

ON THE IMPACT OF VIDEO STALLING AND VIDEO QUALITY IN THE CASE OF CAMERA SWITCHING DURING ADAPTIVE STREAMING OF SPORTS CONTENT

Nicolas Staelens^{*}, Paulien Coppens[×], Niels Van Kets[†], Glenn Van Wallendael[†],
Wendy Van den Broeck[×], Jan De Cock[†], and Filip De Turck^{*}

^{*}Ghent University - iMinds, Department of Information Technology (INTEC), Ghent, Belgium

[×]Vrije Universiteit Brussel - iMinds, Studies on Media, Information and
Telecommunication Research Center, Brussels, Belgium

[†]Ghent University - iMinds, Multimedia Lab, Ghent, Belgium

ABSTRACT

The widespread usage of second screens, in combination with mobile video streaming technologies like HTTP Adaptive Streaming (HAS), enable new means for taking end-users' Quality of Experience (QoE) to the next level. For sports events, these technological evolutions can, for example, enhance the overall engagement of remote fans or give them more control over the content. In this paper, we consider the case of adaptively streaming multi-camera sports content to tablet devices, enabling the end-user to dynamically switch cameras. Our goal is to subjectively evaluate the trade-off between video stalling duration (as a result of requesting another camera feed) and initial video quality of the new feed. Our results show that short video stallings do not significantly influence overall quality ratings, that quality perception is highly influenced by the video quality at the moment of camera switching and that large quality fluctuations should be avoided.

Index Terms— HTTP Adaptive Streaming, Subjective Quality Assessment, Mobile Video, Tablet, Sports

1. INTRODUCTION

Second screens such as tablets and smartphones are becoming omnipresent while watching television¹. The second screens can be used for reasons related to what is happening on the main screen (e.g. find information on the actors or plot, discuss the broadcast on social media) or for other reasons (e.g. online shopping, e-mail) [1].

In order to enrich the overall TV viewing experience and engage the consumers, more custom second screen applications specifically linked to television programs are built. For example, applications have been developed which allow users to participate in TV-quizzes, cast votes in talent shows, and look up additional content and information on the program

being watched. At the same time, an increasing number of video content is also watched on these mobile devices [2] with news, movies, and sports being the most popular content amongst viewers [3]. Especially the latter continues to lead the growth in mobile video traffic [4].

HTTP Adaptive Streaming (HAS) [5] facilitates video delivery over heterogeneous networks to a wide variety of end-devices. In the case of HAS, a video stream is encoded into different *quality levels*. These levels can differ in terms of resolution, frame rate, and/or encoding bitrate. Next, each quality level is splitted into smaller *chunks*, of a fixed duration, which can all be decoded and played independently from another. In HAS, the clients itself are continuously requesting the chunks of the video which should be played. Hence, depending on device and network characteristics, clients can request chunks from different quality levels. For example, when the network is at the verge of being congested, clients can request chunks from a lower quality level in order to ensure playback fluidity [6]. As a result, quality fluctuations and video stallings can occur during playback [7].

The proliferating usage of second screens in combination with video delivery technologies like HAS, opens new opportunities for application developers in order to enhance engagement for remote fans [8] and audiences [9, 10] and enrich their viewing experience.

In this article, we consider a second screen application which allows users to dynamically change the camera angle while watching (live) sports events streamed using HAS to tablet devices. Upon changing the camera angle, new video chunks must be dynamically requested which can result in a video freeze, caused by the time needed to refill the play-out buffer, and a reduction followed by a gradual improvement (ramp up) of the video quality. We investigate the combined effect of video stalling duration and initial video quality level on Quality of Experience (QoE) when changing camera viewpoints. Subjective experiments were conducted using a recently proposed methodology which allows for immersive experiences [11].

¹Also sometimes referred to as 'third screens' if laptop computers are considered to be the second screens.

The remainder of this paper is structured as follows. We start by presenting a selection of related work and best practices for maintaining QoE in the case of HAS in Section 2. Section 3 details the setup of our subjective experiment to assess the influence of video stallings and quality fluctuations during camera switches in live sports event. The results of this study are presented and discussed in Section 4, including the results of the face-to-face interviews which were conducted to further contextualize the experiment. Finally, the article is concluded in Section 5.

