# Towards Subjective Quality of Experience Assessment for Omnidirectional Video Streaming

Raimund Schatz*, Andreas Sackl*,Christian Timmerer†, Bruno Gardlo*
*AIT Austrian Institute of Technology, †Alpe-Adria Universität Klagenfurt / Bitmovin Inc.
{name.surname}@ait.ac.at, christian.timmerer@itec.aau.at

***Abstract*—Currently, we witness dramatically increasing interest in immersive media technologies like Virtual Reality (VR), particularly in omnidirectional video (OV) streaming. Omnidirectional (also called 360-degree) videos are panoramic spherical videos in which the user can look around during playback and which therefore can be understood as hybrids between traditional movie streaming and interactive VR worlds. Unfortunately, streaming this kind of content is extremely bandwidth intensive (compared to traditional 2D video) and therefore, Quality of Experience (QoE) tends to deteriorate significantly in absence of continuous optimal bandwidth conditions. In this paper, we present a first approach towards subjective QoE assessment for omnidirectional video (OV) streaming. We present the results of a lab study on the QoE impact of stalling in the context of OV streaming using head-mounted displays (HMDs). Our findings show that subjective testing for immersive media like OV is not trivial, with even simple cases like stalling leading to unexpected results. After a discussion of characteristic pitfalls and lessons learned, we provide a a set of recommendations for upcoming OV assessment studies.**

***Keywords*—Omnidirectional Video, 360-degree Streaming, Immersive Media, Subjective Testing, Quality of Experience***

## I. Introduction

Universal access to and provisioning of (multi-)media content have become reality, nowadays it is very easy – in real-time – to generate, distribute, share, and consume any media content, anywhere, anytime, and with any device. These kind of real-time entertainment services – specifically, streaming audio and video – are typically deployed over the open, unmanaged Internet and account now for more than 70% of the evening traffic in North American fixed access networks. It is assumed that this number will reach 80% by the end of 2020 [1]. Although Internet capacity is constantly increasing for both fixed and mobile networks, the adoption of streaming audio and video services will continue as well as new applications and services will emerge including – but not limited to – ultra high-definition (UHD) 4K/8K, high dynamic range (HDR), and virtual reality (VR), specifically omnidirectional video (OV) also known as 360-degree video. Consequently, the data volume increases which calls for adaptive streaming techniques to cope with bandwidth fluctuations or – in general – with dynamically changing context conditions.

MPEG Dynamic Adaptive Streaming over HTTP (DASH) [2] is a widely adopted and deployed standard for adaptive media streaming and its usage for omnidirectional video is at the time of writing of this paper subject to standardization efforts as part of the omnidirectional media application format (OMAF). In practice, however, devices for the creation (i.e., 360-degree camera systems including stitching software) and the consumption (i.e., cardboard boxes, head-mounted displays (HMDs)) are becoming available at reasonable prices. The efficient streaming of such content is a big issue since most devices utilize the equirectangular format resulting in a significant increase of the data volume. For example, in order to get a 4K resolution for a field of view of 120-degree, a horizontal resolution of 12K is needed. Additionally, the Quality of Experience (QoE) for omnidirectional/360-degree video streaming scenarios is not yet investigated but expectations are currently very high.

In this paper, we describe a first approach towards a subjective QoE assessment for omnidirectional video streaming. The main focus is on stalling events which are known to have significant impact on the QoE [3]. We describe the methodology used for conducting a subjective QoE assessment of omnidirectional video (OV) streaming and discuss the results. Additionally, we provide an analysis of the lessons learned and propose a roadmap about what is needed from both academia and industry. The remainder of this paper is structured as follows. Section II describes related work and provides a brief overview of standardization activities in this area. The QoE lab experiment is described in Section III and its results in Section IV. After a discussion of these results in Section V, Section VI concludes the paper with a summary of lessons learned as well as potential future work items.

