QUALITY OF EXPERIENCE AND HTTP ADAPTIVE STREAMING: A REVIEW OF SUBJECTIVE STUDIES

M.-N Garcia[1], F. De Simone[2], S. Tavakoli[3], N. Staelens[4], S. Egger[5], K. Brunnström[6], A. Raake[1]

[1] Technische Universität Berlin, Germany, [2] Télécom ParisTech, France, [3] Universidad Politécnica de Madrid, Spain, [4] Ghent University/iMinds, Belgium, [5] Austrian Institute of Technology/ Innovation Systems Department/Technology Experience, Austria, [6] Acreo AB/NetLab, Sweden

ABSTRACT

HTTP adaptive streaming technology has become widely spread in multimedia services because of its ability to provide adaptation to characteristics of various viewing devices and dynamic network conditions. There are various studies targeting the optimization of adaptation strategy. However, in order to provide an optimal viewing experience to the end-user, it is crucial to get knowledge about the Quality of Experience (QoE) of different adaptation schemes. This paper overviews the state of the art concerning subjective evaluation of adaptive streaming QoE and highlights the challenges and open research questions related to QoE assessment.

Index Terms— HTTP adaptive streaming, Quality of Experience, Subjective Quality Assessment

1. INTRODUCTION

Video delivery accounts for the major share of nowadays Internet traffic. A large portion of this traffic is delivered through HTTP server-based streaming services such as Youtube, using TCP as underlying transport protocol. In contrast to more traditional video delivery methods over UDP, where packets can be lost and audio and video distortions are introduced, TCP's packet retransmission property prevents these degradations. However, the delivery channel throughput may vary strongly: When the available bandwidth falls below the video bit rate, the client buffer depletes and the playback is interrupted, resulting in a video *stalling* (i.e. frame freeze). To avoid the negative impact of stalling events on users' Quality of Experience (QoE), streamed videos have to be either encoded at very low bitrates or adapt dynamically to the available bandwidth.

Recently, numerous companies have taken up adaptive streaming approaches, typically referred to as HTTP Adaptive Streaming (HAS). In HAS, the content is stored at the server side in a set of different versions (i.e. representations) called adaptation set. Different representations correspond to different video and audio bitrates. They may differ in terms of encoding quantization settings, spatial resolution, temporal resolution, and audio bitrates. Each representation consists of independently decodable units of fixed duration, termed chunks. The client can switch from one representation to another in the adaptation set, every time it requests a chunk. Since different representations in the adaptation set are assumed to correspond to different levels of visual and auditive quality, we refer to such switches as *quality switches*. More specifically, we use the term encoding switch (ESW), spatial switch (SSW), temporal switch (TSW), and audio switch (ASW), when the switch occurs between two representations that differ in terms of encoding quantization settings, spatial resolution, frame rate, and audio bitrates, respectively. Depending on the client adaptation logic, quality switches allow to adapt the requested bitrate to the available bandwidth. When the streaming starts, the client requests the chunk at an initial bitrate. After an *initial startup delay* during which the playout buffer is filled, the client starts displaying the video, while further chunks are requested to maintain the buffer level. In case the downlink throughput decreases, the client buffer depletes. To prevent stalling, the client requests lower bitrate chunks from the server. Hence, the user may perceive a quality switch from a higher to a lower quality. In turn, when the throughput increases, the client requests higher bitrate chunks, which may again result in a perceivable change in video quality.

In order to develop adaptation logics optimizing user's QoE, it is crucial to understand the perceptual impact of adaptation-related impairments. In fact the adaptation events within HAS represent a novel type of perceptual quality degradation. In traditional video QoE research, audiovisual degradations temporally varying over longer sequences have not received much attention. As a result, the standardized quality assessment methodologies for subjective testing [1, 2] mostly fall short in accounting for these impairments, and in recommending the choice of appropriate contents, presentation modes, and quality scales.

For the first time, in this paper, open questions related to subjective quality assessment of HAS are systematically specified, analyzing why these are still open, to motivate how they may be addressed in future research. To this aim, Sect. 2 reviews related work addressing subjective quality evaluation

of HAS impairments. Based on the summary of current QoE knowledge, Sect. 3 discusses shortcomings of previous studies and assessment methodologies, and identifies novel research questions related to the HAS paradigm not addressed so far. Finally, in Sect. 4, conclusions are drawn that include first guidelines for assessing QoE in HAS systems in future studies.

