OPEN ACCESS
OPUS AUGSBURG

Universität Augsburg
Universitätsbibliothek

Studying the impact of HAS QoE factors on the standardized QoE model P.1203

Michael Seufert, Nikolas Wehner, Pedro Casas

Angaben zur Veröffentlichung / Publication details:

Seufert, Michael, Nikolas Wehner, and Pedro Casas. 2018. "Studying the impact of HAS QoE factors on the standardized QoE model P.1203." In *IEEE 38th International Conference on Distributed Computing Systems (ICDCS), 2-6 July 2018, Vienna, Austria*, edited by Peter Pietzuch, 1636–41. Piscataway, NJ: IEEE. https://doi.org/10.1109/icdcs.2018.00185.

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/

Studying the Impact of HAS QoE Factors on the Standardized QoE Model P.1203

Michael Seufert, Nikolas Wehner, Pedro Casas
AIT Austrian Institute of Technology GmbH
Vienna, Austria
{michael.seufert.fl | nikolas.wehner.fl | pedro.casas}@ait.ac.at

***Abstract*—P.1203 is a recent standardized model for assessing the Quality of Experience (QoE) of HTTP Adaptive Video Streaming (HAS). However, its complex definition does not allow for a straightforward identification of the underlying assumptions. To overcome this issue, this work investigates the impact of the well-known QoE factors of HAS, namely, initial delay, stalling, and adaptation, on the output QoE score of the model. Therefore, parameter studies are conducted using a reference implementation of P.1203, and the model response to variations of the input QoE factors are compared to results of previous QoE studies in order to get a deeper understanding of the standardized model and its inherent weighting of the QoE factors of HAS.**

I. Introduction

Nowadays, video streaming is the most dominant application in the Internet. According to [1], global IP video traffic had a share of 73% of the Internet traffic in 2016. This share will presumably account for up to 82% in 2021. The popularity of video web portals like YouTube, Netflix, or Amazon Video constantly grows. The number of Netflix subscribers has increased from around 60 million subscribers in 2015 [2] to around 117 million subscribers in March 2018 [3]. In the meantime, the Netflix users watch around one billion hours of video content per week [3]. To avoid customer churn, it is the goal of streaming service providers and Internet Service Providers (ISP) to satisfy their customers in terms of service quality and experience.

The perceived streaming experience by the end users can be quantified with the concept of the Quality of Experience (QoE). In terms of video streaming, the QoE states to what extent users are annoyed or delighted with the provided streaming [4]. The currently prevailing streaming technology – HTTP Adaptive Streaming (HAS) – allows to adapt the video bit rate to the network conditions. The goal is to ensure a smooth streaming when end users face throughput fluctuations, e.g., in mobile networks. Thereby, initial delay and stalling, i.e., playback interruptions, can be reduced, which are severe QoE degradations of video streaming. However, due to the bit rate adaptation, the visual quality of the video might vary, which introduces an additional QoE factor [5].

Typically, subjective user studies have to be conducted to assess the QoE of an end user for a given HAS session. Since subjective user studies are expensive, time-consuming, and inconvenient, QoE models offer a simple and practical way to predict the QoE. Usually, the models are derived by analyzing and extrapolating the data of already performed subjective studies. Most often the participants rated their subjective QoE on an ordinal Absolute Category Rating (ACR) [6] scale, which ranges from 1 (bad) to 5 (excellent). Further, all ratings are aggregated as Mean Opinion Score (MOS), which is the mean of the subjective ratings. QoE models take monitored application-layer and/or network-layer streaming parameter as input and commonly return a MOS value that should correspond to the mean user experience. Next to a multitude of proposed QoE models for HAS, recently, a QoE model was standardized by ITU-T as P.1203 [7]. However, its complex definition does not allow for a straightforward identification of the underlying assumptions, e.g., on the importance or contribution of QoE factors to the overall QoE score.

