UNIVERSITÄT AUGSBURG
Universitätsbibliothek

OPEN ACCESS
OPUS AUGSBURG

The impact of adaptation strategies on perceived quality of HTTP adaptive streaming

Sebastian Egger, Bruno Gardlo, Michael Seufert, Raimund Schatz

Angaben zur Veröffentlichung / Publication details:

Egger, Sebastian, Bruno Gardlo, Michael Seufert, and Raimund Schatz. 2014. "The impact of adaptation strategies on perceived quality of HTTP adaptive streaming." In *VideoNext '14: Proceedings of the 2014 Workshop on Design, Quality and Deployment of Adaptive Video Streaming, 2 December 2014, Sydney, Australia*, edited by Mahbub Hassan, Alic C. Begen, and Christian Timmerer, 31–36. New York, NY: ACM. https://doi.org/10.1145/2676652.2676658.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under these conditions:
Deutsches Urheberrecht
Weitere Informationen finden Sie unter: / For more information see:
https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publiz/

The Impact of Adaptation Strategies on Perceived Quality of HTTP Adaptive Streaming

Sebastian Egger[†*]
[†]AIT Austrian Institute of Technology
Innovation Systems - Technology Experience
Modecenterstrasse 17 / Objekt 2
A-1110, Vienna, Austria
sebastian.egger@ait.ac.at

Bruno Gardlo[*]
[*]FTW
Donau-City-Strasse 1
A-1220, Vienna, Austria
gardlo@ftw.at

Michael Seufert[‡]
[‡]Institute of Computer Science
University of Würzburg
Am Hubland, D-97074 Würzburg, Germany
seufert@informatik.uni-wuerzburg.de

Raimund Schatz[*]
[*]FTW
Donau-City-Strasse 1
A-1220, Vienna, Austria
schatz@ftw.at

ABSTRACT

Changing network conditions like bandwidth fluctuations and resulting bad user experience issues (e.g. video freezes) pose severe challenges to Internet video streaming. To address this problem, an increasing number of video services utilizes HTTP adaptive streaming (HAS). HAS enables service providers to improve Quality of Experience (QoE) and resource utilization by incorporating information from different layers. However, these adaptation possibilities of HAS also introduce new perceivable impairments such as the fluctuation of audiovisual quality levels over time, which in turn lead to novel QoE-related research questions. The main contribution of this paper is the formulation of open research questions as well as a thorough systematic user-centric analysis of different quality adaptation dimensions and strategies. The underlying data has been acquired through two crowdsourcing and one lab study. The results provide guidance w.r.t. which encoding dimensions are combined best for the creation of the adaptation set and what type of adaptation strategy should be used. Furthermore it provides insights on the impact of adaptation frequency and the true QoE gain of adaptation over stallings.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Evaluation

Keywords

QoE, HTTP adaptive streaming, crowdsourcing, laboratory studies, video quality, time-varying quality

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in:
VideoNext'14, December 2, 2014, Sydney, Australia.
Copyright 2014 ACM 978-1-4503-3281-1
http://dx.doi.org/10.1145/2676652.2676658

1. INTRODUCTION

Nowadays, streaming video is the most dominant application in the Internet. This dominance will even rise from 57% in 2012 to 69% in 2017 according to [3]. Two thirds of this traffic will then be delivered by content delivery networks like YouTube, Netflix, or comparable services.The major difference to traditional video streaming is the use of HTTP streaming over TCP as delivery method. In contrast to video streaming over UDP, where packet losses resulted in audio and video distortions, the packet retransmission characteristic of TCP results in practical no packet losses but rather long delivery delays. In video services, these long delays translate in temporal impairments such as stalling (i.e., the interruption of playback due to empty playout buffers) or long initial delay, in case large playout buffers have to be initially filled.

Recently, several of aforementioned services have switched to HTTP adaptive streaming (HAS) [20] as its default delivery/playout method. HAS requires the video to be available on the server in an adaptation set that consists of video representations of different bitrates, and split into small chunks each containing a few seconds of playtime. After filling the playout buffer with the initial bitrate the client starts the video playout. Further, it measures the current bandwidth and/or buffer status and requests the next chunks of the video in an appropriate bit rate, such that stalling is avoided and the available bandwidth is utilized best. In terms of user perceived quality, this delivery method results in perceivable quality switches for the end user, in addition to initial delay and stallings.