2. SUBJECTIVE QUALITY PERCEPTION OF HTTP ADAPTIVE VIDEO STREAMING

As mentioned in the introduction, typical visual impairments during HAS are (i) video stallings, caused by buffer starvations, and (ii) fluctuating visual quality, as a consequence of requesting chunks from different quality levels in order to adapt to changing network conditions. Research has already been conducted on the subjective quality perception of video delivery using HAS.

2.1. A Selection of Related Work

In [12], the authors evaluate the influence of increasing, decreasing, and constant video quality taking into account different content types. Results show that subjects are more sensitive to decreasing quality in the case of sports content and that chunk duration influences quality variation visibility in the case of increasing video quality. The authors also conclude that the impact of content type on QoE should not be ignored. This experiment was repeated in [13] and also included buffering scenarios which shows that buffering events have a significant impact on QoE.

Rodríguez *et al.* [14] assess the impact of quality level switching frequency, type, and temporal location. Here, switching type refers to the direction of the switch (switch up/down) and the distance between the quality levels (jumps)². Results show a negative correlation between switching frequency and users' QoE. Also, quality level switches in the beginning of video playback seem to have a higher negative impact on quality perception. The authors also point to the different impact of switches in spatial and temporal resolution.

The subjective study presented in [7] addresses the combined effect of initial loading time, amount and duration of video stallings, and video bitrate on quality perception. It is found that stallings and video bitrate have an additive impact on quality perception. In terms of video stallings, no influence on the content type was found for high bitrate videos on quality ratings. Furthermore, the duration of the initial loading has an almost neglectable influence on perceived visual quality. No influence of the temporal location of the switches

²This also includes the change in spatial and/or temporal resolution.

was found on quality perception. However, it has been shown that sequence duration impacts quality perception [15].

Staelens *et al.* [16] assessed the effects of bitrate switches, switching frequency, switching range, and buffer starvations, on end-users' QoE in the case of mobile video delivery. The authors show that buffer starvations leading to video stallings should be avoided at all times and that gradual quality switches are preferred over large-range switches. Results also highlight that quality level switches are easier detectable in high motion sequences (including sports). In other research [6], the same authors already showed that especially playback interruptions severely hamper immersion while watching television.

Many other subjective studies have also been conducted. However, providing a full list of state-of-the-art research is out-of-scope of this paper. The interested readers are referred to [17] and [18] for a more comprehensive overview of subjective studies in the case of HAS-based video delivery.

2.2. Best Practices for Maintaining QoE

Based on the literature study, the following initial recommendations for maintaining adequate QoE video delivered using HAS can be formulated.

1. Multiple short quality adaptations should be avoided. As such, adaptation algorithms should not be too optimistic, i.e. not always immediately request the optimal video chunk if resources allow.
2. Large magnitude rate switches are not preferred. Hence, instead of switching between low-quality and high-quality levels immediately, adaptation algorithms should switch to an intermediate rate before switching to a higher rate.
3. During playback, video stallings should be prevented. Therefore, it is advised to increase the initial startup delay if, in turn, this could ensure the absence of re-buffering events.

Remark that these recommendations are valid in the general case of watching continuous video content delivered using HAS.

3. EXPERIMENTAL SETUP

In this work, we want to extend the recommendations listed in the previous section to include guidelines on how to handle stream switches corresponding to switching camera viewpoints during live sports events. Hence, we created several videos containing camera viewpoint changes and conducted a subjective experiment in order to collect user feedback.

3.1. Video Content

For this experiment, we used original multi-camera recordings corresponded with broadcast quality from two Formula 1

Table 1. Parameters used for encoding the source video sequences for simulating video distribution over HAS.

Quality level	Resolution (in pixels)	Frame rate (in fps)	Video Bitrate (in kbps)	Audio Bitrate (in kbps)	Profile
1	480x270	25	500	64	Baseline
2	640x360	25	900	64	Baseline
3	960x540	25	2000	96	Main
4	1280x720	25	5000	128	Main
5	1920x1080	25	8500	128	High

ances and two soccer games. We obtained the original recordings in standard broadcast format (1080i50 4:2:2 YUV). As an intermediate step, we deinterlaced the content and converted them to 1080p25 4:2:0 YUV sequences.

