack360 [28] and MIRO360 [29]. An alternative to the head rotation data is to track eye movements while exploring 360° videos [30]. For the present paper, we will focus on head tracking only.

For example, AVTrack [28] is a Python-based framework for capturing exploration behaviour of people while watching 360° videos. It uses the Whirligig player for the playout of 360° videos. Together with the tool, the authors published a dataset providing the head rotation data for 48 subjects watching 20 different omnidirectional videos with 30 s duration in a task-free scenario. Four different methods were proposed to evaluate the head rotation data: percentage of subjects discovered the respective parts of the video, percentage of time spent by the subjects watching specific areas of 360° video, heatmaps and angular yaw speed heatmaps. Heatmaps are found to be a good way for visualizing how people are exploring omnidirectional videos, while angular yaw speed heatmaps can be used to investigate the effect of specific video events at a specific time-stamp on the exploration behavior. MIRO360 [29] is an Android-based tool that can be used to carry out the subjective test with DSIS and ACR test methods using Samsung Gear VR. This tool also records the head rotation data of users while watching the 360° videos. Additionally, after the video is over, there are different questions that can be asked from users related to quality, simulator sickness and presence, while their response can be recorded using a controller.

Recommendations

Video Quality

Based on the studies reviewed above, recommendations/guidelines can be derived for conducting tests for media

[3] based on the test design and target effect, PQ was adapted and 22 questions were selected which fit the goals

quality assessment of 360° videos. In general, it is proposed to assess all major constituents of 360° video QoE, namely audiovisual quality, simulator sickness and presence, collecting head- and if possible eye-tracking data for viewing-behaviour analysis. For media quality assessment, where considered meaningful, known standardized subjective test methods as e.g. recommended by ITU-T or ITU-R will be referred to.

Before starting the test, the test participants should be screened for correct visual acuity using Snellen charts (20/20) and for colour vision using Ishihara charts, as described in ITU-R Rec. BT.500-13. During post-test subject screening, outlier detection can be performed based on the Pearson Linear Correlation Coefficient (PLCC) as described in [31] or based on ITU-R Rec. BT.500-13. Additionally, the Kurtosis coefficient could be computed for each subject to check if their rating score is normally distributed.

For playing out 360° videos, a player must be used that provides smooth playback for 360° videos without stuttering, such as Whirligig [10]. It was shown that for reducing the quality-impact due to the mismatch between video framrate and HMD refresh rate, and hence avoid stuttering, the video framerate should be adapted to the HMD's refresh rate [10]. For interpolating the sequences from the given framerate to the 90 Hz used by most HMDs, FFmpeg MCI and butterflow algorithms are suggested, depending on the amount of motion in the video [10]. To provide at least a reasonable level of perceived quality, video resolution should be at least 4K [3, 13, 14].

For the evaluation of short video sequences ($<$ 20 s), for a more fine-grained video-quality assessment, a double-stimulus test method should be used, such as DSIS. In case that a more absolute rating is sought, an adapted single-stimulus method such as the proposed M–ACR can be used, where the test sequences are repeated twice so that test participants have enough time to explore the omnidirectional videos and assess their quality.

Using longer sequences ($>$ 20 s) enables a more realistic and holistic viewing experience, since subjects can better explore and adjust to the current content. In this case, a double-stimulus method is not suitable. Instead, a single-stimulus test method such as ACR can well be used to avoid a longer duration of the test.

For the collection of ratings, either test participants could record their responses using a controller or report the rating number aloud to an experimenter present in the room, who notes down the rating of the video [13]. Both approaches enable subjects to continuously wear the HMD, avoiding possibly detrimental effects of taking on- and off of the HMD when collecting ratings outside e.g. on a separate screen. It was shown that particular care must be taken when different subjective tests are to be compared, where contextual effects such as outside-the-lab temperatures may have a stronger effect on the results [15].

Simulator Sickness and Presence

Based on the simulator sickness scores, we recommend not to exclude any test participant, reflecting individual effects. For the collection of ratings, for short questionnaires, test participants can record their response either verbally or using controllers. For the long questionnaires, the responses could be recorded on paper, taking breaks between viewings.

Based on the studies reviewed, we recommend using a single-scale question to assess simulator sickness and presence after each stimulus, especially, in case of tests that focus on media quality, and/or which apply short stimuli ($<$30 s). Due to its simplicity, it may complement the quality rating, and could be asked after every PVS. In turn, the full questionnaire may be used for longer stimuli ($>$30 s), and will be more indicative for the effects due to bitrate, framerate, resolution and test duration. Also in case of a media-quality test with short stimuli, the full questionnaire is recommended to be used, however only after each complete test block, then as an indicator to monitor the wellbeing of the test subject over the test duration.

