YouTube QoE on Mobile Devices: Subjective Analysis of Classical vs. Adaptive Video Streaming

Michael Seufert*, Florian Wamser*, Pedro Casas†, Ralf Irmer‡, Phuoc Tran-Gia*, Raimund Schatz†

*University of Würzburg
Institute of Computer Science
Germany
{seufert | florian.wamser | trangia}
@informatik.uni-wuerzburg.de

†FTW
Telecommunication Research Center Vienna
Austria
{casas | schatz}@ftw.at

‡Vodafone Group
Research and Development
United Kingdom
ralf.irmer@vodafone.com

***Abstract*—YouTube is the most popular service in the Internet and is increasingly consumed on mobile devices. With emerging adaptive video streaming technology, the question arises whether it should be also employed in the mobile context, which shows different characteristics in terms of display sizes and reliability of Internet connection. This paper compares YouTube QoE on mobile devices for both classical and adaptive video streaming based on a subjective lab experiment, in which different network conditions were emulated. Our results show that adaptive video streaming provides almost excellent results for the poorest network conditions. Thereby, it clearly outperforms classical video streaming, and thus, should be considered to achieve higher QoE in future mobile streaming applications.**

***Keywords*—QoE; Video streaming; YouTube; HTTP adaptive streaming; Video quality; Mobile networks; Subjective study**

I. Introduction

YouTube is among the most popular and volume-dominant services in today's Internet. More than 1 billion users visit YouTube each month and watch over 6 billion hours of video content during this period.[1] Already almost 40% of the watch time is caused by mobile devices, aiming to go even higher in near future. Understanding the performance and quality of YouTube traffic is thus paramount for ISPs; specially for mobile operators, who must handle the huge surge of traffic with the constraints of cellular networks, while keeping their customers' Quality of Experience (QoE) at acceptable levels.

Since January 2013, YouTube has introduced HTTP adaptive streaming for the desktop version based on MP4 movie fragment boxes. An interruption of the video playback (i.e., stalling), is avoided by changing the quality level according to the network conditions, made possible by the division of the streaming content into small segments (MP4 fragment boxes) with different resolutions and encoding bit rates. From the perspective of mobile network, the question arises whether the adaptive streaming should be also used in a mobile context. There, the devices have other display characteristics and the Internet connections are less reliable, which may require many changes of the quality per video. Ultimately, this must be evaluated by a subjective user study on QoE on mobile devices for the YouTube mobile app.

In this paper, a comprehensive evaluation of user satisfaction for the YouTube app with constant streaming quality (classical streaming) is done in comparison to one with enabled adaptive streaming technology based on the YouTube HTML5 API. Both approaches have different pros and cons. In classical streaming, stalling might occur with short term interruptions of the playback while maintaining constant image quality. In adaptive video streaming on YouTube, the video resolution is adjusted at run-time. Thus, each approach influences the user-perceived quality in a different way.

This paper addresses the research question whether adaptive video streaming yields better QoE results than classical video streaming on mobile devices. We confirm that stalling impacts the user experience of YouTube on mobile devices. Given the small display sizes, we find that video quality changes, as currently performed by the YouTube API, have an almost negligible impact on mobile YouTube QoE. Thus, we show that there is a major difference in terms of user experience when comparing both streaming technologies.

The remainder of the paper is structured as follows. First, related work is summarized in Section II in order to give an overview of work related to HTTP video streaming and its user-perceived influence factors. Second, the test setup and the study is described that used in the course of this paper in Section III. Afterwards, the results are presented in Section IV. Finally, the paper is concluded in Section V.

II. Related Work

The subjectively perceived quality of HTTP video streaming is already well-investigated by the research community. Initial delays and stallings constitute the key performance indicators of video streaming QoE [1], [2]. In [3], authors showed that initial delays should be kept short but most users tolerate them because they are used to them. Stalling, on the other hand, has a huge impact as already little stalling severely degrades the perceived quality [1].