To overcome this issue, the impact of the well-known QoE factors of HAS, namely, initial delay, stalling, and adaptation of the visual quality, on the standardized QoE model P.1203 is studied in this work. Therefore, parameter studies are conducted using a reference implementation of P.1203, and the resulting QoE scores, which are output by the model, are investigated. The model response to variations of the input QoE factors are compared to results of previous QoE studies in order to get an understanding of the standardized model and its inherent weighting of QoE factors.

This paper is structured as follows. Section II outlines related works on the QoE of HAS, as well as QoE models and QoE monitoring approaches. Section III presents the standardized QoE model P.1203 and Section IV outlines the study concept for evaluating the impact of QoE factors on the resulting QoE score. Section V shows the parameter study and discusses the findings with respect to results of previous QoE studies. Finally, Section VI concludes.

II. Related Work

Quality of Experience (QoE) is a concept to quantify the subjectively perceived quality by an end user when utilizing an application or a service. [4] defines QoE as "the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the users personality and current state".

The various QoE influence factors can be separated into four distinct categories, namely context, user, system, and content level. The context level attends to the social and cultural background of the user, as well as to the environment in which the application or service is used, and the purpose for which it is used. The psychological factors like memories of earlier application usage or user expectations are considered on the user level. The system level covers the technical influence factors, e.g., the applied devices and its screen characteristics, the transmission network, and the utilized implementation of the application itself. Finally, in terms of video streaming, the content level specifies the employed video codec, format, resolution, type of video, etc.

The QoE of video streaming is a widely investigated research topic, in particular, the specific QoE influence factors and the generation of QoE models. The main goal of application providers and Internet Service Providers (ISPs) is to satisfy the user with an optimal level of QoE when delivering video streams, and hence, minimize the risk of customer churn. To ensure a high level of user experience, network operators need to know the most important influence factors and their impact on the QoE. [8] reviews methodologies to evaluate the QoE of video streaming. Similarly, [9] gives a tutorial on popular QoE assessment approaches for video streaming. The state of the art on the QoE of HAS is summarized in [5]. [10] surveyed metrics, tools, and measurement methodologies to predict the QoE of video streaming, and [11], [12] not only focused on QoE factors and assessment, but also took QoE management into account. A detailed summary on QoE models and monitoring approaches can also be found in [13]. Most related works agree that stalling, initial delay, and adaptation of the visual quality are considered to be the most important QoE factors. These factors will be discussed in detail in Section V, when their impact on the standardized QoE model P.1203 is investigated.

III. P.1203

The International Telecommunication Union (ITU) released recommendation P.1203 for the standardized quality assessment of HAS in November 2016 [7]. P.1203 is a set of documents that describes models and tools for the quality assessment of progressive video download and adaptive audiovisual streaming services for TCP-type video streaming.

The standard is already in used. [14] collected the decrypted stream, the encrypted stream, and meta information for the videos. Then, they utilized P.1203 to evaluate the played videos by using the decrypted streams and afterwards applied machine learning in order to build a model for the analysis of encrypted streams. In [15], the authors evaluate the performance of P.1203 by applying it on collected data from YouTube and other OTT services. The study reveals that the collected subjective MOS aligns reasonably well with the predicted MOS by P.1203.

The output of the model is a predicted MOS on the ACR scale. The output is shaped by information on the audio and video media encodings as well as by application-layer parameters like the number of stallings, the length of the stalling events, the interarrival time of the stalling events, the initial delay, and the number of quality switches.

Depending on the amount of information obtained before streaming and the complexity of the streaming media, there are four different modes of operation (mode 0-3). The amount of available information increases with the mode number, i.e., in mode 3 the complete bitstream-based media streaming information is available, while mode 0 should be applied when there is only the information available that could be retrieved throughout the streaming due to meta-information.

