Tile-based Streaming of 8K Omnidirectional Video: Subjective and Objective QoE Evaluation

Raimund Schatz*, Anatoliy Zabrovskiy†, Christian Timmerer†‡
*AIT Austrian Institute of Technology, †Alpen-Adria Universität Klagenfurt, ‡Bitmovin Inc.
*raimund.schatz@ait.ac.at, †{*firstname.lastname*}@itec.aau.at, ‡christian.timmerer@bitmovin.com

***Abstract*—Omnidirectional video (ODV) streaming applications are becoming increasingly popular. They enable a highly immersive experience as the user can freely choose her/his field of view within the 360-degree environment. Current deployments are fairly simple but viewport-agnostic which inevitably results in high storage/bandwidth requirements and low Quality of Experience (QoE). A promising solution is referred to as *tile-based streaming* which allows to have higher quality within the user's viewport while quality outside the user's viewport could be lower. However, empirical QoE assessment studies in this domain are still rare. Thus, this paper investigates the impact of different tile-based streaming approaches and configurations on the QoE of ODV. We present the results of a lab-based subjective evaluation in which participants evaluated 8K omnidirectional video QoE as influenced by different *(i)* tile-based streaming approaches (full vs. partial delivery), *(ii)* content types (static vs. moving camera), and *(iii)* tile encoding quality levels determined by different quantization parameters. Our experimental setup is characterized by high reproducibility since relevant media delivery aspects (including the user's head movements and dynamic tile quality adaptation) are already rendered into the respective processed video sequences. Additionally, we performed a complementary objective evaluation of the different test sequences focusing on bandwidth efficiency and objective quality metrics. The results are presented in this paper and discussed in detail which confirm that tile-based streaming of ODV improves visual quality while reducing bandwidth requirements.**

***Index Terms*—Omnidirectional Video, Tile-based Streaming, Subjective Testing, Objective Metrics, Quality of Experience**

I. Introduction

Immersive multimedia applications and services are becoming more and more important both from an economic as well as technical/scientific perspective. In particular, 360-degree or omnidirectional videos (ODV) allow the user to freely change the viewing direction in order to immerse into a scene when consuming the content with head-mounted displays (HMDs). The actual implementation options are manifold and raise various technical issues calling for an overall framework enabling the adaptive delivery of ODV [1].

Current deployments (*e.g.*, YouTube, Facebook) adopt existing projection formats (*i.e.*, equirectantular, cubemap, pyramid) but mainly deliver the ODV using simple viewport-agnostic or viewport-adaptive approaches. Although these approaches are simple and easy to implement/deploy, they introduce a series of issues in terms of storage/network requirements and regarding Quality of Experience (QoE). Tile-based streaming of ODV is a promising solution which increases complexity but provides significant improvements with respect to the aforementioned drawbacks [2].

In the past, we witnessed a plethora of (complex) technical solutions proposed in this domain but we also noticed a significant lack of subjective evaluations in this area, specifically when adopting tile-based streaming approaches for ODV. Therefore, in this paper we conduct a lab-based subjective quality evaluation of such approaches for ODV and its combination with existing objective quality metrics.

The remainder of this paper is as follows. Section II provides background on tile-based streaming and reviews related work. In Sections III and IV we present design and results of our subjective ODV QoE lab study, which is followed by an objective evaluation of the test conditions and sources used in Section V. Finally, we discuss the results and present our conclusions including future work in Section VI.

II. Background and Related Work

A. Background: Tile-based ODV Streaming

The most straightforward approach towards streaming of ODV content over the Internet adopts an equirectangular projection format and simply streams the *entire* 360-degree scene/view in constant quality without exploiting and optimizing the quality for the user's viewport. In practice, other approaches exist utilizing other projection formats (*e.g.*, cubemap, equi-angular cubemap) but they are mainly viewport-agnostic and can be referred to as *monolithic* streaming of ODV content. This approach has the drawback of high network bandwidth consumption (particularly when streaming at ultra-high resolutions like 8K), or low QoE (in case of limited bandwidth). The QoE can be increased by adopting viewport-adaptive streaming schemes, which utilize appropriate projection formats (*e.g.*, pyramid) and allocate higher quality to the user's viewport. However, this approach typically requires additional representations for each viewport which again increases storage/network bandwidth requirements.