Whilst adaptive streaming concepts are known for a long time, their broad commercial usage has only risen recently, and the topic is getting more and more attention within the research

[1]https://www.youtube.com/yt/press/

community. Authors in [4] found that quality adaptation could effectively reduce stalling by 80% when bandwidth decreased in a mobile environment, and was responsible for a better utilization of the available bandwidth when bandwidth increased. However, quality switches have an impact on perceived quality themselves, as they increase or decrease the video quality according to the switching direction [5]. Authors in [6] found that only the time on each quality layer has a significant impact on QoE, but not the number of quality switches. In [7], authors found that resolution is a key parameter for video QoE on small displays. They concluded that low resolutions contributed to enhanced eyestrain of the subjects. [8] presented a model for mapping resolution to MOS. Finally, [9] predicted the MOS of adaptive streaming by temporal pooling of objective per-frame metrics. A more comprehensive survey of the QoE of adaptive streaming can be found in [10].

When it comes to the specific study of YouTube QoE in mobile networks and mobile devices, there are some recent papers worth mentioning. In [11], authors study the characteristics of YouTube traffic for both Android and iPhone mobile devices connected to a cellular network, showing that mobile devices have a non-negligible impact on the characteristics of the downloaded traffic (for example, in terms of video resolution and flow download control behavior). Closer to our work, authors in [12] describe a subjective QoE evaluation framework for mobile Android devices in a lab environment. Additionally, they perform some basic QoE-based study on the classical, non-adaptive YouTube streaming using very low bit rate videos (less than 100 kbps) but neglecting the impact of download throughput in contrast to our study. Authors in [13] go a step beyond and study the QoE of YouTube in mobile devices through a field trial, but completely neglect the analysis and impact of adaptive streaming as we do. Finally, in [14], authors take a further step and introduced an online monitoring system for assessing the QoE of YouTube in cellular networks using network-layer measurements only.

III. Study Description

The subjective experiment conducted in this work consists of 52 participants interacting with two different versions of the YouTube mobile application (classical and adaptive version), while experiencing different bandwidth profiles in the background download data connection.

For the classical streaming use case, the original YouTube mobile application was used with static settings. The video is shown with constant resolution and quality, which can lead to stalling events during playback when the network bandwidth is not sufficient to present the playback quality. The resulting video quality was 720p for all measurements with this mobile application. The adaptive approach takes advantage of the capabilities of the MPEG Dynamic Adaptive Streaming over HTTP (DASH) approach [15] to dynamically adjust the streaming quality, i.e., the video resolution, on a segment basis. The optical design of the original YouTube application has been completely rewritten in a separate application. An Android web view browser element was embedded, such that HTML5 video playback is possible, including adaptive streaming from YouTube according to the DASH approach. More details on this application can be found in [16].

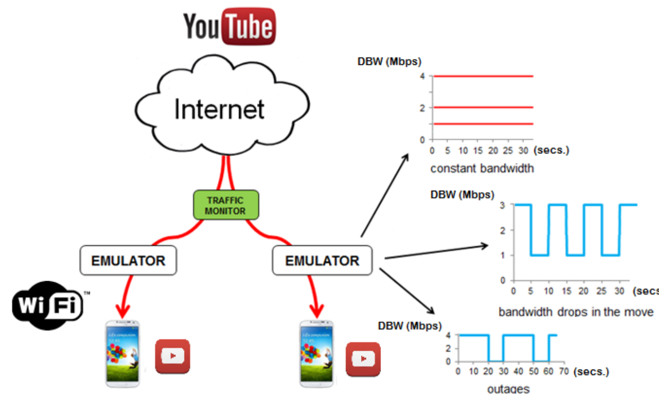


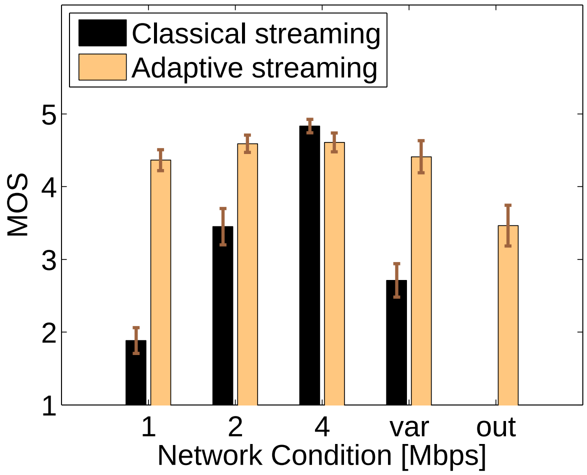
Fig. 1. Testbed for subjective analysis of YouTube QoE in mobile devices.