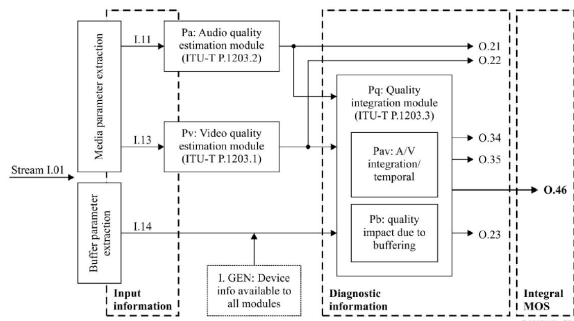


Fig. 1: Recommendation ITU-T P.1203 [7].

Figure 1 depicts the architecture of P.1203 [7]. Overall, there are three modules, namely the audio quality estimation module (Pa), the video quality estimation module (Pv), and the quality integration module (Pq). The module Pq depends on the modules Pa and Pv and returns the overall integral MOS (O.46) at the end. As a result of the modules, the six outputs O.21, O.22, O.23, O.34, O.35, and O.46 are generated.

The module Pa outputs not a single MOS score, but a vector of MOS scores. For the sampling interval of one second, the audio quality is estimated based on the current audio information. For Pa, the used mode inflicts no difference upon the prediction of the MOS scores. Eq. 1 shows the formula for the calculation of a single MOS score for a one second interval. For the calculation of the MOS scores, the current audio bit rate in kbps and the employed audio codec are utilized.

$$O.21 = MOSfromR(100 - (a1A \cdot e^{a2A \cdot Bitrate} + a3A)) \quad (1)$$

The coefficients *a1A*, *a2A*, and *a3A* differ based on the selected audio codec. Finally, the function *MOSfromR* scales the obtained quality value to get a MOS score in the range of 1.05 and 4.9.

Just like the module Pa, Pv also returns a vector of MOS scores. The structure of the module Pv is displayed in Figure 2 [7]. The module consists of the four submodules quantization module, temporal module, upscaling module, and integration module. The integration module combines the returned values from the three other modules to generate the MOS for a one second interval. Eq. 2 shows the formula for the computation of *D* of the integration module, where *Dq* is the result of the quantization module, *Dt* the result of the temporal module, and *Du* the result of the upscaling module.

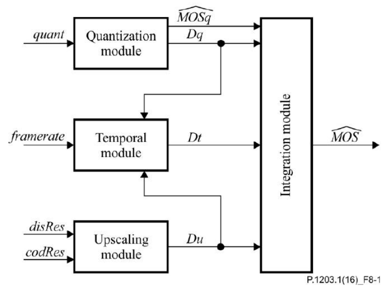


Fig. 2: Structure of the video quality estimation module Pv [7].

The obtained *D* value is then utilized to calculate the MOS for the interval. If *Du* and *Dt* are 0, the MOS equals the returned MOS of the quantization module. Otherwise, the function *MOSfromR* transforms the *D* value into the ACR scale.

$$D = max(min(Dq + Du + Dt, 100), 0) \tag{2}$$

$$O.22 = \begin{cases} MOSq, \text{ if } \; Du = 0 \text{ and } Dt = 0, \\ MOSfromR(100 - D), \text{ otherwise.} \end{cases} \tag{3}$$

Finally, the quality integration module computes the impact on the QoE due to stalling and the impact on the audiovisual quality on the basis of the audio quality and video quality estimation module. Further, the module uses machine learning to predict a MOS for the audiovisual quality estimation.

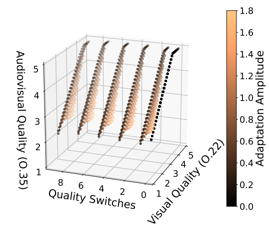
The audiovisual quality per output sampling interval O.34 is depicted in Eq. 4, where av_1, av_2, av_2, and av_4 are fixed coefficients, O.21 is the output vector of the audio quality estimation module, and O.22 is the output of the video quality estimation module.

$$\begin{aligned} O.34 =& max(min(av_1 + av_2 \cdot O.21(t) + \\ & av_3 \cdot O.22(t) + av_4 \cdot O.21(t) \cdot O.22(t), 5), 1) \end{aligned} \tag{4}$$