This trend can not only be observed with YouTube and Netflix, which are prominent examples, but nowadays an increasing number of video applications employ HAS, as it has several more benefits compared to classical streaming. First, offering multiple bit rates of video enables video service providers to adapt the delivered video to the users' demands. As an example, a high bit rate video, which is desired by home users typically enjoying high speed Internet access and big display screens, is not suitable for mobile users with a small display device and slower data access. Second, different service levels and/or pricing schemes can

be offered to customers. For example, a user can increase or decrease the video quality during playback if desired, and can be charged in the end of a viewing session taking into account the consumed service levels [21]. Third and most important, the current video bit rate, and hence the demanded delivery bandwidth, can be adapted dynamically to changing network and server/CDN conditions. If the video is available in only one bit rate and the conditions change, either the bit rate is smaller than the available bandwidth which leads to a smooth playback but spares resources which could be utilized for a better video quality, or the video bit rate is higher than the available bandwidth which leads to delays and eventually stalling, which degrades the Quality of Experience (QoE) severely (e.g., [**?**, 19]). Thus, adaptive streaming is assumed to improve QoE of video streaming. However, from a perceptual point of view there are no clear results or guidelines regarding the performance of different adaptation algorithms in terms of perceived quality.

2. RELATED WORK

Recent work on the performance analysis of HAS solutions can be differentiated along technical and perception based quality assessment. Technical analysis such as [2]mainly concentrate on optimal switching decision strategies for optimization of bandwidth utilization or other network related parameters. The perception based analysis set out to analyse the QoE impact of adaptation parameters and dimensions, and the QoE performance of existing implementations and adaptation strategies. Throughout this paper we are going to concentrate on perception based results.

For proper assessment of video QoE in HAS systems it is important to understand the differences to more classical video QoE results and assessment methodologies which are rather based on signal fidelity. The authors in [1] have shown that adaptation strategy related parameters (stalling, representation switches) have considerable influence on resulting QoE and more importantly, that they have to be considered on a larger time scale (up to some minutes) than video encoding related parameters (resolution, frame rate, quantization parameter, bit rate) which only influence in the order of a few seconds. In addition to those larger time scales, also impairments, such as stalling and quality switches, have not received much attention in more traditional quality assessment methodologies for subjective testing [12]. Hence, proper assessment methodologies that consider such requirements are missing.

Similar to non-adaptive HTTP streaming, waiting time related impairments like initial delay and stalling are of considerable interest in HAS systems as well [4]. The work in [9] has analysed the trade-off between those two impairments directly. In the context of HAS, studies on optimal segment length and buffer size [24], also identified initial delays as an impairment type to be considered. Concerning stalling events during the video playout, it is generally assumed that they decrease QoE severely and that quality adaptation performs better. However, empirical evidence for this assumption is sparse. The data in [17] even suggests that a low number of stalling events is not rated worse than throughput equivalent quality adaptation towards lower quality. Even more interestingly there are to the best of our knowledge no results available that systematically assess the trade-offs between such waiting time related impairments and quality switching.

Results reported in [15] showed that quality switches in active adaptation are perceived as a degradation itself. However, often quality switches are inevitable due to changing bandwidth conditions. If the throughput conditions are very volatile, the questions arises if the adaptation algorithm should immediately switch or better try to hold certain quality levels for a longer time in order to minimize the impairment caused by the switching itself. In that respect [25] investigated the adaptation of image quality for layer-encoded videos. They found that the frequency of adaptation should be kept as small as possible. On the other hand, the results in [18] showed that higher switching frequencies are not penalized in terms of QoE if the duration spent on high quality is sufficiently long. If a variation cannot be avoided, its amplitude should be kept as small as possible. Thus, a stepwise decrease of image quality is rated slightly better than one single decrease as shown in [25]. Also [7] compared smooth to abrupt switching of image quality. They confirmed that down-switching is generally considered annoying. Abrupt up-switching, however, might even increase QoE as users might be happy to notice the visual improvement.