2. WHAT WE DO KNOW

Sect. 2.1 provides an overview of experiments related to the quality assessment of HAS services. The key aspects of the test methods and conditions used in these experiments are summarized in Table 1 and Sect. 2.2. Drawbacks and limitations are discussed in Sect. 3.

2.1. Subjective QoE and HAS

Quality switches. The first question to be answered when considering quality switches is: *when the bitrate has to be lowered (increased), due to restricted (better) network conditions, is it better to reduce (increase) encoding quality, frame rate, or spatial resolution?* Many studies have tried to address this question, even before the spread of HAS solutions, for example in order to develop adaptive transmission strategies relying on Scalable Video Coding. A review of these works is already available in [17]. As a consequence, in the following, the selection and optimization of the adaptation set is not covered. Instead, the paper focuses on studies in which the adaptation set is fixed.

Given an adaptation set where each representation differs in terms of encoding quantization settings, spatial resolution, temporal resolution, and audio bitrate, two key aspects must be considered when designing a perceptually efficient adaptation strategy: the *amplitude* and the *frequency* of the quality switches.

Considering the amplitude of the switch, most of the studies in the literature [8–12] come to the conclusion that *gradual multiple variations are preferred over abrupt variations*. Neverthless, as highlighed in [8], this conclusion may not be generalizable to scenarios where quality levels exhibit a small degree of separation. Some examples of quality amplitude thresholds to delivering a generally acceptable quality have been identified in [9]: for the set of encoding parameters considered in [9], see Table 1 for details, it was found that the amplitude of the switch should not exceed four quantization steps for ESW, one third of the original frame rate for TSW, and half the original frame size for SSW.

Switching frequency also impacts QoE, as highly frequent quality variations may be annoying to the user [8–13]. Additionally, there is an interaction between frequency and amplitude of switches. For example, for ESW or SSW, low-frequency switching can relieve the annoyance of strong quality variations [9]. Nevertheless, the value of the quality levels may still have an impact on the general trend: more frequent switches can be preferred, if the subject is able to view the highest quality video for at least half the duration of the lower rates [8].

Considering the limitations imposed in terms of amplitude and frequency of the switches, one fundamental question can be raised: *is it better to switch quality level or to stay at a constant lower quality?* Many studies have addressed the "to switch or not to switch" dilemma [8–16], with the result that constant quality is usually preferred to varying quality. In particular: short-term spikes may severely degrade the perceived quality [12]; constant (even lower) quality is preferred to decreasing quality (from higher to lower) [10]; constant or nearly constant quality is preferable to frequently varying quality, even if mean quality is lower [9, 13]. In general, providing a bitrate as high as possible does not necessarily lead to the highest QoE [11]. Nevertheless, some exceptions that contradict the general rules must be considered here as well: more than one study pointed out that if the constant quality is too low, any adaptation to the better is preferred [9, 14].

Some conclusions have been also identified concerning the effect of *temporal trends*: according to [8, 10, 14], "all is well that ends well", i.e. the end quality of the video has a definite impact on the perceived quality. Finally, concerning *content dependency*, it has been found in both [9] and [13] that the effect of SSW and TSW varies depending on the content type, even for the same amplitude: while it is difficult to spot quality oscillations when there are frequent scene changes, these are more noticeable in steady shots and when there are strong edges.

Stalling vs. switches. Already back in 2004, the study in [18] has shown that QoE is influenced by both the *duration* and the *frequency* of stalling events. Particularly, subjects prefer a single stalling of longer duration compared to multiple short freezes (confirmed by the study in [8]) and regular (e.g. one stalling event every 3 s) over irregular video stallings. While isolated stallings up to approximatively 400 ms were found to be acceptable to the average end-users [5, 6], many studies agree on the fact that video stalling in HAS should be avoided at all times to improve user's QoE [19–23].