YouTube video ID	Resolution	Avg. video bit rate
6pxRHBw-k8M	720p	1.5 Mbps
iNJdPyoqt8U	720p	1.5 Mbps
kObNpTFPV5c	720p	1.7 Mbps
QS7lN7giXXc	720p	1.7 Mbps
suWsd372pQE	720p	1.6 Mbps

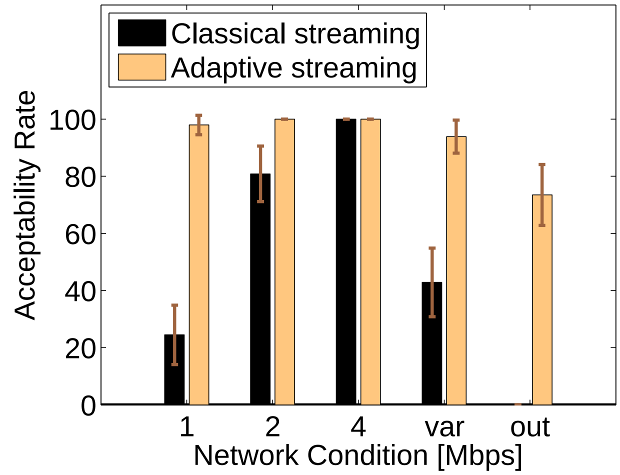
TABLE I. Video content used in the lab experiment. YouTube videos correspond to 4K ultra-HD videos (2160p), which are down-scaled to HD resolution (720p) due to the device's display capabilities.

Figure 1 depicts a high-level diagram of the experimental testbed employed in the subjective tests. Android smartphone devices are used in the study (Samsung Galaxy S4, OS Android 4.4 KitKat). Devices are connected to the Internet through separate WiFi access networks. The downlink traffic between the different evaluated services and the devices is routed through a modified version of the very well known NetEm network emulator [17] so as to control the different access network bandwidth profiles under evaluation. Participants are instructed to evaluate a simple video watching scenario, consisting of watching two minutes HD YouTube videos, considering both the usage of fixed quality HD video streaming and adaptive video streaming (i.e., DASH). Three different bandwidth profiles are instantiated at the network emulators: (i) constant downlink bandwidth: 1 Mbps, 2 Mbps, and 4 Mbps; (ii) fluctuating downlink with varying bandwidth ("var"): downlink bandwidth is periodically increased from 1 Mbps to 3 Mbps for periods of 5 seconds, 3 times per minute. The resulting average downlink bandwidth is 1.5 Mbps; (iii) downlink bandwidth outages ("out"): downlink bandwidth drops from 4 Mbps to 0 Mbps for periods of 10 seconds, twice per minute. In this case, the resulting average downlink bandwidth is 2.7 Mbps.

Tests are performed in a dedicated lab for subjective studies, compliant with the QoE subjective studies standards [18]–[20]. Table I reports the five YouTube videos used (YouTube video IDs), which include four mainly nature-themed clips and a movie trailer. Videos correspond to 4K ultra-HD videos (i.e., 2160p), which are down-scaled to HD resolution (i.e., 720p) due to the device's display capabilities (i.e., screen size and resolution). To better understand the obtained results, the table also reports the average video bit rate of the corresponding videos, which is in all cases around 1.6 Mbps. Each video is



(a) *Rate the overall quality (stability, delay, disturbance, etc.).* The overall quality was rated on a continuous scale ranging from *1 (bad)* to *5 (excellent).*



(b) *Do you consider the overall quality acceptable or unacceptable? Acceptable/Unacceptable ("Unacceptable" means that you would cancel the session and try again later.)* The plot shows the overall acceptability rate, i.e., percentage of users who answered *Acceptable*.

Fig. 2. Rating of overall experience. The advantages of the adaptive streaming approach are notable: users declare a very high quality, even for a download bandwidth of 1 Mbps. In the classical approach, such a low bandwidth configuration results in very poor user experience.

linked to a single bandwidth condition, and it is watched twice using both the classical and the adaptive YouTube applications.

In terms of participants' demographics, 29 participants were female and 23 male, the average age was 32 years old, with 40 participants being less than 30 years old. Around half of the participants were students and almost 43% were employees, and 70% of the participants have completed university or baccalaureate studies.