## II. Standardization and Related Work

### A. Related Developments in Standardization

JPEG started an initative called *Pleno* [4] focusing on images but our focus is on video. In this context, MPEG started a new work item related to *immersive media* referred to as MPEG-I (ISO/IEC 23090) which currently foresees five parts. The first part will be a technical report describing the scope of this new standard and a set of use cases and applications from which actual requirements can be derived. The second part specifies the omnidirectional media application format (OMAF) [5] addressing the urgent need of the industry for a standard in this area. Part three will address immersive video and part four defines immersive audio. Finally, part four will contain a specification for point cloud compression for which a call for proposals is available. OMAF is part of a first phase of standards related to immersive media and should finally become available by the end of 2017, beginning of 2018 while the other parts are scheduled at a later stage around 2020. The current OMAF committee draft comprises a specification of the *i)* equirectangular projection format (others

QoMEX2017 – Erfurt, Germany; 978-1-5386-4024-1/17/$31.00 

might be added in the future), *ii)* metadata for interoperable rendering of 360-degree monoscopic and stereoscopic audio-visual data, *iii)* storage format adopting the ISO base media file format (ISOBMFF/mp4), and *iv)* the following codecs: High Efficiency Video Coding (HEVC) and MPEG 3D audio.

### B. Objective and Subjective Evaluation Methods

Existing objective metrics such as PSNR (or SSIM) are known for their limitations as QoE metrics and are even more controversial for omnidirectional video. However, spherical PSNR [6] and viewport PSNR [7] have been proposed which can be used with Bjøntegaard Delta [8] known from traditional video applications. In [7] subjects are used but only to capture head movements for offline objective evaluation without conducting any formal subjective study. The issue of interactive omnidirectional video delivery with respect to bandwidth efficiency has been proposed in [9] and evaluated using viewport PSNR but lacks of a subjective evaluation.

Related work on stalling events (i.e., temporary interruption of media playback) has extensively evaluated in the literature. The QoE impact when consuming streaming video applications under different conditions has been surveyed for traditional HTTP streaming [3] and more specifically for stalling events in [10] and [11]. In subjective quality assessments, participants were exposed to clips impaired with predefined stalling patterns, often accompanied with a visual stalling indicator (i.e., spinner). In general, the results of these studies show that stalling severely impacts the QoE, much stronger than other typical impairments like initial delay or quality switching. However, the majority of these works addresses stalling in the context of traditional 2D video streaming and not in immersive 3D settings with HMDs.

The QoE impact of freezing events in the context of 3D video streaming has been investigated by Kara et al. [12] where they have extensively analyzed the results of a subjective lab study (N=20) based on watching video content on a auto-stereoscopic display (3D TV). With freeze duration being the main independent variable (30-500ms for each of the three freeze events per clip), their study focuses on the impact of short freeze events, with the main phenomena studied being detectability of freezes, changing subject tolerance to freeze impairments over time, and the influence of scene motion. Their results show that freeze perception varies widely across subjects and strongly depends on scene content (e.g., Y vs. Z-motion). Based on their findings they also identify the challenge that the test design itself might have a strong influence on the results. In contrast to this work, the study presented in this paper focuses on the impact of stalling events in a fully immersive setting, i.e., the user watching an omnidirectional video using a HMD. Furthermore, we focus on the QoE impact of longer stalling durations (1-6s) as well as different numbers of stalling events (0-3 events/clip). In addition to state of the art QoE questions, our test protocol also features additional rating questions regarding immersion and presence, beyond just annoyance by stalling.

In the context of subjective video quality testing using HMDs, the work of Patrick Seeling [13] presents a lab study in which participants assessed the quality of 2D movie segments using a binocular see-through heads-up display. Using the ITU-R BT.500-13 recommendation [14] as general guideline, the author exposed participants to 12 movie segments of varying duration (35-145s) encoded at five different quality levels. The results show that participants tended to overestimate video quality, with higher levels of content dynamics leading to more positive ratings (compared to PSNR and SSIM ground truths), which is attributed to the nature of the heads-up display. The author also highlights the importance of standardized test protocols that allow for comparable quality evaluations when using these new kinds of devices.

