

## Cumulative quality modeling for HTTP adaptive streaming

**Huyen T. T. Tran, Nam Pham Ngoc, Tobias Hoßfeld, Michael Seufert, Truong Cong Thang**

### Angaben zur Veröffentlichung / Publication details:

Tran, Huyen T. T., Nam Pham Ngoc, Tobias Hoßfeld, Michael Seufert, and Truong Cong Thang. 2021. "Cumulative quality modeling for HTTP adaptive streaming." ACM Transactions on Multimedia Computing, Communications, and Applications 17 (1): 22. <https://doi.org/10.1145/3423421>.

### 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/>



# Cumulative Quality Modeling for HTTP Adaptive Streaming

HUYEN T. T. TRAN, The University of Aizu, Japan

NAM PHAM NGOC, VinUniversity, Vietnam

TOBIAS HOßFELD and MICHAEL SEUFERT, University of Würzburg, Germany

TRUONG CONG THANG, The University of Aizu, Japan

---

HTTP Adaptive Streaming has become the de facto choice for multimedia delivery. However, the quality of adaptive video streaming may fluctuate strongly during a session due to throughput fluctuations. So, it is important to evaluate the quality of a streaming session over time. In this article, we propose a model to estimate the cumulative quality for HTTP Adaptive Streaming. In the model, a sliding window of video segments is employed as the basic building block. Through statistical analysis using a subjective dataset, we identify four important components of the cumulative quality model, namely the minimum window quality, the last window quality, the maximum window quality, and the average window quality. Experiment results show that the proposed model achieves high prediction performance and outperforms related quality models. In addition, another advantage of the proposed model is its simplicity and effectiveness for deployment in real-time estimation. Our subjective dataset as well as the source code of the proposed model have been made publicly available at <https://sites.google.com/site/huyenthithanhtran1191/cqmdatabase>.

CCS Concepts: • **Information systems** → **Multimedia streaming**;

Additional Key Words and Phrases: Cumulative quality, quality model, quality of experience, adaptive video streaming

## ACM Reference format:

Huyen T. T. Tran, Nam Pham Ngoc, Tobias Hoßfeld, Michael Seufert, and Truong Cong Thang. 2021. Cumulative Quality Modeling for HTTP Adaptive Streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 17, 1, Article 22 (April 2021), 24 pages.

<https://doi.org/10.1145/3423421>

---

## 1 INTRODUCTION

HTTP Adaptive Streaming (HAS) has become the de facto choice for multimedia delivery [16]. In HAS, a video is encoded into different quality versions. Each version is further divided into a series of segments. Depending on throughput fluctuations, segments of appropriate quality versions will be delivered from the server to the client, which results in quality variations during a session. Therefore, a key challenge in HAS is how to evaluate the quality of a session over time. The evaluation can provide service providers with suggestions to enhance the quality of services [71].

---

Authors' addresses: H. T. T. Tran and T. C. Thang, The University of Aizu, 965-8580 Ikkimachi, Aizuwakamatsu, Japan; emails: {d8192106, thang}@u-aizu.ac.jp; N. P. Ngoc, VinUniversity, Vinhomes Ocean Park, Gia Lam District, Hanoi, Vietnam; email: v.namnp3@vingroup.net; T. Hoßfeld and M. Seufert, University of Würzburg, Chair of Communication Network, Am Hubland, 97074 Würzburg, Germany; emails: {tobias.hossfeld, michael.seufert}@uni-wuerzburg.de. 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 *ACM Trans. Multimedia Comput. Commun. Appl.*, Vol. 17, No. 1, Article 22.

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

<https://doi.org/10.1145/3423421>

Because of the utmost importance of quality assessment, quality modeling for HAS has gained attention from both academic and industry in recent years [7, 36, 68]. Many previous studies have deployed existing quality models to build effective adaptive streaming strategies that aim to provide the highest possible quality to users [5, 22, 64, 69, 70]. In spite of the potential applications, it is currently still in urgent demand of both effective and efficient quality models that could not only accurately reflect the human perceived quality but also be applicable to real-time quality monitoring.

Here, we would like to differentiate three concepts of the quality as follows:

- *Continuous quality* means the instantaneous quality that is continuously perceived at any moment of the session.
- *Overall quality* means the quality of a whole session.
- *Cumulative quality* means the quality cumulated from the beginning up to any moment of the session. Obviously, the concept of overall quality is a special case of cumulative quality.

It should be noted that the concepts of continuous quality and overall quality have been mentioned in Recommendations ITU-R BT.500-13 and ITU-T P.880 [38, 43].

Based on comprehensive surveys of related works on quality modeling for HAS in References [4, 23, 53], it is shown that the continuous and overall quality has been investigated in a large number of previous studies. To the best of our knowledge, however, few existing studies have actually considered the cumulative quality. The work in Reference [45] was the first study on the cumulative quality of a video streaming session, where the authors focused on the impact of quality variations. However, this work employed very short sessions, only 5–15 s.