Finally, in order to provide QoE feedback, participants were instructed to rate their *overall experience* (rate the overall quality) according to a continuous ACR Mean Opinion Score (MOS) scale [18], ranging from "bad" (i.e., MOS = 1) to "excellent" (i.e., MOS = 5). MOS ratings were issued by participants through a custom questionnaire application running on separate laptops, which pops up immediately after a condition was tested. In addition, participants answered/rated the following six questions/features for each video session: (i) *acceptability* (do you consider the overall quality acceptable?); (ii) *initial playback delay* (did you notice any delay at the beginning of the video?) ; (iii) *initial playback delay annoyance* (did you perceive this delay as disturbing?); (iv) *stallings* (did you notice any interruptions or stops during the playback?); (v) *stallings annoyance* (did you perceive these interruptions as disturbing?); (vi) *video image quality* (rate the image quality of the video).

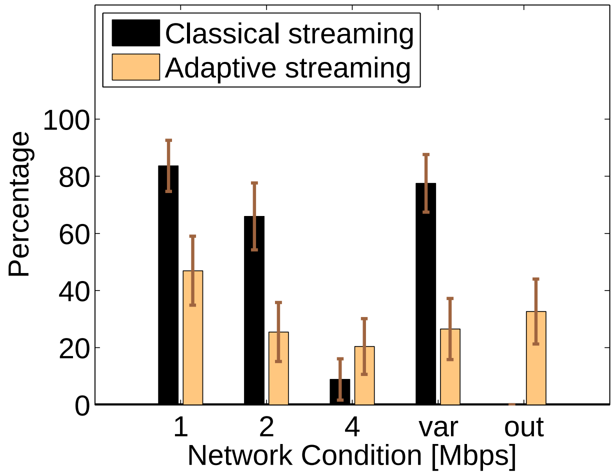
IV. Results

To compare classical and adaptive video streaming in mobile network environments, the results from the two different YouTube apps are compared. In the following, the influence of different network conditions on the overall quality of classical and adaptive video streaming will be analyzed for the tested network conditions. Moreover, the influence on initial delay and stalling will be investigated, and finally, the influence on video image quality will be quantified. Note that the outage condition ("out") was only tested in the adaptive case to obtain and compare results for adaptive streaming, which includes stalling. Hence, there are no results for classical streaming.

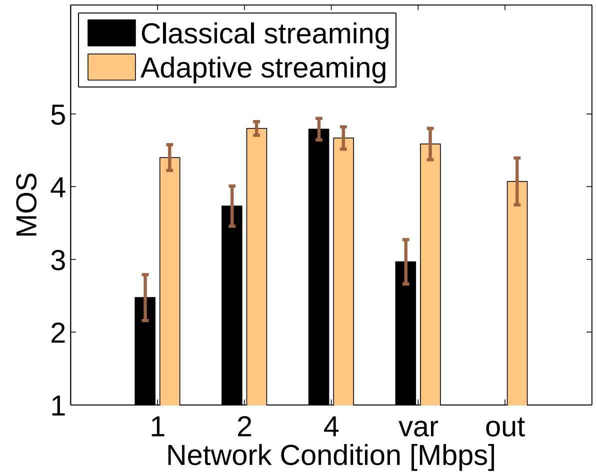
A. Overall Experience Analysis

Figure 2 shows the overall quality of the video streaming, which was rated after each test condition. On the x-axis of Figure 2a, the different network conditions are visible as described in Section III. On the y-axis, the mean opinion score (MOS) of the participants and the 95% confidence intervals are depicted. A strong influence of the bottleneck bandwidth can be observed for classical streaming. The MOS score achieved for a bandwidth of 1 Mbps is very low, close to 1.9. With the bandwidth increasing up to 4 Mbps, the MOS also increases up to 4.83. The variable network condition was rated slightly worse than the 2 Mbps (MOS: 3.45) and only achieves a MOS score of 2.71. Adaptive streaming conditions, in contrast, are rated high almost independent of the network condition. For all three tested constant bandwidth conditions and the variable condition the MOS scores are above 4.36. Only the outage condition has a significantly worse MOS, which is however still comparable to a classical streaming with 2 Mbps.