Whereas, the above mentioned works rather address the client side switching logic, also on the server side certain decisions have to be taken in the content preparation stage. For the encoding of the different representations three dimensions can be considered: spatial, temporal, quantization as well as combinations thereof. A number of studies has analysed related trade-offs for the spatial and temporal dimension [5, 16, 23]. Although the setups as well as the used content differed largely between the different studies their results converge in the way that a combination between the two dimensions performs better than one-dimensional encoding for videos of identical resulting bit rate. In [22], trade-offs between resolution and frame quality are investigated. They found that a small resolution (without upscaling) and high image quality is preferred to a large resolution and low frame quality for a given bit rate. [14] compared different combinations of resolution, frame rate, and pixel bit rate, which resulted in similar average video bit rates. They found that at low bit rates a larger resolution is preferred and thus frame rate should be decreased. At high bit rates, frame rate is more important and pixel bit rate should be decreased to achieve a high perceived quality.

Summarizing, several research questions are still open for video QoE in HAS: **RQ1:** What is the QoE trade-off between quality adaptation and an equivalent stalling event? **RQ2:** Does more frequent quality switching result in worse QoE than less frequent quality switching? **RQ3:** What is the QoE gain of smooth quality switching over abrupt switching? **RQ4**: What is the optimal combination of content encoding dimensions for generating the adaptation set?

3. STUDY DESCRIPTION

In order to investigate the presented research questions, three successive studies were designed. As guideline for the content preparation we selected a mobile access scenario which covers a bandwidth range in which five bandwidth levels exist (5: 2300 kbps, 4: 1400 kbps, 3: 700 kbps, 2: 350 kbps, and 1: 128 kbps). It is assumed that the bandwidth varies along these levels, and thus, the video bit rate has to be adapted to the new bandwidth by a single adaptation step. The video bit rate is always kept equal to the bandwidth level. The adaptation of the video bit rate is

done either in the spatial or image quality domain. Additionally, stalling can happen in case of buffer underflow. Several works, e.g., [14], already found that the content influences the perceived quality. Therefore, we will not explicitly investigate this influence factor in this work but use three different content classes (action, cartoon, and sports) for the experiments. Hence, each bar represents the average across these content classes throughout this paper.

For each video bit rate level the test videos were prepared beforehand with the FFmpeg encoder. As starting point (reference) H.264 videos in 720p were used. For each envisaged bitrate level the respective video chunks were produced according to the encoding settings given in Table 1. The chunk lengths (4 sec, 5 sec and 10 sec cf. Table 1) were chosen to be inline with current HAS solutions [13]. The resulting video chunks were then used as basis for the longer video sequences according to the quality switching patterns displayed in Figure 1 and used in the three studies (cf. Table 2).

	Level	QP	FR	Resolution	Target Bitr.
ref	5	29	25	1280x720	2300
spatial	4	29	25	896x504	1400
	3	29	25	512x288	700
	2	29	25	256x144	350
compr.	4	33	25	1280x720	1400
	3	40	25	1280x720	700
	2	47	25	1280x720	350
multi s-s-c	2	36	25	512x288	350
multi c-c-s	2	40	25	768x432	350

	Level	CRF	FR	Resolution	Target Bitr.
ref	5	26	25	1280x720	2400
spatial	4	26	25	854x480	1250
	3	26	25	640x360	800
	2	26	25	426x240	430
	1	26	25	256x144	195

Table 1: Video encoding parameters for all chunks used in the discussed studies. The upper part of the table corresponds to the chunks used in CS Study 1 and the laboratory study (chunk length = 10 sec), whereas the lower part corresponds to the chunks used in CS Study 2 (chunk length = 5 sec for all quality profiles except the switching frequency profiles where the chunk length was 4 sec).

Regarding the difference between quality adaptation and stalling: In case of video stallings the quality does not change if the bandwidth condition changes, but the download time for a given video part changes. Therefore, if the bandwidth drops, the download takes longer than the video playout time which results in a stalling event. The required stalling duration for a bandwidth change is computed as the difference between download time and playout time. Stalling events with these durations were then introduced in the middle of the video clips.

$$Dur_{stalling} = \frac{bitrate_{high} \cdot (t_{stall})}{bitrate_{low}} - (t_{stall}) \qquad (1)$$

To compute the proper stalling duration Equation (1) can be used, with $bitrate_{high}$ being the initial bitrate that would lead to a stalling event, $bitrate_{low}$ being the bitrate the system would switch down to in order to prevent a stalling event