## III. QoE Lab Experiment

In order to investigate the QoE assessment for omnidirectional video streaming, we set up and conducted a lab study at AIT in Vienna, Austria. In terms of quality impairments we chose to focus on the impact of stalling, since this type of impairment is *a)* highly relevant and common in the context of OV streaming and *b)* straightforward to define and implement. Please note that in this experiment we implemented stalling as temporal stopping of audio and video playback in the form of *freezing* without a visual indicator (i.e., spinner) since this implementation is highly common among current VR movie players and helps avoiding unwanted influences from a specific spinner design rendered in stereoscopic mode. In order to provide a reference we chose to expose participants to stallings not only in the context of HMD-based OV streaming, but also in a traditional TV-based 2D viewing setting. Our experiment addressed the following three research questions:

- How does stalling impact QoE of HMD-based omnidirectional video? (RQ1)
- What are the differences of QoE between HMD-based and traditional TV-based video consumption with respect to stalling? (RQ2)
- What is the impact of content motion levels on OV stalling perception? (RQ3)

### A. User Study Setup and Test Design

For test setup and procedure we used ITU-R BT.500-13 [14] and ITU-T P.913 as general guidelines. To examine the impact of stallings on quality perception our test subjects had to watch short video clips with a duration of 60 seconds each, impaired by stalling patterns (implemented as freezing) on two different viewing device types: *a)* a large TV screen (65 inch diameter) in traditional 2D; and *b)* an Oculus Rift DK2 VR headset with headphones and 3-axis orientation sensor to experience OV streaming in stereoscopic mode.

The source (SRC) video sequence for the TV conditions is a one minute excerpt from the computer animated short film *Sintel*[1] by the Blender Institute. For the OV streaming experience we used two scenes from the ZDF documentary "Vulkane" (Volcanoes)[2], one scene with hardly any camera movement ('VR static') and another scene featuring strong camera movement ('VR move') in order to investigate the impact of content motion on stalling perception. As hypothetical reference circuits (HRCs) we used five stalling/freezing patterns: *i)* no stalling, *ii)* one stalling of six seconds in the middle of the video, *iii)* two and *iv)* three, six seconds long

[1] https://durian.blender.org/

[2] https://www.zdf.de/dokumentation/terra-x/3d-360-grad-immersiver-film-vulkane-100.html

| HRC | Stalling start positions within video | Total stall duration |
|---|---|---|
| no stalling | - | 0 seconds |
| 1x6 seconds | 00:30 | 6 seconds |
| 2x6 seconds | 00:20, 00:40 | 12 seconds |
| 3x6 seconds | 00:15, 00:30, 00:45 | 18 seconds |
| 3x1 seconds | 00:15, 00:30, 00:45 | 3 seconds |

TABLE I. OVERVIEW OF HRCS FEATURING DIFFERENT STALLING PATTERNS.

stallings equally distributed during the video duration, and *v)* three one second long stallings also equally distributed. Table I provides an overview of the HRCs used for the different processed video sequences (PVS).

In total, 27 users participated in our lab user study as our goal was to ultimately obtain valid data from at least 20 users. They were recruited via our test user database, received a monetary incentive in the form of vouchers and were also screened for 2D and 3D vision problems. Note that due to the screening as well as arising problems during the test session (lack of HMD fit, cybersickness, etc.) we had to remove five subjects. Thus, for the remainder of this paper, we use the results from only 22 participants (N=22). Most of our participants had no prior experience with VR or HMDs. Only four participants had experienced VR-related technologies at exhibitions or used VR setups from friends before. For these reasons, we exposed participants to a set of training conditions at the beginning of the test session in order to mitigate the novelty effect as well as to ensure good fit of the device and its optics. Because of time constraints, not all test participants evaluated the VR conditions with the second OV clip. Thus, only 18 test participants evaluated the 'VR move' conditions. To reduce contextual effects, the order of presented test conditions was randomized.

After each test condition, study participants had to answer several rating questions, as part of a single-stimulus with hidden reference procedure. For both device scenarios (2D TV and OV HMD), the users assessed the overall quality ("How do you perceive the overall quality of the video?", with answering options ranging from 5=excellent to 1=bad) and stalling annoyance ("How annoying were interruptions of the video playback?", with answering options ranging from 5=imperceptible to 1=very annoying). Furthermore, in both device settings we asked a question (INV4) adopted from the Igroup Presence Questionnaire (IPQ) [15] to gauge the extent the user's attention was absorbed by the viewing experience. Furthermore, we asked three additional IPQ questions (*G1*, *INV1*, *REAL2*) after each OV condition (see Table II). The aim behind adding the four IPQ questions was to assess the impact of stalling on further experience-related dimensions (immersion, presence) supposed to be relevant for omnidirectional video consumption.