A promising approach is referred to as tile-based ODV streaming utilizing modern video codecs (*e.g.*, HEVC, VP9, AV1, VVC). Tiles divide a video picture/frame into regular-sized, rectangular regions which are independently decodable, enable efficient parallel processing, and provide entry points for local access, which enables the client to request tiles and quality representations depending on the context conditions including the current viewport. Thus, individual tiles can be requested from different quality representations (*e.g.*, those

978-1-5386-8212-8/19/$31.00 ©2019 IEEE

within the viewport with highest possible quality and neighboring, adjacent tiles with lower quality) or not at all.

In this work, we investigate two generic tile-based streaming strategies that provide the basis for our evaluations, namely *full delivery* and *partial delivery* as defined in [2].

Full Delivery: The client requests/receives a full 360-degree video but with different qualities depending on the user's viewport. A basic strategy could be as follows: all tiles within the user's viewport are requested in the highest possible quality representation while tiles outside the user's viewport are requested in the lowest available quality representation.

Partial Delivery: The client requests/receives a partial 360-degree video which has actual video content corresponding only to the tiles within the user's viewport (but potentially with different qualities). However, tiles outside the user's viewport are not requested at all. While this approach enables optimizing towards the available bandwidth, user head movements may lead to the rendering of "blank" tiles which impacts QoE.

B. Related Work

In the past, several subjective ODV quality studies have been conducted. Schatz *et al.* [3] conducted an ODV QoE assessment lab study, focusing on the impact of stalling using two client setups with a HMD and traditional 2D display. Their work addresses a number of pitfalls in ODV subjective testing and shows that (at least in the case of stallings) HMDs and traditional 2D displays yield very similar QoE results. Singla *et al.* [4] subjectively evaluated the quality of various HEVC-encoded ODVs at different bitrates for two different resolutions (Full-/Ultra-HD). They found that encoding quality/bitrate and content clip had significant impact on QoE, with the influence of resolution – while still statistically significant – being marginal due to the resolution of the Oculus Rift HMD as used in the experiments. Furthermore, they found that for the same reason, encoding bit-rate savings were possible at only marginal QoE reduction. Additionally, they also found that participants' head movement activity patterns were only influenced by the actual content, with horizontal head movement (yaw) dominating in the majority of the clips. In a follow-up study, they also found (marginal) rating differences when using DSIS and Modified-ACR for ODV QoE assessment [5]. Zhang *et al.* [6] performed subjective and objective evaluation of a range of 4K ODVs impaired by different encoders at different bitrates. Similar to [5], they present their own subjective testing method, and they compare the resulting scores with those from traditional methods (*i.e.*, SSCQS, SAMVIQ) as well as objective metrics (*i.e.*, PSNR, SSIM, VQM). The aforementioned studies have in common, that they focus on monolithic, viewport-agnostic streaming scenarios, using a single encoding configuration per clip and very short test sequences (around 10s). Thus, it becomes evident that there is a lack of subjective ODV quality evaluation studies that determine the QoE-influence of encoding quality (and other relevant factors) in the context of *viewport-adaptive tile-based streaming* and that use suitable 8K resolution source content (with a duration longer than 10s) along with high-resolution displays.

III. Subjective Test: Description

In order to assess the impact of different tile-based streaming configurations on the ODV experience, we conducted a lab-based QoE study in 2018 at AIT and AAU, respectively.

A. Study Goal and Research Questions

The aim of the study was to compare different tile-based streaming strategies and parametrizations (that result in different streaming bandwidth requirements) in order to identify the best configuration with regard to their QoS/QoE tradeoff. Our scenarios assume ODV viewing with occasional head movement, *i.e.*, over time, a different part of the content's panorama moves into the center of the viewport which requires previously low-quality background tiles to be suddenly streamed with high quality in order to maintain good QoE.