Similar results can be observed when looking at the overall service acceptability in Figure 2b. It shows the percentage of users who rated the service quality under the respective network condition as being acceptable, and the corresponding 95% confidence intervals. For 4 Mbps both streaming approaches perform very well as the available bandwidth is sufficient and the resulting service quality is accepted by all users. When considering lower bandwidths, again the acceptability of classical streaming is depending on the network condition. Streaming over a 2 Mbps link is still accepted by



(a) *Did you notice any delay at the beginning of the video? Yes/No* The plot shows the percentage of noticed initial delays.



(b) *Did you perceive this delay as disturbing? (Please rate "Not disturbing at all" if you did not notice any delay.)* The initial delay was rated on a continuous scale ranging from *1 (very disturbing)* to *5 (not disturbing at all)*.

Fig. 3. Rating of initial delay. Experienced users might be used to initial delays, such that they do not notice them unless delays are unusually long.

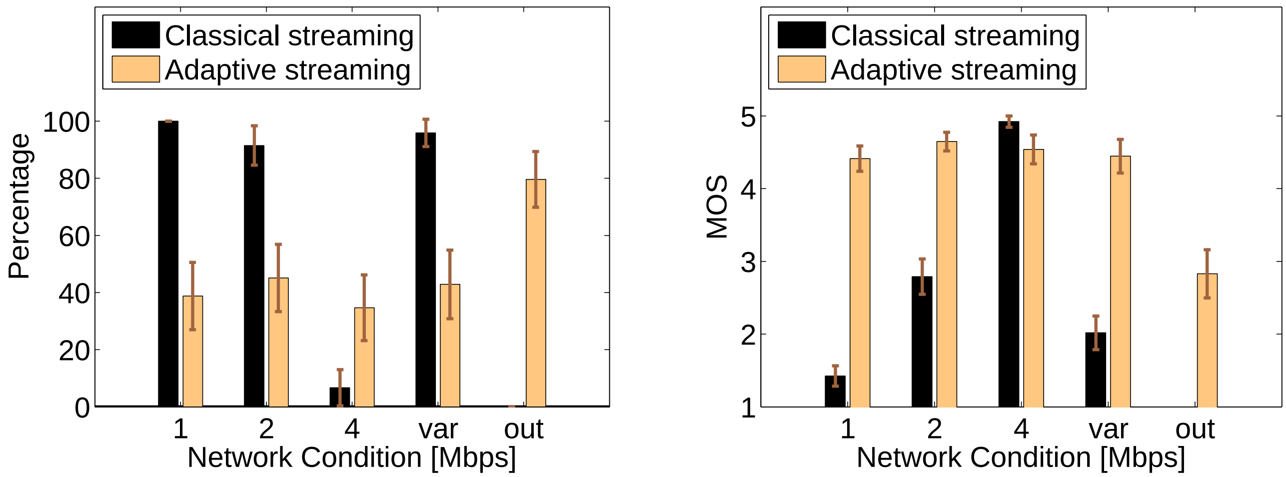
80.85% of the users, whereas only 24.49% of the users accept the quality resulting from classical streaming over a 1 Mbps bottleneck. The varying condition accounts for an acceptability rate of 42.86%. Confirming the results from the MOS scores, adaptive streaming reaches very high percentages of service acceptability for all network conditions. All users are satisfied with the streaming quality at 2 Mbps and 4 Mbps, 97.96% of the users accept the quality at 1 Mbps and 93.88% accept it in the varying condition. Even the outage condition has a high acceptance rate of 73.47%.

The evaluation of both questions already show interesting conclusions. Classical streaming performs well if sufficient bandwidth is available, i.e., surpassing the average video bit rate of around 1.6 Mbps (cf. Table I) by some margin. For low bandwidth conditions, it is outperformed by adaptive streaming. The results show that adaptive streaming can significantly increase the acceptability of video streaming, e.g., from 24.49% up to 97.96% in the 1 Mbps condition. In the following, we will further investigate this remarkable finding by looking more closely into different aspects of video streaming. Therefore, we will analyze the participants' answers to questions related to initial delay, stalling, and image quality of the video streams.