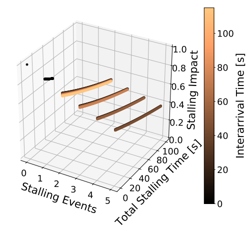
The final audiovisual coding quality is stated by O.35. The computation of O.35 is displayed in Eq. 5. The variable *negBias* is used as a negative bias. The variables *oscComp* and *adaptComp* consider the amplitude of video quality switching, the difference between the maximal video quality and the minimal video quality, the quality change rate, and the longest time on a quality layer. These variables are estimated by using the output of O.22. The $O.35_{baseline}$ is a weighted average of the temporary audiovisual quality O.34, weighted by the current playback time and the video duration.

$$O.35 = O.35_{baseline} - negBias - oscComp - adpatComp \tag{5}$$

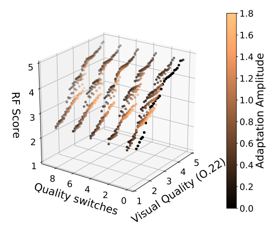
The stalling indicator is expressed by the variable *SI*. Its computation is shown in Eq. 6. The variable is required for the computation of O.23 and O.46. The indicator is based on the number of stalling events *numStalls*, the total stalling length *totalBuffLen*, and the average interarrival time of the stalling events *avgBuffInterval*. If there are less than two stalling events in a video streaming session, *avgBuffInterval* becomes 0. Again, the parameters s_1, s_2, and s_3 are coefficients provided by the recommendation. The variable ranges between



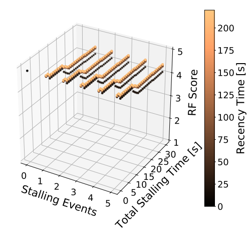
(a) Audiovisual quality (O.35).



(b) Stalling impact (SI).



(c) Random Forest score based on audiovisual quality (RF).



(d) Random Forest score based on stalling impact (RF).

Fig. 3: Composition of intermediate quality scores.

0 and 1, where 1 corresponds to no QoE degradation caused by stalling and 0 corresponds to the opposite.

$$SI = e^{-numStalls/s_1} \cdot e^{-\frac{totalBuffLen}{T \cdot s_2}} \cdot e^{-\frac{avgBuffInterval}{T \cdot s_3}} \tag{6}$$

The algorithm used for the machine learning of the audio-visual quality is a Random Forest classifier. The built model that comes with the recommendation documents consists of 20 trees. Each tree has a maximal depth of 6 and the model consists of 14 features. The features are extracted by analyzing the vectors from O.21 and O.22, as well as using the stalling information. The generated features are then passed on to the Random Forest. Each decision tree returns a MOS score as output. The overall MOS is obtained by taking the average of all MOS scores.

Figure 3 illustrates the composition of the intermediate quality scores, which are used to obatin the final MOS score. The final MOS score for the video streaming session is computed in Eq. 7. The final audiovisual quality O.35 and the stalling indicator are accounted and the result is weighted with 75%, while the prediction of the Random Forest machine learning is weighted with 25%.

$$O.46 = 0.75 \cdot (1 + (O.35 - 1) \cdot SI) + 0.25 \cdot RF_{prediction} \tag{7}$$

Figure 4 visualizes the impact of each intermediate score.

IV. Study Concept

As the definition of the metric is quite complex, the remainder of the paper aims at finding basic relationships between the QoE factors and the output MOS score. Each evaluation focuses solely on a single investigated QoE factor and the corresponding predicted MOS.

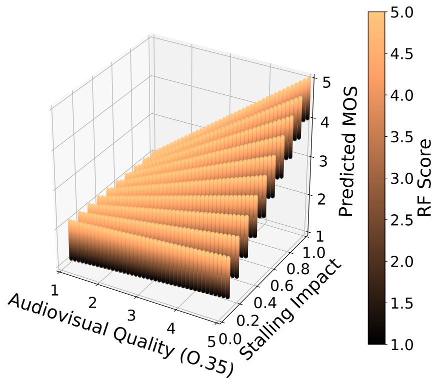


Fig. 4: Predicted MOS by P.1203.