Beyond benchmarking the two most fundamental tile delivery approaches (full vs. partial), the main research goal was to quantify the QoE impact and acceptability of different video quality encoding levels applied to *(a)* within-viewport tiles (in the case of partial delivery) and *(b)* out-of-viewport tiles (in the case of full delivery), since these result in very different video bitrates (and consequently, very different network bandwidth requirements) as well as different visual quality experiences of the per-tile quality adaption process. This aspect is important, since overall bitrate budget for transmission is a critical limiting factor in ODV delivery. Finally, we wanted to test for any influence regarding the ODV content (in terms of camera movement through space) and of user head turn speed on quality perception of the different tile-based streaming configurations. Thus, our experiment addressed the following three research questions concerning tile-based ODV streaming:

- How does full vs. partial delivery impact ODV streaming QoE? Is there a clear user preference? (RQ1)
- What is the QoE impact of different ODV tile quality encoding levels? (RQ2)
- Do camera movement (static vs. moving) or head turning speed exert an influence on quality perception? (RQ3)

Additionally, a main design goal of the study was to maximize comparability of results across subjects by standardizing the experience as far as possible, *i.e.*, by rendering simulated head motion directly into each processed video sequence (PVS) [7]. We considered this a necessary requirement, since due to the dynamic nature of adaptive tile-based streaming, already small differences in, *e.g.*, timings and head movements, can lead to very different visual experiences. Our evaluation mainly focuses on visual quality as a consequence of certain tile-based streaming approaches. Note that this is also common practice within standardization committees, *e.g.*, when evaluating 360-degree test sequences for the (upcoming) Versatile Video Coding (VVC) standard [8][9].

B. User Study Setup and Test Design

We used ITU-R BT.500-13 [10] and ITU-T P.913 [11] as general guidelines for test setup and procedure. The overall test protocol followed the typical design of subjective lab QoE experiments (cf. Table I), with special attention paid to the

TABLE I: Subjective ODV QoE study procedure.

No.	Phase (duration)	Steps
1	Welcome (5 min)	Briefing, informed consent
2	Setup (5 min)	Technical setup and screening
3	Training Task (5 min)	2 PVS incl. rating task
4	QoE session (30 min)	20 PVS incl. post-stimulus questionnaire
5	Debriefing (3 min)	Feedback & remarks

TABLE II: Post-stimulus rating questions.

Code	Question	Scale
Quality	How do you perceive the overall quality of the video?	ACR-7 Continous (cf. ITU P.851[2])
Acceptance	Is this quality acceptable for you for everyday watching?	Binary (yes/no)

technical setup and training phase to ensure that the test task is well understood and can be performed without hassles. Prior to the actual test session, subjects were screened for correct visual acuity using Snellen charts and for color vision using Ishihara charts.

The main part of each test session was the actual QoE assessment of the different PVS, following a single-stimulus with hidden reference procedure. To reduce contextual effects, presentation order of test conditions was randomized, following a partial factorial design with 20 PVS = 2 SRCs (source clips) x 10 HRCs (hypothetical reference circuits). During each of the 20 test conditions, participants had to watch one 30s ODV clip in order to rate its perceived quality and acceptability. We decided to use a clip duration of 30s in order to accommodate for the different head activity phases (still, turning right, still) and give participants enough time to be sufficiently immersed in the content. Note, that for this reason, we deliberately refrained from using double-stimulus methods like DSIS or M-ACR [5] since these methods are more suitable for shorter clip durations (around 10s). Quality and acceptability dimensions (cf. Table II) were rated via a browser-based post-stimulus questionnaire[1].

C. Test Content

We used two test sequences: “Basketball”[3], which lets the viewer observe people’s actions on a basketball playground and “Community”[4], which provides a 360-degree view from the roof of a car driving through suburbs (cf. Table III). The two SRCs primarily differ in terms of level of recording camera motion: SRC 1 (Basketball) has been recorded with a *static* camera; SRC 2 (Community) features *dynamically moving* car-mounted camera. Our assumption was that the resulting different levels and kinds of overall motion flow in the clips might influence viewer’s perception of the different tile-based streaming configurations. The equirectangular ODVs underlying both SRCs feature 30fps, 6x4 tiling (as recommended in [2]), and the 30s excerpts we used are homogeneous in terms of scene content and action. Both SRCs are available in 8K base resolution at very high encoding quality (Basketball – YUV 4:2:0 uncompressed format and Community – mp4 file, HEVC, 150 Mbit/s), which was highly important for preventing any influence of source resolution on perceived encoding quality (as is easily the case with already compressed 4K source clips, especially when having been downloaded from YouTube), since test participants were supposed to view the rendered test clips (PVS) on a large 4k 65” screen (Sony KD-65X8505B) at 1.5m distance.