B. Initial Delay Analysis

Initial delay refers to the time between the video request and the actual start of the video playback. During this period, video data is downloaded to the client's playout buffer and decoded. Depending on the application policy, the video playback is delayed until the playout buffer is filled with certain amount of playtime. Figure 3a shows the percentage of users who noticed initial delay in the video session. Note that initial delay is always present due to the technical reasons described above. Hence, some users might be used to initial delays, such that they do not notice them unless delays are unusually long. This consideration has to be kept in mind when looking at the following percentages of noticed initial delay. For both streaming types, it can be observed that a higher bandwidth leads to less noticed initial delay. This is due to the fact that a higher bandwidth leads to shorter download times and thus shorter initial delays. If the bandwidth was set to 4 Mbps, initial delays were only perceived by 8.89% of the users for classical streaming, which is the lowest overall percentage. Initial delays of adaptive streaming are still noticed by 20.41%. For all other conditions, initial delays of adaptive streaming are less noticed than the corresponding initial delays of classical streaming. At most 46.94% of the people perceive initial delays in the case of a 1 Mbps bottleneck, whereas the percentages for classical streaming are significantly higher, e.g., 83.67% for 1 Mbps.

When looking at the quality ratings of the initial delay, similar results can be observed. Figure 3b shows the MOS values for the disturbance of initial delay ranging from 1 (very disturbing) to 5 (not disturbing at all). Initial delays of adaptive streaming reach high values above 4.40 for all network condition, which indicate that those initial delays are not perceived as being disturbing. For classical streaming the ratings follow the expectations and are high for high bandwidth conditions (MOS 4.79 for 4 Mbps) due to shorter initial delays, and low for low bandwidth conditions (MOS 2.48 for 1 Mbps), which indicates long initial delays. However, the users did not indicate that those initial delays are severe quality degradations, which confirms the findings of [3]. Still, it can be followed that in most cases (except 4 Mbps condition), adaptive streaming achieves a better initial delay compared to classical streaming. A closer look has to be taken at the outage condition. It has a MOS of 4.07, which is less than the MOS of the 1 Mbps condition (MOS 4.40), although less people noticed the initial delay (32.65% in the outage condition compared to 46.94% for 1 Mbps). The reason for this might be the different starting points of the video playback, such that in some cases the initial delays coincided with the outages, which led to a much longer initial delay.



(a) *Did you notice any interruptions or stops during the playback? Yes/No* The plot shows the percentage of noticed stallings.

(b) *Did you perceive these interruptions as disturbing? (Please rate "Not disturbing at all" if you did not notice any interruptions.)* The stalling was rated on a continuous scale ranging from *1 (very disturbing)* to *5 (not disturbing at all)*

Fig. 4. Rating of stalling. During a stalling event, the playback is paused until the buffer is filled with a sufficient amount of video data to resume the playback.

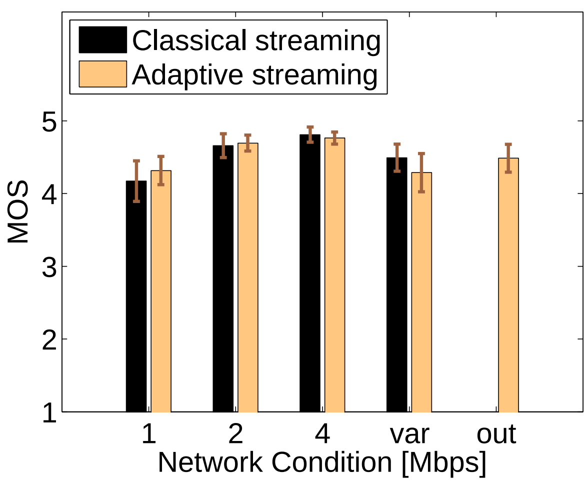


Fig. 5. *Rate the image quality of the video.* The image quality was rated on a continuous scale ranging from *1 (bad)* to *5 (excellent).*