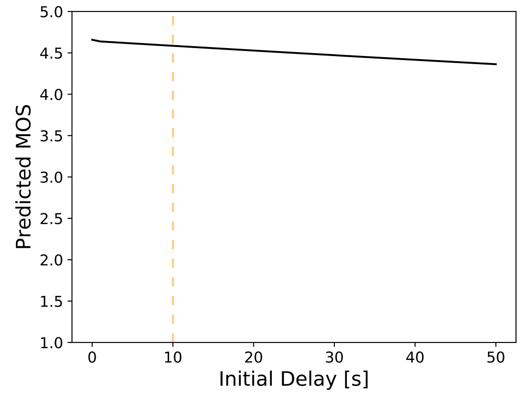


Fig. 5: Impact of initial delay.

The study is performed with a reference implementation of P.1203 provided on Github [16], [17]. No real videos are used for the evaluation, i.e., it is a parameter study. Further, the study does not consider the audio quality O.21, and consequently uses the best possible audio quality (5) throughout the study. The considered video duration is 240 s, which is in the application range of P.1203 (60 - 300 s) [7], and, from both playback options of the standard, a large screen is assumed for video playback, which means that no adjustment for handhelds was added.

To only investigate a single QoE factor, other modules, which are not affected by that QoE factor, are assigned fixed values for simplicity, corresponding to the maximum quality output of the module. For example, for the evaluation of the initial delay and stalling, the best visual quality (5) is assumed for O.22 in order to neglect any QoE degradation of the visual quality. Similarly, there is neither initial delay nor stalling present when evaluating the visual quality O.22 or the audiovisual quality O.35, respectively. For the investigation of the visual quality, constant bit rates (mode 0) and QP values (mode 3) are assumed throughout the video duration, which are then used to obtain the visual quality O.22. Moreover, upscaling due to differences in encoding and screen size and temporal degradations due to lower frame rate than 24 fps are not considered.

Of course, the simple relationships, which are presented next in Section V do not reflect the entire complexity of P.1203. As can be seen in Figure 3, there are always several variables involved in shaping a single output. Thus, it has to be noted that there are joint effects of QoE factors, which are not investigated in this evaluation. Additionally, some modules have limited their output range (e.g., Pa and Pv output only in the range of 1.05 to 4.9), so this evaluation might show a too optimistic MOS prediction. Still, the evaluation is well suited to identify qualitative relationships between the QoE factors of HAS and the predicted MOS.

V. Evaluation

This section presents the impact of the QoE factors of HAS on the output of the P.1203 QoE model. The considered QoE factors include initial delay, stalling, and visual quality. Only one factor was changed at a time, while the other factors were kept at maximum quality. Thus, the impact of a single QoE factor on the predicted MOS can be investigated qualitatively.

A. Initial Delay

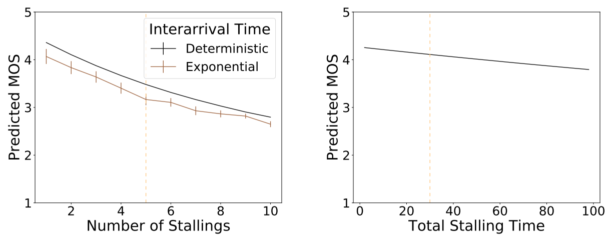
Initial delay resembles the period of time that passes between the video start triggered by the user and the actual video start. The decoding of the video data and the playback of the video is not started until the buffer has been filled with a sufficient amount of video data. Thus, the initial delay is always present when using video streaming.