While it is a common assumption that stalling events are more annoying than quality switches, few studies have actually investigated the trade-off between these two degradation types. For example, in [12] this issue has been considered, but using traces from real networks and not in a systematic way: the general conclusion that video stalling must be avoided at any time has been confirmed. Instead, the results in [8] show that stalling events are not yielding worse quality than quality switches, if a few number of stalling events is compared to quality switches involving low bitrates.

Initial delay vs. stalling and starting bitrate. Stalling can be avoided, to some extent, by employing larger client buffers. However, this affects the initial startup delay of the video. Large-scale experiments, described in [7], have shown that the impact of initial startup delay does not significantly

Table 1. Overview of test conditions considered in the studies reviewed in Sect. 2. Only the studies for which more than two entries could be filled in are reported in the table. Legend: ST = stalling, STSKIP = stalling with frame skipping, ID = initial delay, a = amplitude, d = duration, f = frequency, t = trend, r = resolution, R = representation, V = video, AV = audiovisual, A = audio, lab = laboratory (but not standard conditions), ITU = standard conditions, Ex = experiment, na = not available. A detailed review of these studies can be found in [3].

Ref.	Sources (kind # d r)	Impairments		Method	Viewing conditions
[4]	V, 4, 10s, 288p 25fps	STSKIP (d,f)	d: 80, 160, 280 [ms] , #:1,3,5,8	SAMVIQ	lab, CRT
[5]	V, 4, 10s, 288p 25fps	ST (d)	d: 0.12 , 0.20 , 0.52 , 1.0 , 2.0 , 3.0 [s] , #: 1,2,3,5,8	SS	lab, laptop
[6]	AV, na, 15s, 576i 25fps	ST(d)	d: 280, 320, 360, 480 [ms] , #:1	SS	lab, LCD
[7]	AV, 3, 30s/60s, 360p 25/30fps	ID ST	d: 0, 1, 4, 8, 16, 32 [s] d: 0, 0.5, 1, 2, 4, 8 [s] , #:1	SS	lab, PC/ crowdsourcing
[8]	V, 10, 15s, 720p 30fps	ST(d,f) STSKIP (d,f) ESW (a,f,t)	(d #): (1s 8) (2s 4) (4s 2) (d #): (4s 2) Ad.Set: R1-4 [0.7-6] Mbps, chunk=5s, H.264/SVC a:(R1-R4-R1)(R2-R4-R2) (R3-R4-R3) f:(R1-R4-R1-R4)(R1-R4-R1) t:(R1-R2-R4)(R4-R2-R1) (R1-R3-R4)(R4-R3-R1)	SSCQE and overall rate	lab, smartphone, tablet
[9]	V, 4, 12s, 480x320 30fps	ESW (a,f) SSW (a,f) TSW (a,f) for all SW:	Ad.Set: R1=24QP, R2=[28,32,36,40]QP chunk=[0.2, 0.1, 1, 2, 3, 6]s Ad.Set: R1=480x320r, R2=[240x160, 120x80]r chunk=[0.2, 0.1, 1, 2, 3, 6]s Ad.Set: R1=30fps R2=[15,10,5,3]fps chunk=[1, 2, 3, 6]s H.264/SVC, (a f): (R1-R2-R1-R2) at all chunk sizes	SS, acceptance, y/n stability	lab, smartphone
[10]	V, 3, 10s, na	ESW (a,f,t)	Ad.Set: R1-4 or R1-3=na, chunk=na, SMPEG (a f t): see Fig.8 in [10]	SC	na, PC
[11]	AV, 11, 90s, 720p 30fps	ESW+SSW+ ASW(a,f)	Ad.Set: R1=(720p 3.5Mbps(V) 128kbps(A)), R2=(720p 2.5Mbps(V) 128kbps(A)) , R3= (360x 640 1.5Mbps(V) 128kbps(A)), R4=(360x640 700kbps(V) 128kbps(A)), R5=(180x320 300kbps(V) 160kbps(A)), chunk=4s (a f): see in [11]	SS	na, na
[12]	V, na, 108s, na	ESW (a,f) ESW+ST	Ad.Set: R1-7=[256-2048]kbps, chunk=9s (a f): see Fig. 1 in [12] Ad.Set and (a f): na (real network traces)	SS, quality, definition, fluency, responsivness	lab, na
[13] Ex1	V, 1, 120s, 1080p	ESW (f)	Ad.Set and (a, f): na (real streaming)	na	ITU, HDTV
[13] Ex2	V, 2, 110s, 720p	ESW (f)	Ad.Set: R1-8=[350-3000]kbps, (a,f): na	SS	crowdsourcing
[14]	AV, 7, 14s/40s, 1080p, 24/50fps	ESW (a, f, t)	Ad.Set: R1-4=[600, 1000, 3000, 6000]kbps chunk=[2, 10]s, MPEG-4 (a, f, t): see Table 3 in [14]	SS	ITU, HDTV
[15]	AV, 3, 210s, na	ESW + SSW (a, f)	Ad.Set: R1-6=[5000-100]kbps varying r (na) chunk = 10s, MPEG-2 (a, f): na (real network)	SS, quality, acceptance, delight/annoy. pleasure/arousal expectations	lab (one user/ multiple users), tablet
[16]	V, 4, 5s, 1280x800 30fps	ESW (a, f)	Ad.Set: R1-3=[na]QP, chunk=5s, HEVC (a f) = na (content concatenation)	SS	lab, na