C. Stalling Analysis

Next, users were asked to rate the stallings of the video session. Stalling refers to the interruption of playback due to a playout buffer under-run. When the available network bandwidth is lower than the video bit rate, the buffer empties which will eventually lead to stalling. During a stalling event, the playback is paused until the buffer is filled with a sufficient amount of video data to continue the playback. It follows that the network bandwidth has a big impact on the number and lengths of stalling events during video streaming. Figure 4a depicts the percentages of users which noticed stalling during a video session. It can be seen that stalling events are noticed by almost all users during classical streaming in 1 Mbps, 2 Mbps, and variable condition. Only the 4 Mbps bottleneck is sufficient to avoid stalling of classical streaming. Note that some users have dissenting perceptions of stalling compared to the vast majority. These might be due to missing a stalling event because of distraction, or exceptional stalling because of technical problems. When looking at adaptive streaming of these conditions, interestingly, stalling is noticed by around 40% of the users in each condition. In the outage condition, stalling is perceived by 79.59%. This is surprising because stalling/no stalling could be well distinguished in the classical streaming conditions. It indicates that stalling is not that obvious for adaptive streaming, which will be confirmed by investigating the stalling ratings.

These stalling ratings are depicted in Figure 4b. Again the MOS values for the disturbance of stalling are ranging from 1 (very disturbing) to 5 (not disturbing at all). Stalling of adaptive streaming is not considered disturbing reaching mean opinion scores of at least 4.41 for 1 Mbps, 2 Mbps, 4 Mbps and variable condition. [1] investigated the impact of stalling on the perceived quality and found that there is an exponential relationship between stalling parameters and MOS. They concluded that users tolerated at most one stalling event of up to three seconds length. Thus, it follows that adaptive streaming causes only very short stalling events, short enough that they are not even noticed by all users. The reason could be that quality adaptation (i.e., switch between two quality levels) causes a short glitch of playback, which is considered as stalling. Only for the outage condition, stalling is rated disturbing (MOS: 2.83), which indicates that additionally longer stalling events occurred. For classical streaming, stalling cannot be avoided by adaptation, hence the ratings follow the expectations and fit to the model described in [1]. As length and number of stalling events depend on the network bandwidth, a low bandwidth of 1 Mbps leads to much stalling and a low MOS of 1.42. The variable condition achieves a MOS of 2.02, and a 2 Mbps bottleneck link results in a MOS of 2.79. As most users noticed no stalling for the 4 Mbps condition, no disturbance can be rated by the participants. All in all, it is obvious that adaptive streaming is able to avoid long stalling events in most network conditions. Instead, only playback glitches, i.e., short stalling events, occur during quality adaptation, which are only noticed by some participants and are not considered disturbing.

D. Video Image Quality

In the following, the image quality of video streaming will be investigated in more detail. Figure 5 shows how the participants rated the image quality of each condition. Surprisingly, the image quality is rated good for all conditions. To be more precise, the MOS of the video image quality is at least 4.17 for all conditions, and thus considered to be good. Even though adaptive streaming reduces the resolution for low bandwidth conditions, the resulting image quality is rated almost equal to classical streaming without image quality changes. It follows that the reduction of the resolution was not perceivable in the conducted tests, which could be due to the small display size of the used smartphone.

V. Conclusion

In this work, a lab study was conducted to compare classical and adaptive video streaming in mobile environments. Therefore, different network conditions were emulated and YouTube videos were streamed to a smartphone, either using classical or adaptive streaming. The participants watched each video and rated the quality of the streaming afterwards.

It was shown that adaptive video streaming clearly outperforms classical streaming in terms of user perceived quality. Especially the overall quality was perceived better for low bandwidth conditions. Only for high bandwidth conditions, classical streaming was able to come out slightly ahead. Confirming previous findings, the experiment showed that stalling due to insufficient bandwidth had a severe impact on YouTube QoE on mobile devices. However, adaptation was able to mitigate the impact of stalling. Although the results of the lab study implied that each quality switch (i.e., each quality adaptation) was accompanied by a short glitch of playback, which was noticeable for some participants, it was not as disturbing as normal stalling. Despite the resolution changes during adaptive streaming, the image quality was still rated in a comparable way to classical streaming. Thus, in the tested mobile devices with small displays, the impact of video quality reduction was negligible and was not noticed by the participants.

These findings clearly suggest the use of adaptive video streaming in mobile environments, as it reduces initial delay and stalling while providing a high image quality as perceived by the users. Future mobile video streaming application should therefore rely on adaptive streaming technology to achieve higher QoE and user satisfaction.

Acknowledgment

This work was partly funded in the framework of the EU ICT projects SmartenIT (FP7-2012-ICT-317846) and mPlane (FP7-2012-ICT-318627), and was partly performed within the project ACE 3.0 at the Telecommunications Research Center Vienna (FTW). The authors alone are responsible for the content.