TABLE III: Reference Source Sequences (SRC).

SRC	Source Clip	Camera	Resolution	Begin	End
1	Basketball	Static	8K (8192x4096)	0:00	0:30
2	Community	Dynamic	8K (7680x3840)	3:41	4:11

TABLE IV: Hypothetical Reference Circuits (HRC).

HRC	Delivery	Head turn	Quality (QP)
1	full	slow (4s)	low (46)
2	full	slow (4s)	medium (32)
3	full	slow (4s)	perfect (22)
4	full	fast (1s)	low (46)
5	full	fast (1s)	medium (32)
6	full	fast (1s)	perfect (22)
7	partial	slow (4s)	low (46)
8	partial	slow (4s)	medium (32)
9	partial	fast (1s)	low (46)
10	partial	fast (1s)	medium (32)

The two SRCs were processed assuming ten different HRCs derived from permutations of the following three independent variables: *(i)* full vs. partial delivery, *(ii)* tile encoding quality (QP) and *(iii)* head movement speed (cf. Table IV). As regards encoding quality levels, full delivery features optimal quality (QP=22) for within-viewport tiles, while out-of-viewport tiles per default are delivered either with the same (QP=22) or lower (QP=32 or QP=46) quality. A viewport change (as caused by a shift in viewing direction) typically brings former out-of-viewport tiles into focus, with the subsequent segment being loaded at optimal quality, thus resulting in an (eventually) noticeable change of visual quality of the respective tiles over time. Additionally, we also use two different partial delivery settings where out-of-viewport tiles are left grey, and within-viewport tiles are presented at QP=32 or QP=46. These five settings (3 full + 2 partial)[5] are rendered with two different angular head movement speeds (one vs. four seconds) as part of the generic but very common pattern “still (10s) – turn head 90 degree to the right (1s or 4s) – still (19s or 16s)”. We varied head movement speed, because the faster the turn, the more abrupt low-quality or grey tiles enter the viewport, which might result in different QoE ratings. In order to maximize validity and reproducibility of the test setups, we decided to render all three aforementioned variables (including assumed user head movement pattern) as FoV video clips (with resolution: 1936x1088).

IV. Subjective Test: Results

A total of 35 subjects participated in the subjective test: 14 subjects were female and 21 were male, with an average age of 32 and a median age of 33 years. For analyzing the

[1] http://www.thefragebogen.de/, last accessed: Mar 12, 2019.

[3] http://medialab.sjtu.edu.cn/vr8K/index.html, last accessed: Mar 12, 2019.

[4] https://www.itu.int/en/ITU-T/studygroups/2017-2020/16/Pages/video/jctvc.aspx, last accessed: Mar 12, 2019.

[5] We decided to reduce the number of conditions featuring partial delivery to reduce overall likelihood of user boredom and fatigue.

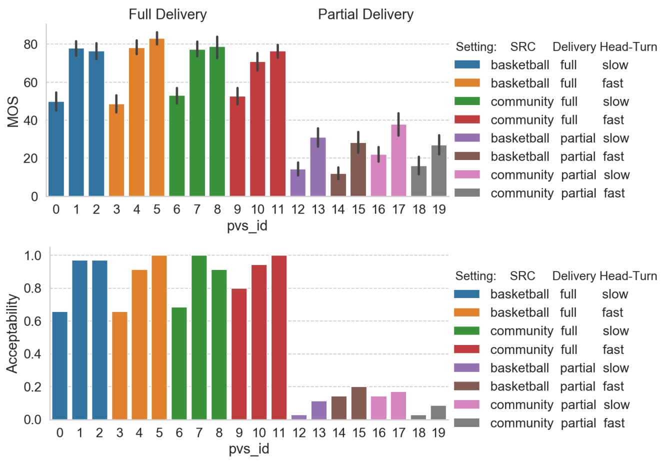

Fig. 1: Mean Opinion Scores (top) for visual quality and acceptance rating proportions (bottom) for all PVSes. MOS is normalized to 0-100, with 0-19% equating "bad" and 81-100% equating "excellent" quality. (MOS CI = 0.95). Each colored group features a combination of delivery type, SRC and head turn speed (see legend), with tile encoding quality being sorted in ascending order within each group.

results data of our subjective QoE evaluation study we used the SensMixed R package[6] and the Python Scientific stack[7].