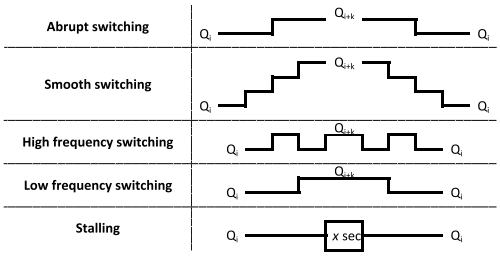

Figure 1: Different ypes of quality switching used across the studies. Q_i and Q_{i+k} correspond to the (quality) levels as described in Table 1

and t_{stall} the point in time where the stalling would happen (assumed in the middle of the clip length). For example, a bandwidth drop from level 5 to level 4 in the middle of a 20 s video clip would be equivalent to constant quality on level 5 and a stalling event of $\frac{2300\,\text{kbps} \cdot 10\,\text{s}}{1400\,\text{kbps}} - 10\,\text{s} = 6.4\,\text{s}$.

Through the remainder of the paper we will us the following notation to describe the quality levels contained in each adaptation pattern (cf. Figure 1): A) for *abrupt* switching (spatial) *2-5* describes a quality switch from (quality) level 1 to level 5 in the middle of the video file, B) *smooth* switching (spatial) *2-5* contains one chunk on level 2, one chunk on level 3, one chunk on level 4 and one chunk on level 5) C) for *high* and *low frequency* switching *2-5-2* denote a quality alternation between level 2 and level 5. In case of low frequency switching only two quality switches occur, whereas for the high frequency case six quality switches take place. For both cases the video duration is 70 sec.

For the combination of encoding dimensions (RQ4) we produced two different versions of the multidimensional combination. For quality level 2 the first one (multi s-s-c) assumed two spatial quality degradations (720p→504p and 504p→288p) followed by an image compression quality degradation (from 288@QP29→288@QP36) resulting in a 288p video with QP36 and a resulting bitrate of 350 kbit/s. The second one (multi c-c-s) assumed two image compression quality degradations (720p@QP29→720p@QP33 and 720p@QP33→720p@QP40) with a further spatial quality degradation (720p→432p) resulting in a 432p video with QP40 and a resulting bitrate of 350 kbit/s. For the creation of the video chunks in CS Study 2 we used slightly different encoding parameters as the lower part of Table 1 shows. However, as we compare results regarding their QoE performance always within a study, this does not harm the validity of the conclusions drawn later in the paper.

Type	Study	Level differences
Abrupt switching	1,2,3	1,2,3,4
Smooth switching	2,3	2,3,4
High frequency switching	2,3	1,2,3,4
Low frequency switching	2,3	1,2,3,4
Stalling	1	1

Table 2: Distribution of adaptation patterns per study

3.1 Study 1: Single quality switch

Currently there is little basic understanding what factors influence overall QoE quality adaptation. Thus, the first study took a bottom-up approach and asked users to rate short video clips which had only one single quality adaptation step (cf. abrupt switching in Figure 1) or a correspond-

ing stalling event (cf. stalling in Figure 1). The main goal of this study was to obtain a general overview of the impact of different adaptation types, resulting in a large number of conditions to be tested. Therefore, the study was conducted via crowdsourcing as described in [10]. To ensure arrival and playout of non-distorted videos, the video under to be evaluated was downloaded to the local cache of the respective crowd worker. After the successful download of the video the playout was started. For the computation of the crowd workers reliability we included consistency questions throughout the test as described in [10].

For the actual test we used more quality levels and combinations as described in Table 1[1], resulting in 85 different test conditions. For three content classes this equaled 255 different test clips. The campaign was run for six months and 10737 ratings were collected. 32% of the ratings were submitted from participants which were classified as reliable[2]. Thus, 3483 reliable ratings of test conditions could be used, resulting in around 40 valid ratings per test condition.

3.2 Study 2: Multiple quality switches

In the follow-up crowdsourcing campaign, the focus was on multiple adaptation steps within a video clip. To limit the number of conditions, only spatial adaptation was considered for the two content classes sport and action. The influence of smooth switching was investigated by comparing step-wise (abrupt) and stair-wise (smooth) adaptation in 20 s video clips (cf. abrupt and smooth switching in Figure 1). Moreover, the influence of adaptation frequencies was tested by comparing clips containing either 2 or 6 quality switches (cf. high and low frequency switching in Figure 1). All in all, 46 test conditions were created in the same way as for the first campaign.