Figure 5 shows the impact of initial delay on the predicted MOS. P.1203 supports video clips with initial delay up to 10 s [7]. It can be seen that initial delay has only a very small linear influence on the resulting QoE score. This is in accordance with current QoE results, as some studies have reported a minor QoE impact. [18] found that the length of the initial delay influenced the QoE. However, they revealed that initial delays of a length up to 16 seconds have an insignificant impact on the QoE. It was confirmed by [19] that initial delays were also considered less important for mobile video users and less critical for having a high QoE. In [20], the authors showed that the impact of the initial delay on the QoE decreases for high resolutions.

B. Stalling

Stalling describes the video playback interruption due to buffer underrun. If the throughput can not afford the video bit rate, the buffer drains. When the buffer level is beneath the threshold, where insufficient data for playback is available, the video playback has to be stopped until the buffer is refilled with a certain amount of video data. Stalling is considered the main factor of QoE degradation [5], because it is processed differently by the human sensory system [21].

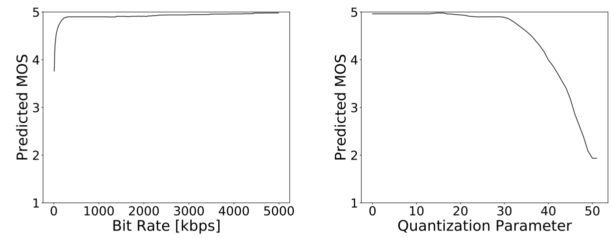
Figure 6a shows the impact of the number of stalling events on the predicted QoE. The number of stalling events is analyzed by using a fixed total stalling time of 15 seconds and either a deterministic (i.e., regular) or an exponential (i.e., memoryless) interarrival time of stalling events. The regular



(a) Number of stalling events.

(b) Total stalling time.

Fig. 6: Impact of stalling.



(a) Bit rate (Mode 0)

(b) Quantization parameter (Mode 3)

Fig. 7: Impact of visual quality.

stalling events occur every T/(N+1) seconds, where T is the duration of the video in seconds and N the number of stalling events. The exponential interarrival times are random with mean T/(N+1) seconds, such that exactly N arrivals lie within the specified video duration. The supported range of P.1203 is 5 stalling events with a maximum event length of 15 s and a maximum total stalling time of 30 s. Moreover, no stalling is allowed to occur in the first 5 s of the video playback [7].

The black plot depicts the deterministic stalling and the brown plot shows the mean QoE and 95% confidence intervals for different exponential stalling patterns. The shape of both plots is rather linear, which does not correspond to the exponential QoE degradation that was reported by [22]. However, it includes the results of [23], which discovered that irregular stalling patterns degrade the QoE more severely than periodic stalling patterns. Still, the difference between both curves is small. Recently, [24] found that the degradation of stalling was worse when the presentation quality was higher. This effect could also be found in the P.1203 model.

Figure 6b depicts the impact of the total stalling time. The evaluation considers only two stalling events, which are regularly distributed and have the same length. The stalling events occurs exactly at one third and two thirds the video duration, i.e., at 80 and 160 s. The considered total stalling times range from 2 to 100 seconds. Note that P.1203 officially supports only a total stalling time of up to 30 s [7]. It can be seen that P.1203 shows a linear relationship between the total stalling time and the predicted QoE score. This is generally in line with previous results of [22], which found that the QoE level decreases with an increasing stalling duration. However, the overall impact in P.1203 is small, in particular, it is smaller than the effect of the number of stalling events.

C. Visual Quality

With HAS, the videos are commonly encoded with several different bit rates and resolutions. When adapting the bit rate, the user perceives a change in the visual quality, which eventually has an impact on the QoE.

Figure 7 shows the impact of the visual quality on the predicted MOS score. Figure 7a investigated the mode 0 of P.1203 when only metadata are available. In this case, the prediction is based on the bit rate of the segments. The considered bit rates range from 10 to 5000 kbps. Typically, the higher the bit rate, the higher the visual quality, which is also reflected in P.1203. P.1203 considers a bit rate of 4500 kbps sufficient to achieve an excellent visual quality. Lower bit rates result in lower predicted MOS. However, it can be seen that the QoE score is only significantly reduced for bit rates below 500 kbps. Interestingly, also with very low bit rates the predicted MOS is still good (≥ 3.5) if the streaming is smooth.