Conclusions

In this paper, we have reviewed some of the relevant literature on the subjective evaluation of 360° video QoE. From the results and those obtained in a number of own tests, we derived a first set of recommendations that could be used for conducting subjective tests for a holistic evaluation of 360° video QoE. To this aim, the evaluation was decomposed into four indicators of omnidirectional video QoE, namely media quality, simulator sickness, presence and viewing behavior. In future work, the initial recommendations will be substantiated by further tests and methodological comparisons. Here, a particular focus will be on the usage of full sets of questionnaires versus the usage of shortened approaches, and in particular on the role of simulator sickness and wellbeing.

Acknowledgments

EFRE
EUROPE FOR THURINGIA
EUROPEAN REGIONAL DEVELOPMENT FUND
EUROPEAN UNION

References

[1] *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)(S. Möller, P. Le Callet and A. Perkis, eds.)*, Lausanne, 1.1 ed., Qualinet White Paper on Definitions of Quality of Experience – Output Version of the Dagstuhl Seminar 12181, 2012.

[2] ITU–T Rec. P.10/G.100, *Vocabulary for performance and quality of service*, International Telecommunication Union, Geneva, 2017.

[3] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung, and T. C. Thang, "A subjective study on qoe of 360 video for vr communication," *IEEE 19th International Workshop on MMSP*, pp. 1–6, 2017.

[4] H. Duan, G. Zhai, X. Yang, D. Li, and W. Zhu, "Ivqad 2017: An immersive video quality assessment database," in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, May 2017, pp. 1–5.

[5] A. Singla, S. Fremerey, W. Robitza, and A. Raake, "Measuring and comparing qoe and simulator sickness of omnidirectional videos in different head mounted displays," in *9th International Conference on QoMEX*, May 2017, pp. 1–6.

[6] C. Li, M. Xu, X. Du, and Z. Wang, "Bridge the gap between vqa and human behavior on omnidirectional video: A large-scale dataset and a deep learning model," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM 18. New York, NY, USA: Association

for Computing Machinery, 2018, p. 932940. [Online]. Available: https://doi.org/10.1145/3240508.3240581

[7] F. Lopes, J. Ascenso, A. Rodrigues, and M. P. Queluz, “Subjective and objective quality assessment of omnidirectional video,” in *SPIE Conference Series*, Sep. 2018.

[8] S. Mahmoudpour and P. Schelkens, “Visual quality analysis of judder effect on head mounted displays,” in *2019 27th European Signal Processing Conference (EUSIPCO)*, Sep. 2019, pp. 1–5.

[9] A. Covaci, R. Trestian, E. a. B. Saleme, I.-S. Comsa, G. Assres, C. A. S. Santos, and G. Ghinea, “360 mulsemedia: A way to improve subjective qoe in 360 videos,” in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM 19. New York, NY, USA: Association for Computing Machinery, 2019, p. 23782386. [Online]. Available: https://doi.org/10.1145/3343031.3350954

[10] F. Hofmeyer, S. Fremerey, T. Cohrs, and A. Raake, “Impacts of internal hmd playback processing on subjective quality perception,” *Electronic Imaging*, vol. 2019, no. 12, pp. 219–1, 2019.

[11] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang, and Z. Chen, “Subjective panoramic video quality assessment database for coding applications,” *IEEE Transactions on Broadcasting*, pp. 461–473, June 2018.

[12] Bo Zhang, Junzhe Zhao, Shu Yang, Yang Zhang, Jing Wang, and Zesong Fei, “Subjective and objective quality assessment of panoramic videos in virtual reality environments,” in *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, July 2017, pp. 163–168.

[13] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake, “Comparison of subjective quality evaluation for hevc encoded omnidirectional videos at different bit-rates for uhd and fhd resolution,” in *Proceedings of the on Thematic Workshops of ACM Multimedia*, 2017. [Online]. Available: http://doi.acm.org/10.1145/3126686.3126768

[14] A. Singla, W. Robitza, and A. Raake, “Comparison of subjective quality evaluation methods for omnidirectional videos with dsis and modified acr,” *Electronic Imaging*, vol. 2018, pp. 1–6, 2018.

[15] A. Singla, W. Robitza, and A. Raake, “Comparison of subjective quality test methods for omnidirectional video quality evaluation,” in *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*, Sep. 2019, pp. 1–6.

[16] B. J. Sitter, J. Yang, J. Thielen, N. K. Naismith, and J. Lonergan, “78-3: Screen door effect reduction with diffractive film for virtual reality and augmented reality displays,” 2017.

[17] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Transactions on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[18] T. Tominaga, T. Hayashi, J. Okamoto, and A. Takahashi, “Performance comparisons of subjective quality assessment methods for mobile video,” in *2010 Second International Workshop on QoMEX*, June 2010, pp. 82–87.

[19] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness,” *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[20] Y. Y. Kim, H. J. Kim, E. N. Kim, H. D. Ko, and H. T. Kim, “Characteristic changes in the physiological components of cybersickness,” *Psychophysiology*, vol. 42, no. 5, pp. 616–625, 2005.