In this campaign, a new online reliability computation as introduced in [6] was used. This resulted in a lower number of unreliable ratings (compared to study 1) and hence reduced campaign execution times. Within 25 days, 1593 ratings could be collected. 86% of the ratings were considered reliable which results in 1377 ratings, i.e., around 30 ratings per condition.

3.3 Study 3: Laboratory Study

In order to confirm the results from the crowdsourcing studies in a controlled test environment, a lab study was conducted. 34 participants (16 male, 18 female with a mean age of 38,1 years) rated 60 conditions. The encoding settings for the chunks were identical as for the CS Study 2 (cf. Table 1). In terms of resulting quality profiles, abrupt switching, smooth switching and high and low-frequency switching profiles as described in Figure 1 were used. Additionally, the multidimensionally encoded clips were used.

4. RESULTS

The results of the aforementioned studies are presented in bar plots that depict differences between the different switching strategies and statistical results of our analysis of variance (ANOVA) in the related tables.

[1]due to space limitations we only include the clip settings and discuss the results relevant for comparison with the other studies presented in this paper

[2]This number seems pretty low for conventional QoE testing, however such low numbers are common in crowdsourcing studies as described in [10].

RQ1: What is the QoE trade-off between quality adaptation and an equivalent stalling event? To quantify QoE differences between quality adaptation and stalling events , we compare pairs consisting of a video with a quality switch (compression or resolution) and a video with stalling. The duration of the stalling event equals the additional playout (or download) time caused by the lower bitrate after the switch (cf. Section 3) In Study 1 we considered only quality switches based on resolution changes from level 5 (reference) to level 4 and 3. Figure 2 depicts the results, where each quality switch is compared to the corresponding stalling pattern. These results show that a quality switch with a magnitude of one level (5→4 and 4→3) yields comparable results as the respective stalling impaired video. The used quality switches of only one quality level might seem to be of limited amplitude only, however we decided to use these quality switches as larger quality switches would result in very long stalling events. E.g. a quality switch from 5→3 would result in a stalling of approx. 22 sec, which is pretty high for videos of 20 sec length.

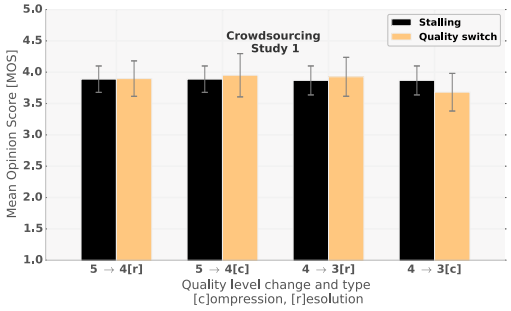

Figure 2: Stalling vs. adaptation: stalling event is not worse than quality adaptation

Seq.	5-4[s]	5-4[c]	4-3[s]	4-3[c]
F	0.002	0.091	0.090	0.930
P	0.961	0.764	0.764	0.339

Table 3: Statistical test results for crowdsourcing study 1 (CS1) shown in Figure 2

RQ2: Does more frequent quality switching result in worse QoE than less frequent quality switching? In order to answer this research question we compared video sequences with two quality switches to sequences with 6 quality switches but the same video duration (cf. high and low frequency switching in Figure 1). In addition the amplitude of the quality switch was varied. In Study 1, quality switches were achieved by changing video resolution whereas in Study 2 uses changes in compression settings. The results in Figure 3 show that more frequent switching did not lead to worse QoE ratings in both studies. These results are contradictory to results presented in [25] and [18] so far. Our most plausible explanation is the fact that for our experiments the chunk lengths were 10 sec (Study 3) and 4 sec (Study 2), whereas [25] and [18] used considerably lower chunk lengths. Hence, even the high-frequent switches in our studies were not really annoying the subjects.