In order to prepare the content for delivery over HAS, we encoded audio and video into different quality levels corresponding to the parameters listed in Table 1. These parameters are a combination of the encoding guidelines specified in Apple’s Technical Note (TN) 2224 and parameters collected from our project partners³.

Video was encoded as H.264/AVC whereas audio was encoded as AAC, all using the *ffmpeg* tool. Finally, we segmented the video into chunks with a duration of two seconds.

3.2. Simulating Camera Viewpoint Changes

During the subjective experiment, we want to ensure all subjects evaluate exactly the same video sequence on the tablet device in order to ensure experiment repeatability. Therefore, instead of giving test persons the control to change the camera viewpoint themselves, we simulate user input automatically. Using the multi-camera video sources described in the previous section, we created 25 clips with a duration of exactly 50 seconds. All the clips were created so that they contained one camera viewpoint switch. This switch occurred between 15 and 30 seconds after video playback started. In our test app, we then included different icons representing the different camera viewpoints as shown in Figure 1. Each time the camera viewpoint changed during video playback, the corresponding icon was automatically highlighted. This was used to indicate to the test subjects that the camera change was intentional.

We consider five different scenarios, as depicted in Figure 2, to evaluate the trade-off between video stalling duration (1, 2 or 4 seconds) and initial video quality after the camera switch. The idea is that stalling duration influences initial video quality, i.e. a longer stalling results in better quality and vice versa.

Ramping the quality up or down takes place in steps of two seconds, corresponding to individual chunk durations. Scenario 5 is a special version of scenario 2 in that it includes

³These correspond to parameters currently used by their mobile TV viewing app.



Fig. 1. Screenshot of the subjective test app.

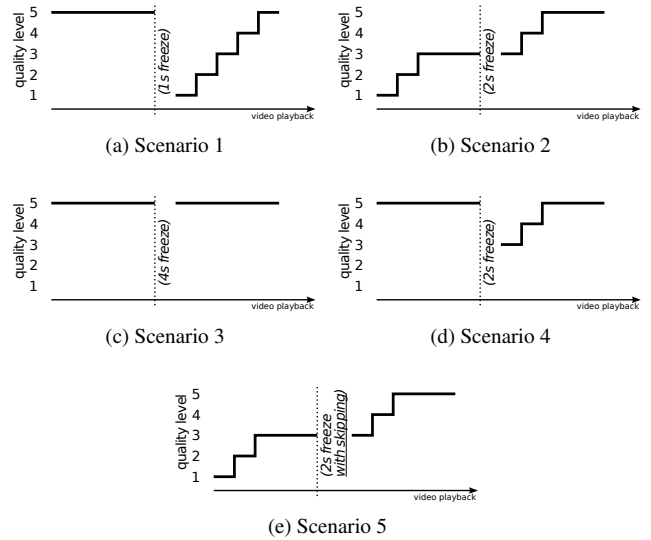


Fig. 2. Scenarios to assess trade-off between stalling duration and starting video quality. The dotted line indicates the occurrence of a camera switch.

stalling with skipping. As such, the duration of the stalling is compensated with a jump in time (corresponding with the freeze duration) in order to maintain the live character of the

video stream.

The sequences were created by concatenating the corresponding individual chunks into one longer playable sequence. In total, this resulted in 125 different video sequences to be evaluated subjectively.

3.3. Subjective Assessment Methodology

Pinson *et al.* [11] recently proposed a novel subjective audiovisual quality assessment methodology which allows for both immersion and conducting the experiment under more natural viewing conditions. The methodology differs from the conventional ones (e.g. methodologies defined in ITU-T Rec. P.940) in the fact that (1) each source video is only shown once during the experiment, (2) longer duration video sequences are considered, and (3) distractor questions are used to shift subjects' focus from pure quality evaluation. The methodology does, however, require a higher number of participants.

Following the authors' guidelines, we created five distinct datasets so that each source video was only shown once per set, but all datasets together counted for all combinations of sources and scenarios. Pinson *et al.* report stable results when using 30 to 40 test subjects in case of four or five impairment scenarios. Hence, for this experiment, we target 40 participants so that each impaired video is evaluated by 8 subjects. One dataset took around 35 minutes to complete.