Figure 1 shows the MOS (visual quality) and acceptability rating results, respectively. When comparing full with partial delivery (RQ1), we see a clear preference of full (pvs_id $<$ 12) over partial delivery (pvs_id $\geq$ 12) by almost all participants. Although we did not test the best encoding level (QP=22) in partial delivery mode, even the conditions featuring medium quality (QP=32) are associated with "poor" to "fair" MOS (and QP=46 even with "bad") and very low acceptance ratios ($<$20%). This result was also confirmed by participants' comments in the debriefing phase, with the majority stating that plain grey out-of-focus tiles are simply too intrusive in order to be acceptable as bandwidth-saving strategy. For the remainder of this section we therefore limit our analysis to full delivery conditions only. Since for full delivery, acceptance ratios were highly correlated with MOS, we will only focus on the latter metric.

As concerns the impact of tile encoding quality (RQ2), full delivery HRCs featuring the lowest quality setting (QP=46, see pvs_id=$\{$0,3,6,9$\}$ in Figure 1) were rated significantly worse than those featuring medium or high quality (QP=$\{$32,22$\}$). In contrast, we found the difference between medium and high quality encoding for out-of-focus tiles to be not statistically significant, but a trend is recognizable. Still, given the small difference in perceived quality/acceptability, these results suggest that QP=32 might be a good balance between reducing out-of-focus tile bitrates and maintaining good QoE when streaming in full delivery mode.

TABLE V: Mixed model analysis for MOS response variable (full delivery conditions only) split into *Fixed Effects* (top) and *Random Effects* (bottom). Stars indicate significance of the respective F-test and Chi-square tests' p-values.

Fixed Effects	**F**	**p**	
quality	164.728	0.000	***
headmov	4.727	0.031	*
content	0.724	0.396	
quality:headmov	0.552	0.577	
quality:content	5.489	0.005	**
headmov:content	4.727	0.031	*
Random Effects	**Chi**	**p**	
quality:user	0.223	0.637	
headmov:user	0.000	1.000	
content:user	0.000	1.000	
quality:headmov:user	0.000	1.000	
quality:content:user	0.000	1.000	
headmov:content:user	0.000	1.000	
user	51.542	0.000	***

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Above results are confirmed by our mixed model ANOVA results (full delivery conditions only), which – among other factors – identifies (encoding) "quality" as factor with highly significant influence on MOS (p<0.001, see Table V). As regards the influence of *head movement speed* and *content type* (camera static vs. dynamic/moving) on the QoE (RQ3), we could confirm head movement speed has exerting statistically significant influence on MOS ratings: fast head turns resulted in worse QoE compared to slow ones as expected, since fast head movements cause a more pronounced presence of low quality tiles in the active viewport. However, this effect is not as significant as the others (p=0.031). To our surprise, we could not identify significant impact of factor "content" per se, *i.e.*, no significant MOS variance can be explained by changing the SRC clip alone. However, the ANOVA results show significant interactions of SRC choice with "quality" (p=0.005) and "headmov" (p=0.031), which means that the type of content at least influenced the impact of out-of-focus tile encoding quality and head turn speed on the QoE. According to our post-hoc analysis (LS-Means) the differential impact of encoding quality on MOS was significantly more pronounced in the case of the rather static SRC 1, which is in line with the general observation that high levels of motion tend to mask quality impairments and artifacts [12].

In addition, we also tested for user-related influences by analyzing the random effects part of our mixed-effects model (cf. Table V). Indeed, the model shows significant "assessor effects" on quality ratings (p<0.001). However, we could not detect any systematic influence of user-related age, gender, or education level. Furthermore, no user reported symptoms of cybersickness, which can be explained by the fact that our setup featured a large screen as display and not a fully immersive HMD.

V. Objective Evaluation: Results

The aim of this section is to present the objective evaluation of the processed video sequences (PVS) utilizing common metrics, *e.g.*, Peak Signal-to-Noise Ratio (PSNR), structural similarity (SSIM) index, and Video Multimethod Assessment Fusion (VMAF), and compare objective results with those from the subjective evaluation.