Figure 7b considers mode 3 of P.1203, in which complete bitstream-based media streaming information has to be available. In this case, the QoE prediction is based on the quantization parameter of the video encoding, which ranges from 0 (no compression, high visual quality) to 51 (strong compression, low visual quality). The QoE score decreases fast down to the lowest MOS score of 1 when the quantization parameters increases. The results for both mode 0 and mode 3 correspond to results from related works that the QoE changes according to the visual quality [25], [26].

Figure 8 investigates the impact of the number of quality switches, i.e., the adaptation frequency, for different intensities of adaptation, i.e., adaptation amplitudes. The evaluation is based on specified O.22 patterns with the same average visual quality. Note that P.1203 computes the number of quality switches by analyzing the difference between consecutive O.22 samples. If the absolute value of the difference is greater than 0.2, a quality switch is noted. The evaluated patterns are generated by specifying a baseline visual quality level, which is 3 on the O.22 scale. This baseline pattern results in a predicted MOS of 4 because no stalling is added. A high quality level and a low quality level are generated by adding or subtracting the adaptation amplitude from the baseline quality level on the O.22 scale, respectively. Four adaptation amplitudes were selected that range from 0.2 (black) to 1.6 (yellow). All patterns contain an odd number of regular quality changes (1 to 9) and change the quality between the high and the low quality level, such that the average quality level on the O.22 scale is equal to the baseline pattern. The plot shows that the number of quality changes only has a minor impact on the predicted QoE. In contrast, the adaptation amplitude is more important. These findings are in line with [27], [28].

VI. Conclusion

This work investigated the impact of variations of different QoE factors of HAS on the QoE score of the standardized QoE

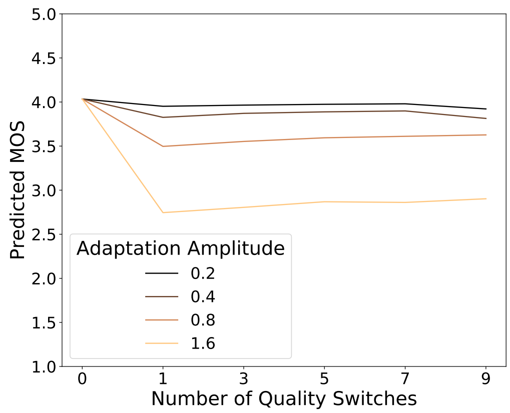


Fig. 8: Impact of quality switches.

model P.1203. Although the computation of the QoE score is quite complex, it was found that well-known results of previous QoE studies are qualitatively represented in the model. However, the quantitative effect of each QoE factor might be implemented differently compared to other QoE studies. More detailed investigations are planned in future work, especially to better understand the combined effects of QoE factors on the standardized model, and the underlying assumptions with respect to the inherent weighting and importance of the different QoE factors of HAS. This helps to further improve the usability of the QoE model P.1203 and to properly interpret the predicted QoE scores.

References

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2016-2021," Cisco, Tech. Rep., 2017.

[2] ReelnReel, "20 Netflix Streaming Statistics That Will Blow Your Mind," 2017. [Online]. Available: https://www.reelnreel.com/20-netflix-streaming-statistics-that-will-blow-your-mind

[3] DMR, "110 Amazing Netflix Statistics and Facts (March 2018)," 2018. [Online]. Available: https://expandedramblings.com/index.php/netflix_statistics-facts

[4] P. Le Callet, S. Möller, and A. Perkis (eds), "Qualinet White Paper on Definitions of Quality of Experience," European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Tech. Rep., 2013, version 1.2.

[5] 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, no. 1, pp. 469–492, 2015.

[6] International Telecommunication Union, "ITU-T Recommendation P.910: Subjective Video Quality Assessment Methods for Multimedia Applications," 2008.