In this study, our goal is modeling the cumulative quality of HTTP adaptive video streaming. We first carry out a subjective test to measure the cumulative quality of long sessions of 6 minutes. Then, the impacts of quality variations, primacy, and recency are investigated. Based on the obtained results, a cumulative quality model (called CQM) is proposed. In the proposed model, a sliding window of video segments is the basic unit of computation. It should be noted that, in the following, the term “window” means either the conceptual sliding window or a window at a certain location. Experiment results show that the quality of the last window, the average window quality, the minimum window quality, and the maximum window quality are key components of the cumulative quality model. Also, it is found that the proposed model outperforms 10 existing models in both cumulative and overall quality prediction. Moreover, the proposed model is applicable to real-time quality monitoring thanks to its low computation complexity. To the best of our knowledge, the proposed model is the first cumulative quality model for actual streaming sessions.

The remainder of this article is organized as follows. Section 2 discusses the related work and our contributions. Because the proposed model is based on an analysis of subjective results, the subjective test is presented in Section 3. Then, Section 4 presents the proposed cumulative quality model. In Section 5, we evaluate the performance and computation complexity of the proposed model and compare it to ten existing models. Also, some remarks on cumulative quality prediction are presented. Finally, conclusions are drawn in Section 6.

## 2 RELATED WORK AND CONTRIBUTIONS

In this section, we will discuss the works related to three types of quality, namely, (1) continuous quality, (2) overall quality, and (3) cumulative quality. Also, our contributions in this study will be presented at the end of this section.

## 2.1 Continuous Quality

The recommendation ITU-R BT.500-13 describes the Single Stimulus Continuous Quality Evaluation (SSCQE) method for subjective assessment of the continuous quality [38]. In this method, test sessions are displayed in a random order. Each subject, while watching a video, is asked to continuously move a slider along a continuous scale so that its position reflects his/her selection of quality at that instant. All subjects' quality ratings at each instant of each video are averaged to compute a mean opinion score (MOS) of that instant.

The work in Reference [6] is the first study on the continuous quality of a streaming session. Note that, in this article, the authors use the term "time-varying quality" to refer to "continuous quality." To measure the continuous quality, the authors conducted a subjective test similar to the SSCQE method. Then, a continuous quality model is proposed, taking into account the impact of the recency. In particular, a Hammerstein-Wiener model was employed to predict the continuous quality of 5-minute-long sessions. As this work is focused on continuous quality, the model mainly depends on the quality values of the last 15 s.

Reference [31] uses machine learning to predict initial delay, stalling, and video quality from the network traffic in windows of 10 s. The considered features are derived from IP or TCP/UDP headers only. ViCrypt [52] detects quality degradations on encrypted video streaming traffic in real-time within 1 s by using a streamlike analysis approach with two continuous sliding windows and a cumulative window. The features are based on packet-level statistics of the network traffic, and allow to accurately recognize initial delay and stalling [52], as well as video resolution and the average bitrate [65].

Reference [17] presents a continuous quality predictor using an ensemble of Hammerstein-Wiener models, while [2, 15] developed neural-network-based continuous quality models. As discussed in Recommendation ITU-R BT.500-13 [38], the continuous quality values of a session can be pooled to predict the overall quality. However, effective pooling strategies are currently under study [3, 38, 50].

## 2.2 Overall Quality

The overall quality perceived by the end-users can be quantified with the concept of Quality of Experience (QoE). In terms of video streaming, the QoE states to what extent users are annoyed or delighted with the provided streaming [18, 26].

In Reference [20], it was found that the impact of the initial delay of the video stream is not severe, whereas the impact of stalling, i.e., playback interruptions, is significant. To model the impact of the interruptions, previous studies generally used some statistics such as the number of interruptions [30, 55, 67], the average [55], the maximum [55], the sum [30, 49, 67], and the histogram [61] of interruption durations. To ensure a smooth streaming when end-users face throughput fluctuations, e.g., in mobile networks, HAS allows to adapt the video bit rate to the network conditions. Thereby, initial delay and stalling 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, called quality variations [57].

Existing studies on overall quality were mostly limited to short sessions (about 1–3 minutes) [19, 21, 61, 63]. These studies mainly focused on the impact of the quality variations [4]. This impact is generally modeled by some statistics of segment quality values and switching amplitudes (i.e., differences between consecutive segment quality values) such as average [9, 11, 47, 63], standard deviation [63], minimum [19], median [19], histogram [61], and time duration on different quality levels [21].