deteriorate quality perception. Different amounts of startup delay are tolerated, depending on the specific type of application [24] and, overall, end-users are willing to tolerate larger startup delays, if this results in less video stalling [6]. The initial delay also depends on the bitrate of the chunks that are downloaded to fill the buffer at the beginning of the streaming. If chunks at high (low) bitrate are downloaded, the initial delay will be long (short), but the starting video quality high (low). Few studies have investigated this aspect: in [12], the authors found that a low startup bitrate followed by slow in-

crease ("ramp-up") of quality clearly degrades the QoE.

Combined effects. At the best of our knowledge, there are no systematic studies in the literature where video streams with initial delay, quality switches and stalling events, occurring all in the same video rendering instance, have been used as test material to collect users' quality feedback.

2.2. Summary of applied test methods and conditions

Studies reported in Table 1 are really diverse in terms of test design. Indeed, a large amount of parameters can be adjusted

in the case of adaptive streaming, for instance the number of quality levels, the duration and frequency of quality switches, the type of quality switches, etc. As a consequence, it is impossible to test all parameters in a single test and this forces the test designer to restrict her/his set of test conditions.

Despite of the diversity of studies, some trends can be observed: the most commonly used type of quality switches is ESW. Most of the tests have been conducted in test laboratories, which allows for a more systematic and controlled assessment, cf. Sect. 3. High Definition (1080p, 720p) display resolutions are increasingly used, which match real usage of adaptive streaming. In terms of testing methodology, Single-Stimulus (SS) methods are widely applied instead of the Single Stimulus Continuous Quality Evaluation (SSCQE) method [1].

The biggest difference between studies lies on the amount and duration of source sequences (SRCs) used in the tests. While all studies agree on the influence of the spatio-temporal complexity of the contents on the perceived quality, only three studies have used more than four SRCs [8, 14, 25]. Few studies have considered test sequences longer than one minute [11–13, 15, 26].

3. WHAT WE DO NOT KNOW

A multitude of research questions on HAS is still open. Many questions have been tackled in previous research without actually resolving them; in Sect. 3.1, these questions are summarized in terms of "lack of valid test data", referring to either incomplete reporting of results or inappropriate conduction of tests. Moreover, there are some completely new research questions that have not been addressed so far, or may result from a lack of appropriate test methods; these are discussed in Sect. 3.2.

3.1. Questions due to lack of valid test data

Lack of results can be due to (i) a limited number of tests conducted to address the question, or (ii) shortcomings of the reported studies, such as missing information in the respective publication, evident limitations in the considered set of test conditions, or methodological shortcomings in terms of how tests have been conducted.

Research questions that are still open in spite of some evidence are: What is the combined effect of stalling events and initial loading or quality switches? Does initial loading yield better perceived quality than a slow quality increase at the beginning of the sequence? These tradeoffs have been studied only in [12], but not in a systematic way. How do different dimensions (i.e. ESW, SSW, TSW, ASW) affect the quality-impact due to switching? In this regard, as mentioned in Sect. 2.2, few studies have considered the variation of multiple quality factors: video bitrate and spatial resolution in [25] and additionally audio in [11]. Since only partial data is

available, it is still unclear whether findings on HAS-specific quality perception depend on the properties of the underlying content.