An important issue when quality testing OV is the fact that the viewer can freely change his/her viewing direction. Compared to traditional video quality assessment, this additional degree of user freedom creates the challenge of the experimenter not being in full control which part of the world represented by the video the user actually sees at a given point in time, endangering reproducibility of results. We had to decide between two general solutions to this problem: *a)* the user has no control over the viewing direction (e.g., fixed viewpoint) or *b)* let the user look around, but use some means (e.g., instructions, content design) to influence head movement as wanted. We decided for the latter approach since a fixed view feels completely unnatural when using HMDs. To reduce head motion we told participants that "it is fine to look around for orienting yourself, but you can safely look straight ahead as this is where most of the action of the movie actually is". This solution worked fairly well for our study, also since it was in line with the way the content was designed.

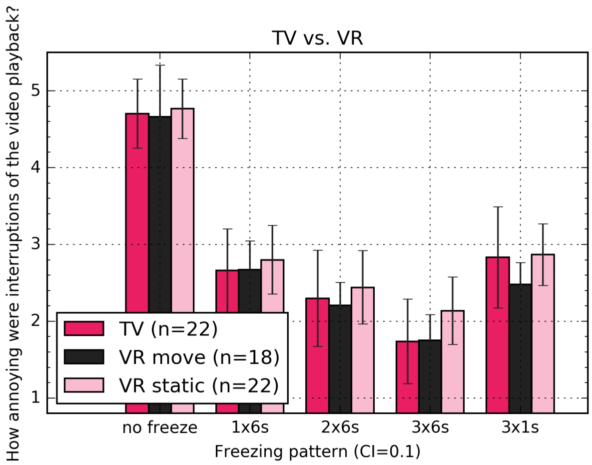


Fig. 1. Impact of the Different Stalling Patterns on Subjective Assessment Ratings (5=imperceptible, 1=very annoying).

## IV. STUDY RESULTS

Figure 1 addresses research question RQ1 and depicts the impact of the different stalling patterns on the subjective annoyance assessment. For both device contexts (TV and OV), already a single stalling event leads to a significant drop regarding annoyance, as known from other studies [10], [11]. Surprisingly, there are no significant differences between the ratings for TV and for OV. In both situations, the occurrence of stalling events lead to similar annoyance assessment ratings (RQ2). Initially, we assumed that in the case of HMD-based OV consumption, stalling would had a stronger impact than in the case of watching on the TV which is supposed to be the less immersive setting. Also the impact of OV content turned out to be less relevant than expected as MOS results do not differ significantly between the two OV SRCs featuring very different levels of camera motion (RQ3). Figure 3 shows the amount of motion in terms of average motion vector sizes over time (extracted on a frame-level basis and aggregated on second-level) for the two panoramic movie clips. The graph confirms that the 'VR move' clip with high camera movement (predominantly vertical panning) exhibits significantly more motion than the 'VR static' clip (mean average motion vector size 8.5 vs. 2.2). Therefore, we initially expected that stalling would have significantly higher QoE impact for the 'VR move' clip.

Figure 2 (top) shows the impact of the various stalling patterns on the subjective involvement assessment ratings (IPQ INV4 question) (RQ1). In contrast to the explicit stalling

| IPQ item | Shortcut | English question | English anchors (7 items scales) |
|---|---|---|---|
| INV4 | Attention captivated by virtual environment | I was completely captivated by the virtual world. | fully disagree – fully agree |
| G1 | Sense of being there | In the computer generated world I had a sense of ”being there” | not at all – very much |
| INV1 | Awareness of real environment | How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)? | extremely aware – not aware at all |
| REAL2 | Experience similar to real environment | How much did your experience in the virtual environment seem consistent with your real world experience? | not consistent – very consistent |

TABLE II. QUESTIONS ADOPTED FROM THE IGROUP PRESENCE QUESTIONNAIRE (IPQ) [15].