As shown in Figure 1, we developed our own mobile app to conduct the experiment on fourth generation iPad devices. Before the start of the experiment, subjects received detailed instructions on how to evaluate the different sequences. Also, one training sequence was used to familiarize the subjects with the test software. Concerning the distractor questions, we created a pool of seven questions. After each video sequence, three distinct distractor questions were randomly selected and shown to the subjects. These questions were then followed by the two target questions:

1. How would you rate the overall quality of the video fragment (blockiness, blurring, camera transition, ...)?
[5-Excellent; 4-Good; 3-Fair; 2-Poor; 1-Bad]
2. Did you notice a difference in quality during playback (blockiness, blurring, ...)?
[5-Imperceptible; 4-Perceptible but not annoying; 3-Slightly annoying; 2-Annoying; 1-Very annoying]

Subjects were handed the tablet device and instructed to conduct the experiment in their own typical home environment.

3.4. Test subjects

As mentioned in the previous section, we recruited 40 test subjects in total. Male subjects were aged between 16 and 56 years old with an average age of 28. Female subjects, which counted for 50% of the total number of participants,

were aged between 21 and 57 years old with an average age of 29.

Supplementary to conducting the subjective experiment, 20 face-to-face interviews were organized with the test subjects in order to contextualize the experiment. After conducting the subjective experiment, these respondents were questioned more in-depth about their perception of the video quality and stallings in the test and their experiences with and expectations towards video quality and stallings in general.

4. RESULTS

In this section, we present the results of our subjective experiment and the face-to-face interviews.

4.1. Impact of Video Stalling and Initial Video Quality

In order to determine the reliability of our subjective test data, we first measured the Standard deviation of Opinion Scores (SOS) parameter a as suggested in [19]. For our experiment, we obtain a SOS parameter a equal to 0.2273, which is in the parameter range for subjective user studies considering video streaming [19].

In Figure 3, the Mean Opinion Score (MOS) for each of the scenarios is plotted with a confidence interval (CI) of 95%. The graph shows that scenarios 3 and 4 receive the highest MOS, whereas scenarios 2 and 5 are rated the lowest. Based on the MOS values, the scenarios are ranked as follows (from highest to lowest): 3 - 4 - 1 - 2 - 5.

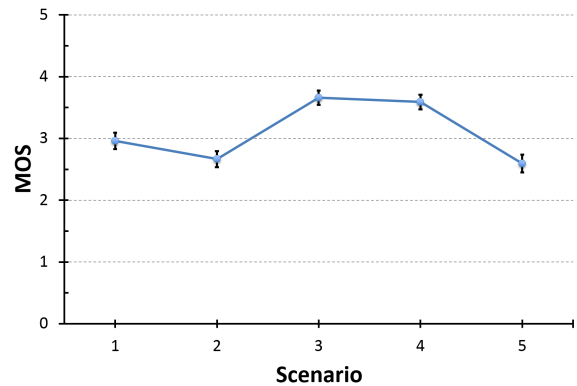


Fig. 3. MOS ratings for each scenario (CI=95%)

Scenarios where the camera switch occurs while watching the video at the highest quality level (cfr. scenarios 1, 3 & 4) are, in general, rated better quality. Furthermore, the results show that the initial video quality after the camera switch has a slightly higher impact on quality ratings compared to the duration of the stalling. In our case, a longer stalling duration is tolerated if this results in better video quality. We performed a Mann-Whitney U Test to check for significant differences

amongst the quality ratings for the different scenarios. Based on our subjective data, we found no significant difference between scenario 3 and scenario 4.

Scenarios 2 and 5 are rated lowest quality. In this case, camera switching occurs while watching video at medium quality. Comparing scenarios 2 and 4 shows that the quality rating is highly impacted by the video quality at the moment of the camera switch. In scenario 4, switching occurs when watching the video at the highest quality which results in better perceived overall quality.

Scenario 5 is identical to scenario 2, except that freezing with skipping is implemented. Hence, after the freeze, the video is automatically skipped two seconds. This technique is often used for maintaining the 'live' character of the video. Based on a Mann-Whitney U Test, we did not find a significant difference between scenarios 2 and 5. As such, in our case, freezing without and freezing with skipping are not rated differently.