PSNR and SSIM are well-known objective quality metrics which are easy to calculate (*e.g.*, with FFmpeg) and are often

[6] https://cran.r-project.org/web/packages/SensMixed/index.html, last accessed: Mar 12, 2019.

[7] https://www.scipy.org/, last accessed: Mar 12, 2019.

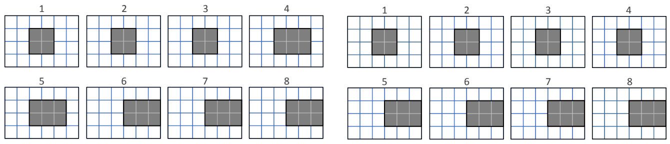

Fig. 2: 10-4-16 secs (slow head turn) scenario (left) and 10-1-19 secs (fast head turn) scenario (right). Segment length 4s; last segment has 2s.

used in order to get a first impression of the picture/video quality. For PSNR we adopted weighted PSNR (wPSNR) [13]. VMAF[8,9] is a relatively new metric which reflects the viewer's perception of streaming services and it has been recently updated to support also 4K content.

For the objective evaluation we used only an excerpt of the PVS. That is, we used only those segments representing head movements taking into account the presence of an adaptive bitrate/streaming logic including buffering because the encoding quality of visible tiles before/after the head movements is at the maximum possible quality level. Taking into account also the segments before/after the head movements – which are basically identically with the SRC – would thus significantly impact the object evaluation. The segment structure for the two head movements are shown in Figure 2.

Full delivery: each segment represents 4s of the actual content and grey tiles represent the viewport (always in best quality; QP=22) while white tiles are outside the viewport (always in lower quality; QP={32,46}).

Partial delivery: within-viewport tiles have qualities (QP={32,46}) while out-of-viewport tiles are represented as grey tiles (*e.g.*, those tiles are not available to the client).

Apparently, head movement begins in the middle of seg.#3 (*e.g.*, after 10s) and would theoretically end after (i) one second (head movement "fast"), which is still within seg.#3 and (ii) four seconds (head movement "slow"), which would be in the middle of seg.#4. In principle, this would allow having the best possible quality for the next segment (*e.g.*, seg.#4 for head movement "fast") and next but one segment (*e.g.*, seg.#5 for head movement "slow") assuming perfect network conditions and zero buffering at the client. As such a scenario is impractical, we decided to add a delay within the PVS to reflect current practice as shown in Figure 2. The area percentage of out-of-focus tiles in the viewport reaches 41% for 10-4-16s scenario and 70% for 10-1-19s scenario.

The results for wPSNR, SSIM, and VMAF of the PVS are shown in Table VI for segments 3-5. As expected, the partial delivery mode has lower quality than the full delivery mode which confirms the results from the subjective evaluation. Interestingly, the difference for SSIM between full and partial is not that high which may indicate that SSIM is probably not a suitable metric in such a scenario. The quality metrics for "all tiles" provide a baseline reference when all tiles within the viewport have the same quality (QP={32,46}). Apparently, the quality is much lower than full delivery mode but still higher than partial delivery mode. The difference between head movements for "all tiles" is obviously imperceptible.

We further note that head movement "fast" has consistently lower quality than "slow" for each test condition (*e.g.*, full/partial delivery and low/medium quality) where, specifically PSNR and VMAF show differences that may also result in differences for subjective quality which is only partially confirmed by our subjective tests. However, head movement "fast" has more tiles in lower quality for a longer time (cf. Figure 2) which also explains this difference.

In a next step we calculate wPSNR, SSIM, and VMAF for the entire 8K equirectangular test sequences with QP={32,46}) for all tiles as such a configuration corresponds to the current deployment practice. It allows for a comparison with tiled streaming configurations as reported above when assuming the same bitrate budget. To calculate VMAF we used downscaled videos as the current VMAF implementation only supports resolutions up to 4K. The results of this evaluation are shown in Table VII ("all tiles..."). The PSNR is consistently lower than when using tiled streaming with full delivery but higher compared to partial delivery (cf. Table VI) assuming the same bitrate budget.