[7] ——, "ITU-T Recommendation P.1203: Parametric Bitstream-based Quality Assessment of Progressive Download and Adaptive Audiovisual Streaming Services over Reliable Transport," 2016. [Online]. Available: https://www.itu.int/rec/T-REC-P.1203/en

[8] O. B. Maia, H. C. Yehia, and L. de Errico, "A Concise Review of the Quality of Experience Assessment for Video Streaming," *Computer Communications*, vol. 57, pp. 1–12, 2015.

[9] Y. Chen, K. Wu, and Q. Zhang, "From QoS to QoE: a Tutorial on Video Quality Assessment," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 1126–1165, 2015.

[10] P. Juluri, V. Tamarapalli, and D. Medhi, "Measurement of Quality of Experience of Video-on-Demand Services: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 401–418, 2016.

[11] M. G. Martini, C. T. Hewage, M. M. Nasrall, and O. Ognenoski, "QoE Control, Monitoring, and Management Strategies," in *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, C. W. Chen, P. Chatzimisios, T. Dagiuklas, and L. Atzori, Eds. John Wiley & Sons, 2015, pp. 149–168.

[12] T. Zhao, Q. Liu, and C. W. Chen, "QoE in Video Transmission: A User Experience-Driven Strategy," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 285–302, 2017.

[13] M. T. Seufert, "Quality of Experience and Access Network Traffic Management of HTTP Adaptive Video Streaming," Doctoral Thesis, University of Würzburg, 2017. [Online]. Available: https://opus.bibliothek.uni-wuerzburg.de/files/15413/Seufert_Michael_Thomas_HTTP.pdf

[14] S. Göring, A. Raake, and B. Feiten, "A Framework for QoE Analysis of Encrypted Video Streams," in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[15] S. Satti, C. Schmidmer, M. Obermann, R. Bitto, L. Agarwal, and M. Keyhl, "P.1203 Evaluation of Real OTT Video Services," in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[16] A. Raake, M. N. Garcia, W. Robitza, P. List, S. Göring, and B. Feiten, "A Bitstream-based, Scalable Video-quality Model for HTTP Adaptive Streaming: ITU-T P.1203.1," in *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, 2017.

[17] S. Göring and W. Robitza, "ITU-T Rec. P.1203 Standalone Implementation," 2018. [Online]. Available: https://github.com/itu-p1203/itu-p1203

[18] 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 *Proceedings of the 4th Intl. Workshop on Quality of Multimedia Experience (QoMEX)*, Yarra Valley, Australia, 2012.

[19] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the Influence of Rebuffering Interruptions on the User's Quality of Experience During Mobile Video Watching," *IEEE Transactions on Broadcasting*, vol. 59, no. 1, pp. 47–61, 2013.

[20] M.-N. Garcia, D. Dytko, and A. Raake, "Quality Impact Due to Initial Loading, Stalling, and Video Bitrate in Progressive Download Video Services," in *Proceedings of the 6th International Workshop on Quality of Multimedia Experience (QoMEX)*. Singapore: IEEE, 2014.

[21] A. Raake and S. Egger, "Quality and Quality of Experience," in *Quality of Experience: Advanced Concepts, Applications and Methods*, S. Möller and A. Raake, Eds. Springer, 2014.

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

[23] Q. Huynh-Thu and M. Ghanbari, "Temporal Aspect of Perceived Quality in Mobile Video Broadcasting," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 641–651, 2008.

[24] K. Zeng, H. Yeganeh, and Z. Wang, "Quality-of-experience of Streaming Video: Interactions between Presentation Quality and Playback Stalling," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016.

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

[26] 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 *Proceedings of the 5th International Workshop on Quality of Multimedia Experience (QoMEX)*, Klagenfurt, Austria, 2013.

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

[28] M. Seufert, T. Hoßfeld, and C. Sieber, "Impact of Intermediate Layer on Quality of Experience of HTTP Adaptive Streaming," in *Proceedings of the 11th International Conference on Network and Service Management (CNSM)*, Barcelona, Spain, 2015.