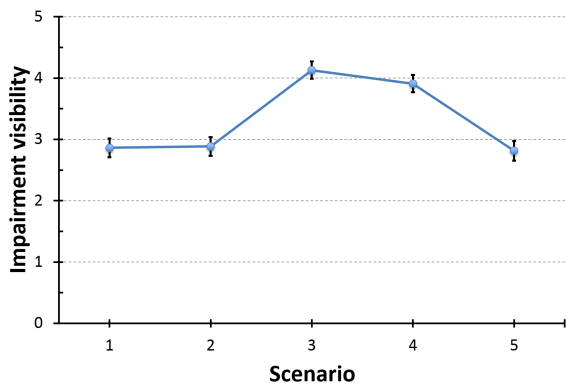


Fig. 4. Average impairment visibility ratings for each of the scenarios. (CI=95%)

In Figure 4, we plot the results of the second quality question concerning quality impairment visibility which confirm our observations.

These results indicate that quality perception is not significantly influenced by stalling duration and that the quality levels of the video before and immediately after the camera switch have an additive effect on perceived quality. Therefore, as an additional best practice guideline for maintaining adequate QoE, it is advised to maintain at least the same quality level of the video after switching camera viewpoint. Starting with a decreased quality level after a camera switch will result in a lower quality rating. In order to keep the video stalling duration limited which is important in the case of live sports events, pre-fetching strategies and workahead buffers could be implemented as suggested by Krishnamoorthi *et al.* [20].

4.2. Contextualisation

In line with the subjective test results, the tolerance amongst the test subjects towards video stallings is higher than a decrease in video quality. Most respondents say their viewing experience is more negatively impacted by a too low video quality, since this can make it difficult to see what is happening onscreen. If video stallings occur, it is still possible to catch up with the video. An extra element that partly explains the higher tolerance towards video stallings, is the fact that many people are used to video stallings in case of (mobile) video streaming. Various respondents say they learned to accept this or they try to handle it (e.g. by letting a video load for some minutes first). In case of a camera switch, some say they even expect a video stalling.

However, in case of a video stalling, two elements are important. First, there is the length of the video stalling. One to two seconds is considered acceptable, but a longer stalling is not. This somewhat contradicts the subjective test findings, which show that quality perception is not significantly influenced by stalling duration. Second, there is freezing with and without skipping. Also in slight contradiction to the subjective test results, freezing with skipping is perceived as highly annoying, since this can cause respondents to miss certain parts of a video (e.g. goal in a soccer game). This clearly shows the added value of conducting face-to-face interviews complementary to gather subjective quality ratings.

In general, however, many respondents admit they are willing to give up a bit of the video quality to be able to watch a video without stallings. Also, a short and low decrease in quality is tolerated when changing camera viewpoint, as long as it is short and gets better afterwards and it does not occur very often. In addition to this, the respondents say that the bigger the difference in video quality is within one video fragment, the less they tolerate this. This is in line with the results of the subjective experiments. A video with a constant lower quality is even preferred over a video in which the video quality fluctuates between very good and very poor.

Lastly, the respondents were asked if they have different video quality expectations for different types of video content. First, there is the length of the video. The longer the video, the higher the video quality expectations. Second, a couple of respondents also pointed out that they have lower video quality expectations for live broadcasts.

5. CONCLUSIONS

In this work, we considered the combined effect of video stalling and video quality in the case of camera switching while watching simulated live sports events delivered over HAS on tablet devices.

We conducted a subjective experiment following a newly proposed methodology which enables immersion. Furthermore, face-to-face interviews took place in order to further

contextualise our experiment.

The results show that especially the quality level of the video right before and immediately after a camera switch has a significant effect on quality evaluation. The influence of stalling duration is negligible. As such, during camera switching, QoE can be maintained by ensuring video quality does not degrade immediately after the switch.

This was confirmed in the face-to-face interviews, in which it is revealed that large fluctuations in video quality within a video are not tolerated. The interviews also revealed a higher tolerance towards video stallings compared to a low video quality. However, most of the respondents also say they are willing to give up a bit of the video quality if this means a video can be watched without stallings.

These results were obtained by simulating camera switches. In order to further quantify QoE, it would definitely be interesting to conduct similar experiments under more realistic conditions, i.e. enabling subjects to change camera view-points dynamically. It is believed that this would lead to new insights, especially concerning the influence of video stalling duration.