Finally, we show average bitrates used in this evaluation resulting from the QP settings used during encoding (Table VIII). It shows that "Community" has a much higher bitrate than "Basketball" which is due to the nature of its content, *e.g.*, moving camera with higher temporal information than static camera which has lower temporal information at roughly the same spatial information for both sequences. This is confirmed when calculating spatial and temporal information (SI/TI) as follows: Community (SI=46.3/TI=14.9), Basketball (SI=48.0/TI=1.7).

Overall, we observe following results and user preferences taking into account subjective and objective evaluations. In general, users prefer tile-based ODV streaming in full delivery mode followed by non-tiled streaming (*e.g.*, current deployment practice) while partial delivery mode is not acceptable at all as it results in very low QoE. However, in full delivery mode, out-of-viewport tiles still significantly impact the QoE when their encoding quality is low. For medium quality (QP=32), perceived quality is almost as good as for high quality (QP=22) enabling a bitrate reduction of around 50%.

VI. Conclusions, Limitations, and Outlook

In this paper, we addressed the QoE of viewport-adaptive tile-based omnidirectional video (ODV) streaming on behalf of combined subjective and objective evaluation. Our experimental setup addresses – among others – the challenge of presentation consistency (posed by the dynamic behavior of viewport-adaptive ODV streaming) by rendering relevant media delivery aspects (including assumed user head movements and tile quality adaptation) into the processed video sequences used for evaluation. Our results quantify the influence of factors like tile encoding quality, head movement speed, *etc.*and confirm that tile-based ODV streaming indeed has the potential to

[8] https://github.com/Netflix/vmaf, last accessed: Mar 12, 2019.

[9] https://medium.com/netflix-techblog/toward-a-practical-perceptual-video-quality-metric-653f208b9652, last accessed: Mar 12, 2019.

TABLE VI: wPSNR, SSIM and VMAF of FoV videos (results for segments: 3-4-5).

	$wPSNR$		$SSIM$		$VMAF$	
Delivery, Quality (QP), Head turn	**Basketball**	**Community**	**Basketball**	**Community**	**Basketball**	**Community**
all tiles with QP46, slow	30.8	31.7	0.8724	0.9076	48.62	49.36
all tiles with QP46, fast	30.9	31.9	0.8729	0.9114	45.56	49.76
full, low (46), slow	39.3	40.5	0.9783	0.9884	92.37	97.42
full, low (46), fast	36.5	37.2	0.9599	0.9745	83.47	84.77
all tiles with QP32, slow	39.5	40.1	0.9726	0.9749	91.14	93.77
all tiles with QP32, fast	39.6	40.3	0.9730	0.9757	87.85	93.49
full, medium (32), slow	48.1	48.9	0.9952	0.9966	97.58	99.99
full, medium (32), fast	45.0	45.7	0.9913	0.9924	95.26	99.08
partial, low (46), slow	22.8	22.3	0.8420	0.8843	39.98	41.36
partial, low (46), fast	20.5	20.9	0.8115	0.8676	29.82	27.32
partial, medium (32), slow	23.3	23.6	0.9270	0.9434	72.73	77.62
partial, medium (32), fast	20.7	21.3	0.8812	0.9140	55.89	52.54

TABLE VII: wPSNR SSIM and VMAF of 8k equirectangular videos (results for segments: 3-4-5).

	$wPSNR\|SSIM\|VMAF$					
Quality (QP), Head turn	**Basketball**			**Community**		
all tiles with QP46	32.7	0.9281	76.12	33.2	0.9381	77.49
low (46), slow	35.0	0.9564	86.90	35.1	0.9591	86.12
low (46), fast	34.6	0.9523	85.33	34.7	0.9559	84.39
all tiles with QP32	41.2	0.9809	94.98	41.1	0.9797	98.10
medium (32), slow	43.6	0.9880	96.36	43.0	0.9859	99.28
medium (32), fast	43.1	0.9869	96.14	42.6	0.9850	99.04

TABLE VIII: Average bitrate of 8k equirectangular videos (all 8 segments, 30 sec.).