Due to the lack of comprehensive data, further (systematic) investigation is required to understand how the considered quality levels [8], chunk lengths [14] and directions of the variation [23] impact the users' perception and QoE in terms of the amplitudes and frequencies of switches.

Finally, the variation of quality with different viewing devices [8, 15], network, and service types has not been systematically addressed in the literature.

3.2. New problems: Open questions & lack of methods

HAS-related quality comes along with research questions not previously addressed in the literature. In addition, substantial methodological issues may arise when these *new* questions shall be addressed in respective tests. But also when tackling *known* questions, the usage of inappropriate test methods designed for quite different purposes may lead to invalid data. Examples include the question of how long test sequences should be, whether tests should be conducted in a controlled environment, in the field [6], or via crowdsourcing, and what test instructions should be given to the subjects. Even more fundamental issues are related to the general suitability of methods, and the possible need for completely new approaches.

Never-tackled individual and combined degradations. A first straight-forward effect not addressed in any of the reviewed sources is the combined quality-impact due to initial delay, stalling events and quality switches that all occur in one sequence. Also, for benchmarking different approaches to live video streaming, a comparison between HAS-type and UDP-type time-varying quality will be of interest, for example assessing the quality-difference between freezing with and without skipping for longer sequences. Finally, no research has been reported so far on the implications of audiovisual asynchrony in HAS-type streaming.

Dealing with varying quality and long test sequences. With adaptive streaming, it is probable that the quality varies over periods up to several minutes.

As standardized method, only the Single Stimulus Continuous Quality Evaluation (SSCQE) method, specified in [1] and for instance used in [8, 27], is addressing long test sequences by enabling instantaneous evaluation of quality for the whole sequence. Recency and hysteresis effects of the human behavioral responses while continuously evaluating the time-varying video quality were observed [28]. Since subjects are asked to continuously rate the quality, it is unclear whether they completely focus on the degradations or, since the sequences are long, whether they are immersed into the content and forget to rate the quality. Both aspects complicate the analysis of the time-continuous ratings.

The authors in [26] have developed a new test method where subjects continuously view 5 min videos impaired dur-

ing 10 s every 16 s. During the non-impaired 6 s, subjects are asked to rate the quality of the previous impaired 10 s segment. It has been claimed that this method provides a more realistic HAS environment and feeling for the observer than other standardized methods. However, the duration of impaired video segments used in their experiment are still limited to 10 s being far from some of the durations of actual events.

Apart from the SSCQE method, current assessment standards have been produced for short test sequences of constant quality. It is therefore questionable whether these standards are applicable for test sequences longer than 10 s with varying quality. In particular, with long sequences, the attention of the viewers may vary from one content to another and within a single content, depending on the semantic component of the content, its attractiveness, and its popularity. There are also several issues linked to the test design: it may be problematic to repeat the same SRC, since the interest and therefore attention of the viewers may decrease with subsequent viewings of the same SRC. In addition, the interaction between test conditions and the spatio-temporal (ST) complexity of the contents is higher than with short test sequences. The variability of the ST complexity of the content should therefore be considered when designing the tests. In particular, it may be appropriate to use many SRCs per test case.

In this respect, a new subjective assessment methodology making use of a larger source content pool has been suggested in [29]. In this methodology, subjects watch each source only once. By means of statistically sampling and processing the data, conclusion can be generalized on the overall quality variation level of HAS, instead of its impact on a specific content.

Finally, with the use of long sequences, the number of test sequences per test becomes highly limited. Thus, the trade-off between the number of test conditions, SRCs, and SRC duration is more complex than with short test sequences. Also, presenting long sequence videos without audio may be awkward and boring for viewers. As a consequence, video-only tests may have to be avoided in favor of audiovisual tests.