RQ3: What is the QoE gain of smooth quality switching over abrupt switching? For this comparison, abrupt switches were achieved by one switch in the

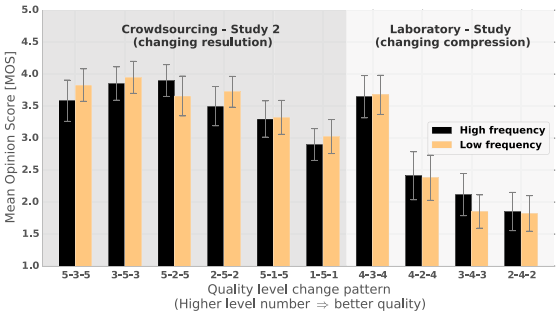

Figure 3: Switching frequency has no (measureable) negative impact

	Crowdsourcing Study 2				Laboratory Study					
Seq.	4-3-4	4-2-4	3-4-3	2-4-2	5-3-5	3-5-3	5-2-5	2-5-2	5-1-5	1-5-1
F	0.02	0.01	1.65	0.02	1.44	0.29	1.49	1.36	0.17	0.48
P	0.891	0.912	0.203	0.881	0.234	0.591	0.227	0.247	0.896	0.489

Table 4: Statistical test results for Figure 3

middle of the video, whereas smooth switches were implemented through several smaller quality switches of one level only throughout the video duration. Surprisingly, smooth switching does not yield statistically significant better QoE results as can be seen in Figure 4. Only for down switching (5-2 in CS 2 and 4-2 in the Lab) smooth switching performs slightly better in terms of QoE compared to abrupt switching. These results are in contrast to several results from related work such as [18]. We explain that by the fact that the switching period was rather long in our videos (4 sec in Study 1 and 10 sec in Study 1) compared to switching periods in the related work. We chose these chunk lengths as they are typical for current HAS implementations.

RQ4: What is the optimal combination of content encoding dimensions for generating the adaptation set? In our studies, the single-dimension videos were created with achieving the bitrate target by only reducing the respective dimension. In contrast, the multi-dimension videos used a combination of two dimensions to reach the bitrate target. Independent of the direction or the amplitude of the switch, Figure 5 shows that the multi-dimensional adaptation results in better QoE scores across both study contexts. As already mentioned at the beginning of this section, one can observe that that the crowdsourcing results tend to be less critical than the lab results which featured lower scores on average. Our results provide a good indication that for the creation of the adaptation set, a combination of multiple encoding dimensions yields best results. However, these results are only a first indication and further studies are needed to answer this question sufficiently.

An additional finding from Figure 5 is that identical video samples from both studies (3→2 and 2→3) show that the results form Study 3 (light brown and orange bars) are rated considerably worse than in the CS Study 2 (darker bars). These differences are based on the fact that crowd workers seem to be less critical in terms of media quality as also shown in [8, 10]. This is explained by the fact that for lab QoE assessments trainings of the subjects before the actual tests take place as recommended e.g. in [12] and [11], which leads to user ratings that are well distributed along the scale. This is not possible in the crowdsourcing studies: here, such trainings cannot be implemented due to the limited number

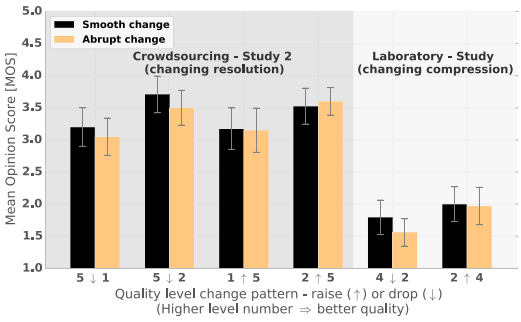

Figure 4: Smooth switching does not provide significantly better QoE than abrupt switching

	Crowdsourcing Study 2				Laboratory Study	
Seq.	5-1	5-2	1-5	2-5	4-2	2-4
F	0.54	1.13	0.01	0.18	1.95	0.03
P	0.466	0.289	0.915	0.669	0.168	0.881

Table 5: Statistical test results for Figure 4

of video sequences that can be used for each session as required to avoid loss of crowd worker attention and ensure reliable results [10]. Hence, rating scale usage is not as well distributed as in Study 3. In terms of absolute MOS values such differences can be critical. As we are comparing results of different adaptation strategies only *within* studies in terms of their perceived quality, this effect does however not harm the validity of the conclusions drawn within this paper regarding QoE differences between switching strategies.