	$avgBitrate(Mbit/s)$	
Quality (QP), Head turn	**Basketball**	**Community**
all tiles with QP46	1.3	3.5
low (46), slow	24.1	38.5
low (46), fast	21.9	35.8
all tiles with QP32	12.8	24.0
medium (32), slow	30.7	52.5
medium (32), fast	29.1	50.4
high (22)	57.3	114.1

achieve substantial bandwidth savings (>50%) while only marginally compromising QoE. Furthermore, we could show that while different subjective and objective metrics used generally agree, there is number of cases where they also show deviant behavior.

Nonetheless, our subjective study comes with several limitations, which should be addressed in future work. Due to practical constraints and to avoid any influence of user fatigue, we tested with only two (yet fundamentally different) SRCs as well as only one head motion pattern (left to right, albeit with two speeds). Furthermore, we deliberately used a large 2D display (instead of an HMD) and rendered head-motion into each PVS (instead of self-induced motion) in order to maximize reproducibility of results at the expense of realism and level of immersion and, thus, generalizability. This approach was chosen to restrict the variety of possible viewing trajectories as observed in real life settings (e.g. [4][5]) to enable a confined evaluation of in-viewport video quality. Similar approaches have been used in the past (cf. VV-360 use cases [7][8][9]). We acknowledge that a detailed study to compare results obtained from different setups (*e.g.*, 2D TV vs. HMD, rendered head motion vs. self-induced, *etc.*) is subject to future work, specifically with respect to peripheral vision. In this sense, we still believe that the results of our subjective and objective evaluation represent a good starting point for further research in this area and that its results can be applied to QoE optimization of ODV streaming systems as well as future ODV quality evaluation studies featuring, *e.g.*, advanced tile-streaming strategies (including viewport prediction) or a larger variety of content types and head movement patterns.

ACKNOWLEDGMENT

This work was supported in part by the Austrian Research Promotion Agency (FFG) under the Next Generation Video Streaming project “PROMETHEUS.

REFERENCES

[1] C. Timmerer and A. C. Begen, “A Framework for Adaptive Delivery of Omnidirectional Video,” *Electronic Imaging*, vol. 2018, no. 14, 2018.

[2] M. Graf, C. Timmerer, and C. Mueller, “Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP: Design, Implementation, and Evaluation,” in *Proc. of ACM MMSys’17*, Taipei, Taiwan, 2017, pp. 261–271.

[3] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo, “Towards Subjective Quality of Experience Assessment for Omnidirectional Video Streaming,” in *Proc. of QoMEX2017*, 2017, pp. 1–6.

[4] 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 *Proc. of QoMEX2017*, 2017.

[5] 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, no. 14, pp. 1–6, 2018.

[6] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang, and Z. Fei, “Subjective and objective quality assessment of panoramic videos in virtual reality environments,” in *Proc. of ICMEW2017*, 2017, pp. 163–168.

[7] P. Hanhart, Y. He, Y. Ye, J. Boyce, Z. Deng, and L. Xu, “360-Degree Video Quality Evaluation,” in *Proc. of PCS2018*, 2018, pp. 328–332.

[8] A. Segall, V. Baroncini, J. Boyce, J. Chen, and T. Suzuki, “Joint Call for Proposals on Video Compression with Capability beyond HEVC,” October 2017. [Online]. Available: http://phenix.it-sudparis.eu/jvet/doc_end_user/current_document.php?id=3361

[9] P. Hanhart, J. Boyce, K. Choi, and J.-L. Lin, “JVET common test conditions and evaluation procedures for 360° video,” October 2018. [Online]. Available: http://phenix.it-sudparis.eu/jvet/doc_end_user/documents/12_Macao/wg11/JVET-L1012-v1.zip

[10] ITU-R Rec. BT.500-13, “Methodology for the subjective assessment of the quality of television pictures,” Geneva, CH, 2012.

[11] ITU-T Rec. P.913 (03/2016), “Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment,” Geneva, CH, 2016.

[12] A. C. Bovik and A. Singhal, “The essential guide to video processing.” *J. Electronic Imaging*, vol. 19, no. 1, 2010.

[13] J. R. Ohm, G. J. Sullivan, H. Schwarz, T. K. Tan, and T. Wiegand, “Comparison of the Coding Efficiency of Video Coding Standards — Including High Efficiency Video Coding (HEVC),” *IEEE TCSVT*, vol. 22, no. 12, pp. 1669–1684, 2012.