annoyance ratings, the impact of stalling events on this immersion-related metric is much weaker than in the case of the stalling annoyance ratings. Although the INV4 score differences are not significant, they show certain trends: as expected, the TV condition captivated participants' attention consistently to a lower extent than the OV conditions and user involvement in the static VR scene is less susceptible to disruption than for the higher motion VR scene. Figure 2 (bottom) compares the MOS results of all four IPQ questions (see Table II) for the 'VR move' conditions. Compared to the INV4 IPQ scores, the impact of stalling amount on the other three IPQ scores is even less pronounced. Together with the fact that score variances and the resulting error bar sizes are fairly large (see also Figure 4) and that the related scores are highly correlated (PLCCs ranging from 0.79 to 0.92), these results suggest that either immersion and presence are not affected by stalling at all (which we consider unlikely) or that participants had difficulties relating these four questions to the experienced test conditions in the intended way.

Additionally, we also analyzed the impact of user-related factors in order to identify possible causes for the above results. Therefore, we split users into two age groups (younger than 35 years, and older or equal than 35 years), but differences between the resulting MOS splits were marginal. Furthermore, we collected participants' feedback and comments during and after each test session. Most interestingly, half of the users provide negative comments about the HMD resolution. This is reasonable, since we used the Development Kit 2 (DK2) of the Oculus VR headset which in contrast to later versions has a screen resolution of only 960x1080px for each eye. We also had several comments about the lack of comfort when wearing the HMD, emerging feeling of exhaustion as well as missing options to mechanically adapt the headset exactly to the individual shape of the head. However, splitting participants according to their comments also does not explain any scoring variance, i.e., we found no impact of comments made vs. not made on ratings. Notably, in addition to complaints there were also comments from our test users expressing their positive surprise with the VR experience. Thus, although user comments did not explain any significant variance in our case, they highlight relevant issues that might have generally increased the variance of scores and that should be definitely taken into account for future HMD-based QoE assessment setups.

## V. DISCUSSION

In general, the results of the study presented in this paper suggest that stalling events seriously impact the QoE of OV streaming in very similar ways as they do in the context of traditional modes of watching (TV, mobile, etc.). From this perspective it seems that our experimental setup is capable of reliably quantifying QoE (as influenced by technical impairments like stalling) for such a more immersive and interactive type of medium. On the other hand, the number of unexpected results found in the study data and our observations made during the test sessions (see previous section) suggest that QoE testing for OV streaming is more challenging than for traditional setups and that testing methodologies currently

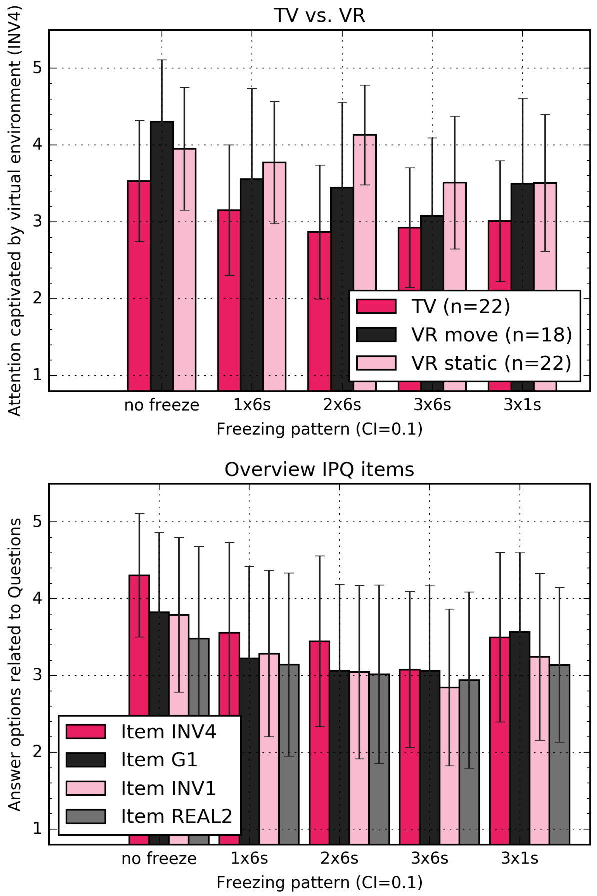


Fig. 2. Top: Impact of Stalling Pattern on Perceived Involvement ("I was completely captivated by the virtual world", 5=fully agree to 1=fully disagree). Bottom: Scores of the four IPQ questions as influenced by stalling for the 'VR move' condition (5=strong *sense of presence/immersion*, 1=weak *sense of presence/immersion*).