5. CONCLUSION

In this paper we have investigated the QoE impact of temporally changing quality in the context of http adaptive video streaming services. In order to cover the most relevant parts of the fairly large 'haystack' of different adaptation combinations we have conducted three complementary studies in crowdsourcing and lab contexts. In terms of quality profiles we included abrupt quality switches, smooth quality switches, high and low frequency switching as well as respective stalling conditions.

The results of this systematic approach provide new insights on user perception of time-varying video quality as well as on optimal adaptation strategies: In terms of content preparation our results show that a multi-dimensional

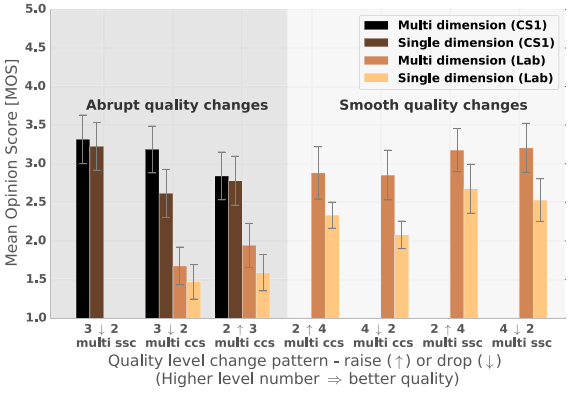

Figure 5: Multi-Dimension Adaptation performs better

	Crowdsourcing Study 2			Laboratory Study					
Seq.	3-2	3-2	2-3	3-2	2-3	2-4	4-2	2-4	4-2
F	0.16	6.58	0.07	1.49	3.48	4.8	9.55	5.36	9.87
P	0.688	0.012	0.878	0.226	0.066	0.032	0.003	0.024	0.003

Table 6: Statistical test results for Figure 5

adaptation strategy yields better QoE scores for identical video bitrates than single-dimensional strategies. In contrast to the widespread assumption that smooth switching performs better in terms of QoE, the presented results reveal that smooth switching performs only slightly better in the case of adapting towards lower quality.

Another unexpected result was the finding that frequent quality adaptation is not perceived considerably worse than videos where are less often quality changes happen. This might be based on that fact that we used chunk lengths typically used in current HAS implementations of 4 sec and 10 sec, which are considerably longer than those mostly used in related work.

Finally, the comparison between QoE perception of videos with quality adaptation vs. videos with one stalling of according length showed that, against the common assumption that stalling is always perceived much worse, in the tested cases stalling performed equally well with video adaptation in terms of resulting QoE.

The presented results are a first thorough analysis of several different impairment profiles very likely to appear in real-world HTTP adaptive streaming scenarios. They provide guidelines for QoE optimal adaptation strategies and additionally have revealed some new insights that have to be further analysed. In terms of quality adaptation vs. stalling, quality switches of larger amplitude together with longer stallings or multiple short stalling events would be a meaningful extension of the presented results. Similarly, additional quality switching amplitudes and periods are needed for better understanding the relationship between frequent quality adaptation and resulting QoE.

6. REFERENCES

[1] Alberti, C., Renzi, D., Timmerer, C., Mueller, C., Lederer, S., Battista, S., and Mattavelli, M. Automated QoE Evaluation of Dynamic Adaptive Streaming over HTTP. In *Proceedings of QoMEX 2013* (Klagenfurt, Austria, 2013).

[2] Arsan, T. Review of Bandwidth Estimation Tools and Application to Bandwidth Adaptive Video Streaming. In *Proceedings of HONET 2012* (Istanbul, Turkey, 2012).

[3] Cisco. Cisco Visual Networking Index: Forecast and Methodology, 2012-2017. Tech. rep., Cisco, 2013.

[4] Egger, S., Hossfeld, T., Schatz, R., and Fiedler, M. Tutorial: Waiting Times in Quality of Experience for Web based Services. In *IEEE QoMEX 2012, Yara Valley, Australia* (2012).

[5] et al., N. C. User perception of adapting video quality. *In: International Journal of Human-Computer Studies 64*, 8 (2006), 637–647.

[6] Gardlo, B., Egger, S., and Seufert, M. Crowdsourcing 2.0: Enhancing Execution Speed and Reliability of Web-based QoE Testing. In *Proc. IEEE ICC, Sydney, Australia* (June 2014).

[7] Grafl, M., and Timmerer, C. Representation Switch Smoothing for Adaptive HTTP Streaming. In *Proceedings of the 4th International Workshop on Perceptual Quality of Systems (PQS 2013)* (Vienna, Austria, 2013).