References

[1] T. Hoßfeld, R. Schatz, M. Seufert, M. Hirth, T. Zinner, and P. Tran-Gia, "Quantification of YouTube QoE via Crowdsourcing," in *IEEE International Workshop on Multimedia Quality of Experience - Modeling, Evaluation, and Directions (MQoE)*, Dana Point, CA, USA, 2011.

[2] R. K. P. Mok, E. W. W. Chan, X. Luo, and R. K. C. Chan, "Inferring the QoE of HTTP Video Streaming from User-Viewing Activities," in *ACM SIGCOMM Workshop on Measurements Up the STack (W-MUST)*, Toronto, ON, Canada, 2011.

[3] 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 *4th International Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, 2012.

[4] J. Yao, S. S. Kanhere, I. Hossain, and M. Hassan, "Empirical Evaluation of HTTP Adaptive Streaming Under Vehicular Mobility," in *10th International IFIP TC 6 Networking Conference: Networking*, Valencia, Spain, 2011.

[5] B. Lewcio, B. Belmudez, A. Mehmood, M. Wältermann, and S. Möller, "Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission," in *IEEE International Workshop Technical Committee on Communications Quality and Reliability (CQR)*, Naples, FL, USA, 2011.

[6] T. Hoßfeld, M. Seufert, C. Sieber, and T. Zinner, "Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming," in *6th International Workshop on Quality of Multimedia Experience (QoMEX)*, Singapore, 2014.

[7] H. Knoche, J. D. Mccarthy, and M. A. Sasse, "How Low Can You Go? The Effect of Low Resolutions on Shot Types in Mobile TV," *Multimedia Tools and Applications*, vol. 36, 2008.

[8] L. Janowski and P. Romaniak, "QoE as a Function of Frame Rate and Resolution Changes," in *3rd International Workshop on Future Multimedia Networking (FMN)*, Krakow, Poland, 2010.

[9] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp, "To Pool or not to Pool: A Comparison of Temporal Pooling Methods for HTTP Adaptive Video Streaming," in *5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.

[10] 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, 2015.

[11] J. J. Ramos-Muñoz, J. Prados-Garzon, P. Ameigeiras, J. Navarro-Ortiz, and J. M. López-Soler, "Characteristics of Mobile YouTube Traffic," *IEEE Wireless Communications*, vol. 21, 2014.

[12] I. Ketykó, K. De Moor, T. De Pessemier, A. J. Verdejo, K. Vanhecke, W. Joseph, L. Martens, and L. De Marez, "QoE Measurement of Mobile YouTube Video Streaming," in *3rd Workshop on Mobile Video Delivery (MoViD)*, Florence, Italy, 2010.

[13] G. Gómez, L. Hortigüela, Q. Pérez, J. Lorca, R. García, and M. C. Aguayo-Torres, "YouTube QoE Evaluation Tool for Android Wireless Terminals," *EURASIP Journal on Wireless Communications and Networking*, vol. 164, 2014.

[14] P. Casas, M. Seufert, and R. Schatz, "YOUQMON: A System for On-line Monitoring of YouTube QoE in Operational 3G Networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, 2013.

[15] International Standards Organization/International Electrotechnical Commission (ISO/IEC), "23009-1:2012 Information Technology – Dynamic Adaptive Streaming over HTTP (DASH) – Part 1: Media Presentation Description and Segment Formats," 2012.

[16] F. Wamser, M. Seufert, P. Casas, R. Irmer, P. Tran-Gia, and R. Schatz, "YoMoApp: a Tool for Analyzing QoE of YouTube HTTP Adaptive Streaming in Mobile Networks," in *European Conference on Networks and Communications (EuCNC)*, Paris, France, 2015.

[17] S. Hemminger, "Network Emulation with NetEm," in *6th Australias National Linux Conference (LCA)*, Canberra, Australia, 2005.

[18] International Telecommunication Union, "ITU-T Recommendation P.800: Methods for Subjective Determination of Transmission Quality," 1996.

[19] ——, "ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications," 2008.

[20] ——, "ITU-T Recommendation P.1501: Subjective Testing Methodology for Web Browsing," 2013.