Acknowledgement

This research was conducted within the LISSA MIX-ICON project and in close collaboration with Videohouse nv. Within this project, a interactive second screen application for live events was built. The outcome of this paper was implemented in the adaptive streaming algorithm of this second screen application.

References

- [1] Nielsen, "Living Social: How Second Screens are Helping TV Make Fans," Report, August 2014.
- [2] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016," White Paper, February 2012.
- [3] Ooyala, "Q2 2013 Video Index," Report, September 2013.
- [4] Adobe, "The U.S. Digital Video Benchmark - Q4 2013," Report, February 2014.
- [5] T. Stockhammer, "Dynamic Adaptive Streaming over HTTP – Standards and Design Principles," in *Proceedings of the second annual ACM conference on Multimedia systems*. 2011, MMSys '11, pp. 133–144, ACM.
- [6] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën and, B. Vermeulen, P. Lambert, R. Van de Walle, and P. Demeester, "Assessing quality of experience of IPTV and Video on Demand services in real-life environments," *IEEE Transactions on Broadcasting*, vol. 56, no. 4, pp. 458–466, December 2010.
- [7] M.-N. Garcia, D. Dytko, and A. Raake, "Quality Impact due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video Services," in *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 125–130.
- [8] P. Centieiro, "Bringing the Sport Stadium Atmosphere to Remote Fans," in *Proceedings of the 21st ACM International Conference on Multimedia*. 2013, MM '13, pp. 1063–1066, ACM.
- [9] R. van Brandenburg and A. Veenhuizen, "Immersive Second-Screen Experiences using Hybrid Media Synchronization," in *Media Synchronization Workshop*, 2013.
- [10] L. Barkhuus, A. Engström, and G. Zoric, "Watching the Footwork: Second Screen Interaction at a Dance and Music Performance," in *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*. 2014, CHI '14, pp. 1305–1314, ACM.
- [11] M. Pinson, M. Sullivan, and A. Catellier, "A New Method for Immersive Audiovisual Subjective Testing," in *Eighth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2014.
- [12] S. Tavakoli, K. Brunnström, K. Wang, B. Andrén, M. Shahid, and N. Garcia, "Subjective Quality Assessment of an Adaptive Video Streaming Model," in *Image Quality and System Performance XI*, 2014.
- [13] M. Shahid, J. Sjøgaard, J. Pokhrel, K. Brunnström, K. Wang, S. Tavakoli, and N. Gracia, "Crowdsourcing based Subjective Quality Assessment of Adaptive Video Streaming," in *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014.
- [14] D. Rodríguez, Z. Wang, R. Rosa, and G. Bressan, "The Impact of Video-Quality-Level Switching on User Quality of Experience in Dynamic Adaptive Streaming over HTTP," *EURASIP Journal on Wireless Communications and Networking*, no. 1, 2014, to appear.
- [15] P. Fröhlich, S. Egger, R. Schatz, M. Muhlegger, K. Masuch, and B. Gardlo, "QoE in 10 Seconds: Are Short Video Clip Lengths Sufficient for Quality of Experience Assessment?," in *Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2012, pp. 242–247.
- [16] N. Staelens, J. De Meulenaere, M. Claeys, G. Van Wallendael, W. Van den Broeck, J. De Cock, R. Van de Walle, P. Demeester, and F. De Turck, "Subjective Quality Assessment of Longer Duration Video Sequences Delivered over HTTP Adaptive Streaming to Tablet Devices," *IEEE Transactions on Broadcasting*, vol. 60, no. 4, pp. 707–714, December 2014.
- [17] M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, K. Egger, S. and Brunnström, and A. Raake, "Quality of Experience and HTTP Adaptive Streaming: A Review of Subjective Studies," in *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 137–142.
- [18] 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*, to appear.
- [19] T. Hoßfeld, R. Schatz, and S. Egger, "SOS: The MOS is Not Enough!," in *Third International Workshop on Quality of Multimedia Experience (QoMEX)*, Sept 2011, pp. 131–136.
- [20] V. Krishnamoorthi, N. Carlsson, D. Eager, A. Mahanti, and N. Shahmehri, "Quality-adaptive Prefetching for Interactive Branched Video using HTTP-based Adaptive Streaming," in *Proceedings of the ACM International Conference on Multimedia*. 2014, MM '14, pp. 317–326, ACM.