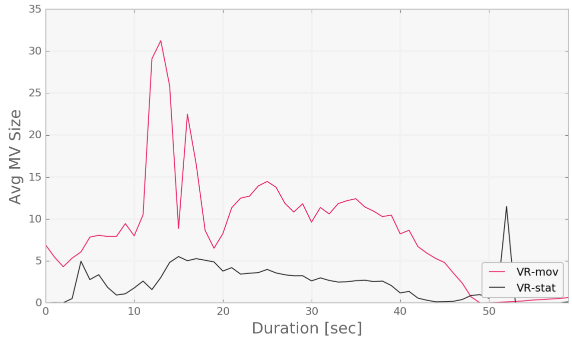


Fig. 3. Average motion vector sizes over time for the two different panoramic video clips (static vs. moving camera) reflecting motion intensity.

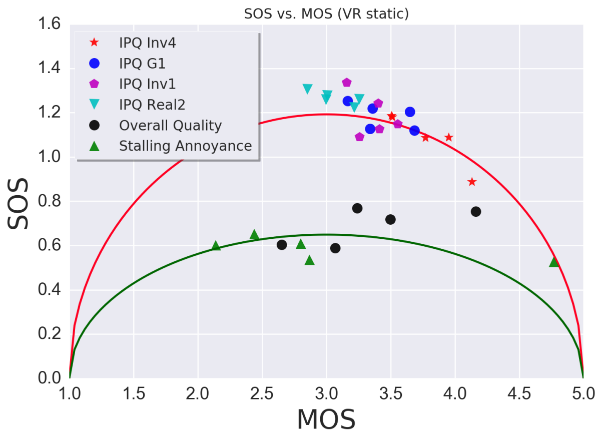


Fig. 4. SOS vs. MOS[16] plot for all rating items used in the 'VR static' conditions. High SOS indicates high scoring variance for a specific question in a specific test condition. The curves are fittings to IPQ-INV4 (red) and stalling annoyance (green) MOS-SOS pairs.

used need to be substantially improved in order to ensure reproducible research on this subject.

As regards the surprisingly low score differences between the three test scenarios (TV, OV static camera, OV moving camera), we see various plausible explanations: besides the possibility that stallings might actually be perceived by users with the exactly same severity in the three different scenarios, the relatively high rating variance even in the case of the supposedly easy-to-answer stalling annoyance question (see Figure 1) indicates that for the case of testing stallings (or freezing) in TV and OV, one should not expect the same amount of consensus among observers as, e.g., for traditional 2D image quality evaluation. Although one could argue that in related studies like [17] and [12] observer agreement is higher than in our experiment, but this can be clearly attributed to the nature of stimuli (panoramic images at different encodings in [17]) and experimental designs (freeze event number and positions did not vary in [12]) used. In addition, the fact that similar issues (i.e., high variance of ratings, large error bars) were experienced in the context of video quality testing on binocular see-through glasses [13] with a similar number of users (N=20) confirms our explanation. Therefore, we recommend that the number of subjects for QoE evaluations of immersive video applications should be substantially higher (e.g., N=40–50) in order to enable greater statistical significance when comparing across, e.g., different content types and viewing devices.

From an assessment methodology point of view, a number of issues needs to be addressed and ultimately, standardized in the case of OV. A major point is the fact that subjects can freely move their head and, thus, change view orientation during video playback. This essential characteristic of OV needs to be maintained for a realistic experience of the content. However, at the same time this also affects rating reliability, specifically in cases where content characteristics and impairments – e.g. encoding artifacts – are not homogeneously distributed over the whole viewing sphere. This might be less of a problem when testing the impact of global impairments like stallings, but more critical when evaluating dynamic optimization techniques like progressively increasing resolution of selected picture tiles according to current gaze direction and viewport as foreseen by MPEG-OMAF. Therefore, head movement and viewing direction should be addressed with deliberate measures, either by providing participants with specific instructions, by content design (i.e., provide obvious points of interest that attract attention) or by tracking head movements to create saliency maps (like in [17]) which can then be used to cluster test results according to viewing direction change patterns.