Which question(s) should be asked to the subjects in order to a) better capture the impact of quality variation and b) better include the effect of stalling (which actually does not degrade the picture quality) in the quality ratings? Also, with long sequences, it may be appropriate to have complementary questions, for instance related to the semantic content or to the likability of the content, in addition to the typical overall quality question. This would allow to better analyze the influence of the source content on perceived quality [29].

4. CONCLUSION

The survey in this paper clearly shows that the research on HAS is still very active. It also highlights that there are many remaining open questions in this domain. Fig. 1 illus-
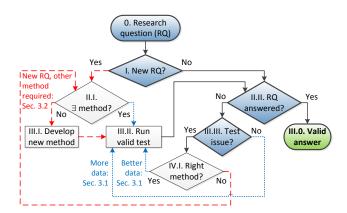

Fig. 1. Flow chart illustrating the types of open questions in HAS assessment. Items highlighted on the right side of the figure represent the research reviewed in this paper. The lighter parts on the left side represent the required steps to answer the remaining open questions.

trates a framework for classifying and systematically addressing these open questions. First step consists of identifying whether the research question is new or not. If not (right side of the figure), it should be checked whether the question has been properly answered or not, due to for instance testing issues and inappropriate test method. In the latter case, new tests should be run. This is the case of the open questions listed in Sect. 3.1.

It is of course not always straightforward to know whether there have been testing issues or if the test method is inappropriate. However, there are some general rules to follow when designing a test which are worth being reminded here. First, the methodology must be chosen carefully along with the test material. This will ensure that the outcome of the study is really answering the research question(s). For the results to be useful and interpretable, it is important to accurately document the test conditions. As a general rule in empirical sciences, the documentation should be in such a way that the experiments can be replicated (*reproducible research*). A detailed description may not be possible in all type of publications (due to, for example, length restrictions), but in that case a more detailed complementary report should be made available. Many articles we surveyed lack this kind of detailed information, making it difficult to generalize the sometimes contradictory conclusions.

If a research question has never been addressed (left side of Fig. 1), as the ones listed in Sect. 3.2, it should be first identified whether existing test methods can be applied to answer this question. If not, a new method should be designed. This is often the case in HAS assessment, and the research is still, to some extent, hindered due to the lack of proper standardized subjective assessment methodologies. In Sect. 3.2 we tried to explain why new methodologies are required. We have also reviewed newly proposed testing methodologies.

Next step consists of building on these new methodologies in order to validate them, extend them, and therefore answer remaining open research questions in HAS assessment.

References

[1] ITU-R, "Methodology for the Subjective Assessment of the Quality of Television Pictures," ITU-R Recommendation BT. 500, Jan 2012.

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

[3] Qualinet VQEG collaboration on adaptive streaming, "http://qualinetdash.pbworks.com," 2014.

[4] Ricardo R Pastrana-Vidal, Jean-Charles Gicquel, Catherine Colomes, and Hocine Cherifi, "Frame Dropping Effects on User Quality Perception," in *Proceedings of the 5th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*, Lisbon, Portugal, 2004.

[5] S. van Kester, T. Xiao, R. E. Kooij, K. Brunnström, and O. K. Ahmed, "Estimating the impact of single and multiple freezes on video quality," *Proc. SPIE 7865, Human Vision and Electronic Imaging XVI*, 2011.

[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] T. Hoßfeld, R. Schatz, Ernst W. Biersack, and L. Plissonneau, "Internet video delivery in YouTube: From Traffic Measurements to Quality of Experience," in *Data Traffic Monitoring and Analysis: From Measurement, Classification, and Anomaly Detection to Quality of Experience*, Lecture Notes in Computer Science, Volume 7754. Springer Berlin Heidelberg, March 2013.

[8] A. K. Moorthy, L.K. Choi, A. C. Bovik, and G. de Veciana, "Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies," *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 6, pp. 652–671, 2012.

[9] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Flicker Effects in Adaptive Video Streaming to Handheld Devices," *Proc. 19th ACM International Conference on Multimedia*, pp. 463–472, 2011.

[10] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz, "Subjective Impression of Variations in Layer Encoded Videos," *Proc. 11th International Conference on Quality of Service*, pp. 137–154, 2003.