[21] H. K. Kim, J. Park, Y. Choi, and M. Choe, “Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment,” *Applied Ergonomics*, vol. 69, pp. 66 – 73, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S000368701730282X

[22] P. J. Gianaros, E. R. Muth, J. T. Mordkoff, M. E. Levine, and R. M. Stern, “A questionnaire for the assessment of the multiple dimensions of motion sickness.” *Aviation, space, and environmental medicine*, vol. 72 2, pp. 115–9, 2001.

[23] B. Keshavarz and H. Hecht, “Validating an efficient method to quantify motion sickness,” *Human Factors*, vol. 53, no. 4, pp. 415–426, 2011, pMID: 21901938. [Online]. Available: https://doi.org/10.1177/0018720811403736

[24] J. E. Bos, S. C. de Vries, M. L. van Emmerik, and E. L. Groen, “The effect of internal and external fields of view on visually induced motion sickness,” *Applied Ergonomics*, vol. 41, no. 4, pp. 516 – 521, 2010, special Section - The First International Symposium on Visually Induced Motion Sickness, Fatigue, and Photosensitive Epileptic Seizures (VIMS2007). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0003687009001586

[25] J. F. Golding, “Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness,” *Brain Research Bulletin*, vol. 47, no. 5, pp. 507 – 516, 1998. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0361923098000914

[26] A. Singla, R. R. R. Rao, S. Gring, and A. Raake, “Assessing media qoe, simulator sickness and presence for omnidirectional videos with different test protocols,” in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2019, pp. 1163–1164.

[27] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[28] S. Fremerey, A. Singla, K. Meseberg, and A. Raake, “Avtrack360: An open dataset and software recording people’s head rotations watching 360-degree; videos on an hmd,” in *Proceedings of the 9th ACM MMSys*, 2018.

[29] P. Prez and J. Escobar, “Miro360: A tool for subjective assessment of 360 degree video for itu-t p.360-vr,” in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, June 2019, pp. 1–3.

[30] E. J. David, J. Gutiérrez, A. Coutrot, M. P. Da Silva, and P. L. Callet, “A dataset of head and eye movements for 360 videos,” in *Proceedings of the 9th ACM Multimedia Systems Conference*, ser. MMSys 18, 2018, p. 432437.

[31] T. Tan, M. Mrak, R. Weerakkody, N. Ramzan, V. Baroncini, G. Sullivan, J.-R. Ohm, and K. McCann, *HEVC subjective video quality test results*. United Kingdom: Institution of Engineering and Technology, 2014, location: Amsterdam, Netherlands Conference date: 11-15 Sept. 2014.

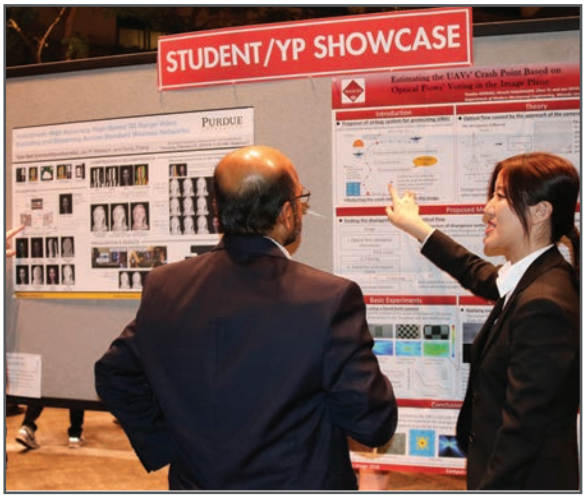
JOIN US AT THE NEXT EI!

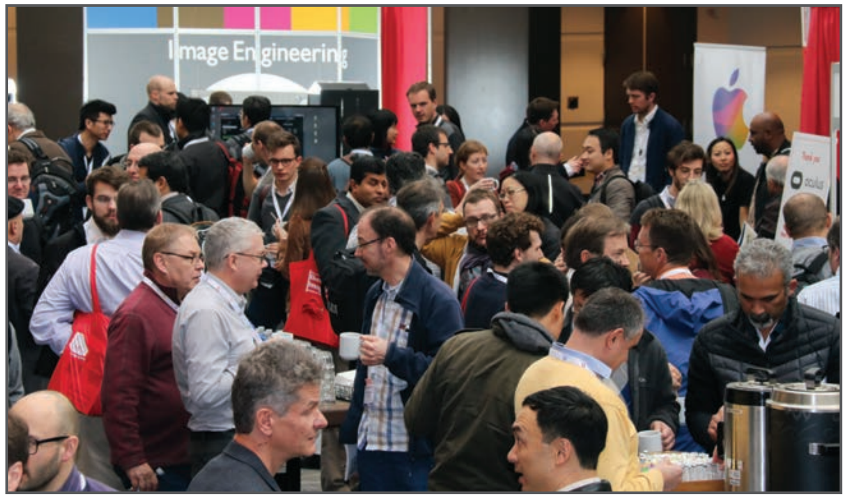
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org