[8] Hossfeld, T. On Training the Crowd for Subjective Quality Studies. *VQEG eLetter 1* (Mar. 2014).

[9] Hossfeld, T., Egger, S., Schatz, R., Fiedler, M., Masuch, K., and Lorentzen, C. Initial Delay Vs. Interruptions: Between The Devil And The Deep Blue Sea. In *Proc. QoMEX 2012, Yarra Valley, Australia* (July 2012).

[10] Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., and Tran-Gia, P. Best practices for qoe crowdtesting: Qoe assessment with crowdsourcing. *Multimedia, IEEE Transactions on 16*, 2 (Feb 2014), 541–558.

[11] International Telecommunication Union. Subjective video quality assessment methods for multimedia applications. *ITU-T Recommendation P.910* (April 2008).

[12] International Telecommunication Union. Methodology for the Subjective Assessment of the Quality of Television Pictures. *ITU-R Recommendation BT.500* (2012).

[13] Lederer, S., Müller, C., and Timmerer, C. Dynamic adaptive streaming over http dataset. In *Proceedings of the 3rd Multimedia Systems Conference* (2012), ACM, pp. 89–94.

[14] Lee, J.-S., De Simone, F., and Ebrahimi, T. Subjective Quality Evaluation via Paired Comparison: Application to Scalable Video Coding. *IEEE Transactions on Multimedia 13*, 5 (2011), 882–893.

[15] Lewcio, B., Belmudez, B., Mehmood, A., Wältermann, M., and Möller, S. Video Quality in Next Generation Mobile Networks – Perception of Time-varying Transmission. In *Proceedings of CQR 2011* (Naples, FL, USA, 2011).

[16] McCarthy, J. D., Sasse, M. A., and Miras, D. Sharp or Smooth?: Comparing the Effects of Quantization vs. Frame Rate for Streamed Video. In *Proceedings of CHI 2004Subjective Preference of Spatio* (Vienna, Austria, 2004).

[17] Moorthy, A. K., Choi, L. K., Bovik, A. C., and de Veciana, G. Video quality assessment on mobile devices: Subjective, behavioral and objective studies. *Selected Topics in Signal Processing, IEEE Journal of 6*, 6 (2012), 652–671.

[18] Ni, P., Eg, R., Eichhorn, A., Griwodz, C., and Halvorsen, P. Flicker effects in adaptive video streaming to handheld devices. In *Proceedings of the 19th ACM international conference on Multimedia* (2011), ACM, pp. 463–472.

[19] Oyman, O., and Singh, S. Quality of Experience for HTTP Adaptive Streaming Services. *IEEE Communications Magazine 50*, 4 (2012), 20–27.

[20] Roettgers, J. Don't Touch That Dial: How YouTube is Bringing Adaptive Streaming to Mobile, TVs, 2013.

[21] Sackl, A., Zwickl, P., and Reichl, P. The Trouble with Choice: An Empirical Study to Investigate the Influence of Charging Strategies and Content Selection on QoE. In *Proceedings of the 9th (CNSM)* (Zurich, Switzerland, 2013).

[22] Wang, D., Speranza, F., Vincent, A., Martin, T., and Blanchfield, P. Toward Optimal Rate Control: A Study of the Impact of Spatial Resolution, Frame Rate, and Quantization on Subjective Video Quality and Bit Rate. In *Proceedings of the SPIE VCIP 2003* (Lugano, Switzerland, 2003).

[23] Wang, Y., Chang, S.-F., and Loui, A. C. Subjective Preference of Spatio-temporal Rate in Video Adaptation Using Multi-dimensional Scalable Coding. In *Proceedings of the IEEE ICME 2004* (Taipei, Taiwan, 2004).

[24] Yao, J., Kanhere, S. S., Hossain, I., and Hassan, M. Empirical Evaluation of HTTP Adaptive Streaming Under Vehicular Mobility. In *Proceedings of the 10th International IFIP TC 6 Networking Conference: Networking 2011* (Valencia, Spain, 2011).

[25] Zink, M., Schmitt, J., and Steinmetz, R. Layer-encoded Video in Scalable Adaptive Streaming. *IEEE Transactions on Multimedia 7*, 1 (2005), 75–84.