[11] R.K.P. Mok, X. Luo, E.W.W. Chan, and R.K.C. Chang, "QDASH: A QoE-aware DASH System," *Proc. 3rd Multimedia Systems Conference*, pp. 11–22, 2012.

[12] L. Yitong, S. Yun, M. Yinian, L. Jing, L. Qi, and Y. Dacheng, "A study on Quality of Experience for adaptive streaming service," *IEEE International Conference on Communications Workshops*, pp. 682–686, 2013.

[13] D. C. Robinson, Y. Jutras, and V. Craciun, "Subjective Video Quality Assessment of HTTP Adaptive Streaming Technologies," *Bell Labs Tech. Journal*, vol. 16, 2012.

[14] S. Tavakoli, K. Brunnström, K. Wang, B. Andrén, M. Shahid, and N. Garcia, "Subjective quality assessment of an adaptive video streaming model," *Proc. SPIE*, vol. 9016, 2014.

[15] B. Villa, K. De Moor, P. Heegaard, and A. Instefjord, "Investigating Quality of Experience in the Context of Adaptive Video Streaming: Findings from an Experimental User Study," in *Norsk Informatikkonferanse NIK 2013*, E. Tøssebro and H. Meling, Eds., pp. 122–133. Akademika forlag, 2013.

[16] A. Rehman and Zhou Wang, "Perceptual experience of time-varying video quality," *Fifth International Workshop on Quality of Multimedia Experience*, pp. 218–223, 2013.

[17] J.-S. Lee, F. De Simone, T. Ebrahimi, N. Ramzan, and E. Izquierdo, "Quality assessment of multidimensional video scalability," *Communications Magazine, IEEE*, vol. 50, no. 4, pp. 38–46, April 2012.

[18] R.R. Pastrana-Vidal, J.C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," *Proc. SPIE 5292, Human Vision and Electronic Imaging IX*, 2004.

[19] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the Quality of Experience of HTTP Video Streaming," in *IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 2011, pp. 485–492.

[20] A. Floris, L. Atzori, G. Ginesu, and D. D. Giusto, "QoE Assessment of Multimedia Video Consumption on Tablet Devices," in *IEEE Globecom Workshop: Quality of Experience of Multimedia Communications*, 2012, pp. 1329–1334.

[21] L. Atzori, A. Floris, G. Ginesu, and D. D. Giusto, "Quality Perception when Streaming Video on Tablet Devices," *Journal of Visual Communication and Image Representation*, 2013.

[22] T. Hoßfeld, D. Strohmeier, A. Raake, and R. Schatz, "Pippi Longstocking Calculus for Temporal Stimuli Pattern on YouTube QoE," *Proc. 5th Workshop on Mobile Video*, pp. 37–42, 2013.

[23] F. Dobrian, V. Sekar, A. Awan, I. Stoica, D. Joseph, A. Ganjam, J. Zhan, and H. Zhang, "Understanding the impact of video quality on user engagement," *Proc. ACM SIGCOMM*, pp. 362–373, 2011.

[24] 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," *Fourth International Workshop on Quality of Multimedia Experience*, pp. 1–6, 2012.

[25] R.K.P. Mok, E.W.W. Chan, X. Luo, and R.. K.C. Chang, "Inferring the QoE of HTTP Video Streaming from User-viewing Activities," *Proc. 1st ACM SIGCOMM Workshop on Measurements Up the Stack*, pp. 31–36, 2011.

[26] S. Tavakoli, J. Gutierrez, and N. Garcia, "Subjective Quality Study of Adaptive Streaming of Monoscopic and Stereoscopic Video," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 4, pp. 684–692, April 2014.

[27] N. Cranley, P. Perry, and L. Murphy, "User perception of adapting video quality," *Int. J. Hum.-Comput. Stud.*, vol. 64, no. 8, pp. 637–647, Aug. 2006.

[28] C. Chen, L.K. Choi, G. de Veciana, C. Caramanis, R.W. Heath, and Al.C. Bovik, "A dynamic system model of time-varying subjective quality of video streams over HTTP," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3602–3606, 2013.

[29] M.H. Pinson, M. Sullivan, and A. Catellier, "Immersive audiovisual subjective testing," *Proc. VPQM*, 2014.