Another critical methodological aspect that emerged from our results relates to the number and type of rating questions used for QoE assessment. As already highlighted in Section IV, the sensitivity of the four IPQ items' scores to the degree of stalling (see Figure 2) was surprisingly low compared to the results of more traditional questions directly addressing stalling annoyance. We would have expected at least a significant difference between test subjects' sense of immersion and presence in stalling vs. non-stalling conditions which was not the case. Results of a more detailed analysis of the IPQ user ratings revealed that all IPQ items are highly correlated and per-user score means vary considerably across subjects. The two most plausible explanations for the observed rating behavior are that a) participants were overwhelmed by the overall amount of questions they had to answer after each condition (despite using comfortable electronic questionnaires) and b) participants did not fully understand these more 'exotic' questions and had difficulties to relate them to the test situation experienced. Both causes are a challenge for OV quality assessment because the experience of immersive media like OV has many components one wants to quantify, with some being of more abstract nature than others (compare 'immersion' with 'stalling annoyance'). As lessons learned we therefore recommend to consciously limit the number of rating questions per condition (like in [12]) and to provide extensive explanations to the rating questions used to ensure correct understanding of the concepts targeted.

Another essential, yet missing prerequisite for reproducible research in this domain is access to OV content that is truly suitable for subjective QoE testing – as it is already the case for the domain of 2D video quality testing[3]. Currently, every research group uses a different set of OV clips (downloaded from the Internet in most cases), which limits results comparability and reproducibility as well as experimental results quality, since most publicly available clips were produced for demonstration/entertainment and not for quality testing purposes. Therefore, we hope that in near future, datasets with

[3]See for example http://live.ece.utexas.edu/research/quality/live_video.html or http://ivc.univ-nantes.fr/en/databases/1080i_Videos

free, uncopyrighted (or creative commons protected) OV clips would be made available to the community that feature high production quality, high visual quality (resolution, compression), are interesting to watch, and that cover a variety of relevant content dimensions (camera/object motion types and levels, scene content, recorded realistic vs. artificial/rendered visual style, distribution of points of interest within panorama, projection type, etc.). In this context, the importance of the audio dimension should not be underestimated, since it is know to be an essential component of any immersive experience [18].

Additionally, test setup and execution have to take into account that practical issues participants face with VR and HMDs are the norm, not the exception. Also in our study, users reported issues with the HMD (low perceived resolution, low wearing comfort, signs of cybersickness, etc.) which can lead to higher fatigue or distraction and in turn, less reliable ratings. Newer HMDs might come with less problems, but we strongly recommend to track all these practical sources of bias and noise carefully, taken them into account during analysis and to invest substantial effort to provide a comfortable study setup. Furthermore, (lack of) prior usage experience can in general influence the quality expectations and the resulting quality assessment [19]. In this respect, the novelty of OV streaming as well as HMD-based viewing might be a worthwhile research subject in terms of exploring changes of quality perception and user experience over time, from the introduction of a new technology and the transformation of expectations until its widespread adoption.

## VI. Conclusions

In this paper we have described a first approach towards QoE assessement in a fully immersive setup of omnidirectional video streaming using stereoscopic HMDs as the main use case. Our main focus was on the impact stalling events, including comparisons with traditional 2D streaming, which resulted in interesting findings we extensively discussed. We could show that for reproducible research on this topic the multimedia quality community should work towards *a)* publicly available VR/OV content datasets designed for QoE assessment rather than entertainment; *b)* a detailed OV test protocol with validated questionnaires and appropriate means for voting addressing the specific challenges and constraints of quality testing in VR setups; *c)* a test-bed allowing for immersive media service deployments enabling consistent and coherent experiences across devices and platforms; and *d)* in most cases, a larger number of subjects than for traditional QoE user studies in the audio-visual domain.

In addition, we aim to study wider ranges of stalling patterns (particularly short freezes) and variations in encoding quality as relevant for adaptive streaming scenarios. Furthermore, we will benchmark different approaches towards seamless integration of electronic evaluation questionnaires in HMD-based VR testing applications.

## Acknowledgment

This work has been partially supported by the project NAFI 4.0 funded by the Vienna Business Agency. The authors would like to thank their colleague Georg Regal for his valuable input.

## References

[1] Sandvine, “2016 Global Internet Phenomena Report: Latin America & North America,” 2016, Online: *http://sandvine.com/*.

[2] I. Sodagar, “The MPEG-DASH Standard for Multimedia Streaming Over the Internet,” *IEEE MultiMedia*, vol. 18, no. 4, pp. 62–67, 2011.

[3] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, “A Survey on Quality of Experience of HTTP Adaptive Streaming,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 469–492, Firstquarter 2015.

[4] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, “JPEG Pleno: Toward an Efficient Representation of Visual Reality,” *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, Oct 2016.

[5] B. Choi, Y.-K. Wang, M. M. Hannuksela, and Y. Lim, “ISO/IEC CD 23000-20 Part 20: Omnidirectional Media Application Format (OMAF),” ISO/IEC JTC 1/SC 29/WG 11, Committee Draft, Jan. 2017, work in Progress.

[6] C. W. Fu, L. Wan, T. T. Wong, and C. S. Leung, “The Rhombic Dodecahedron Map: An Efficient Scheme for Encoding Panoramic Video,” *IEEE Transactions on Multimedia*, vol. 11, no. 4, pp. 634–644, June 2009.

[7] M. Yu, H. Lakshman, and B. Girod, “A Framework to Evaluate Omnidirectional Video Coding Schemes,” in *2015 IEEE International Symposium on Mixed and Augmented Reality*, Sept 2015, pp. 31–36.

[8] G. Bjontegaard, “Calculation of average psnr differences between rd-curves,” Video Coding Experts Group (VCEG), Proposal, Mar. 2001, proposal.

[9] P. Rondao Alface, J.-F. Macq, and N. Verzijp, “Interactive Omnidirectional Video Delivery: A Bandwidth-Effective Approach,” *Bell Lab. Tech. J.*, vol. 16, no. 4, pp. 135–147, Mar. 2012.

[10] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, “Initial delay vs. interruptions: Between the devil and the deep blue sea,” in *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, 2012, pp. 1–6.

[11] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, “Quantifying the Influence of Rebuffering Interruptions on the User’s Quality of Experience During Mobile Video Watching,” *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, Mar. 2013.

[12] P. A. Kara, W. Robitza, M. G. Martini, C. T. Hewage, and F. M. Felisberti, “Getting used to or growing annoyed: How perception thresholds and acceptance of frame freezing vary over time in 3D video streaming,” in *Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[13] P. Seeling, “Towards Quality of Experience Determination for Video in Augmented Binocular Vision Scenarios,” *arXiv:1406.0912 [cs]*, Jun. 2014.

[14] ITU-R Recommendation BT.500-13, “Methodology for the subjective assessment of the quality of television pictures.” Geneva, Switzerland: International Telecommunication Union, 2012.

[15] H. Regenbrecht and T. Schubert, “Real and illusory interactions enhance presence in virtual environments,” *Presence: Teleoper. Virtual Environ.*, vol. 11, no. 4, pp. 425–434, Aug. 2002.

[16] T. Hossfeld, R. Schatz, and S. Egger, “SOS: The MOS is not enough!” in *2011 Third International Workshop on Quality of Multimedia Experience*, Sept 2011, pp. 131–136.

[17] E. Upenik, M. Rerabek, and T. Ebrahimi, “A Testbed for Subjective Evaluation of Omnidirectional Visual Content,” in *32nd Picture Coding Symposium*, Nuremberg, Germany, Dec. 2016.

[18] S. Serafin and G. Serafin, “Sound Design to Enhance Presence in Photorealistic Virtual Reality,” in *Proc. 10th International Conference on Auditory Display (ICAD), Sydney, Australia, July 6-9 2004*, 2004.

[19] A. Sackl, R. Schatz, and A. Raake, “More than i ever wanted or just good enough? user expectations and subjective quality perception in the context of networked multimedia services,” *Quality and User Experience*, vol. 2, no. 1, p. 3, 2017.