For long sessions, the primacy and recency are also important factors to be considered. Here, the primacy (recency) factor refers to the influences of quality degradations near the beginning

(end) of a session. The authors in Reference [57] found that the primacy and recency both have significant impacts on the overall quality of a session. Reference [54] studies different temporal pooling methods, which emphasize different aspects (e.g., recency, lowest quality), for aggregating objective quality metrics into an overall quality score. In Reference [49], the authors proposed an overall quality model, taking into account the impacts of the quality variations, primacy, and recency. Specifically, a session is divided into three temporal intervals. In each interval, the impact of quality variations is modeled by the frequencies of switching types. Each switching type is defined based on resolutions and frame rates. To take into account the impact of the primacy and recency, each interval is simply assigned a weight to represent its contribution to the overall quality of the session. The experiment results then revealed that the first interval has the highest weight, and so the largest contribution to the overall quality.

In the latest stage of ITU-T P.1203 standardization for quality assessment of streaming media, a model (called P.1203) is recommended for predicting the overall quality, where session durations are from 1 to 5 minutes [41]. The P.1203 model also takes into account the impacts of quality variations, primacy, and recency. Then, to model the impact of quality variations, the authors used the average of the segment quality values in each temporal interval and various statistics calculated over a whole session, such as the total number of quality direction changes and the difference between the maximum and minimum segment quality. To take into account the impact of the primacy and recency, the authors used a weighted sum of all segment quality values in the session.

### 2.3 Cumulative Quality

To the best of our knowledge, the only previous study on the cumulative quality of a streaming session is in Reference [45], where the authors presented some qualitative observations regarding the impact of quality variations. However, the authors employed simple simulated sessions of very short durations (5–15 s) with only one to three segments. It is found that, when there is a quality variation with a small switching amplitude, the cumulative quality is quite stable. Meanwhile, a large switching amplitude results in a significant change of the cumulative quality. From these observations, the authors proposed a cumulative quality model, in which a piecewise linear function of switching amplitudes was used to quantify the impact of the quality variations.

The preliminary work of our cumulative quality research was presented in Reference [59]. In this article, the previous work is extended significantly in several aspects. First, we carried out more subjective tests with new videos and so the dataset is now doubled. Second, factors in the model are extensively studied with one-way analysis of variance (ANOVA). Third, different window sizes are analyzed and used for different window quality statistics. Fourth, two additional pooling modes of window quality values are investigated to validate the efficiency of the proposed model. Fifth, the model performance is explored in detail and the best setting is recommended. Finally, the evaluation is extended with seven more related models, two more test sets, and in-depth analysis of models' performances with respect to the length of sequences as well as models' computation complexity.

The contributions of our work have two general categories. First, we build a dataset that is specific to the cumulative quality. Our dataset helps to investigate how existing overall quality models perform cumulative quality prediction. Second, we propose a new cumulative quality model that can well predict the cumulative quality of streaming sessions. In particular, the distinguished features of our study are as follows.

- First, a subjective test was specifically designed for measuring the cumulative quality of HAS sessions. In our test, there are in total 72 test sequences generated from six 6-minute-long videos. The total time required for rating these sequences was approximately 160 hours.

Table 1. Features of Source Videos

Video	Content	Type	Video parameters
Video#1	Slow movements of characters	Animated video, Movie	720p, y4m, 24 fps
Video#2	A story about Sintel and her friend, a dragon.	Animated video, Movie	720p, y4m, 24 fps
Video#3	Conversations of characters	Natural video, Movie	4K, y4m, 24 fps
Video#4	A talk show host analyzing news	Natural video, News	720p, mp4, 24 fps
Video#5	A documentary about the science experiment	Natural video, Documentary	720p, mp4, 24 fps
Video#6	A soccer match	Natural video, Sport	720p, mp4, 24 fps

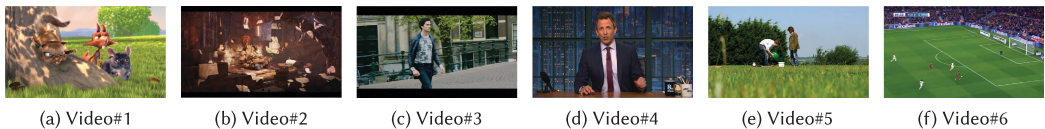


Fig. 1. Screenshots of source videos.

- Second, through statistical analysis, insights into the impacts of three factors of quality variations, primacy, and recency are provided. In particular, it is found that the impacts of the quality variations and recency are significant. However, no significant impact of the primacy is observed.
- Third, we proposed a new cumulative quality model that takes into account the impacts of the quality variations and recency. Experiment results show that the proposed model is able to predict well the cumulative quality of streaming sessions.
- Fourth, a comparison of the proposed model with ten existing models was conducted. This is the first time a large number of quality models have been investigated for cumulative quality prediction. Experiment results show that the proposed model outperforms the existing models.
- Fifth, a performance evaluation of the models for overall quality prediction was also conducted using two open test sets. The results show that the proposed model achieves the highest prediction performance for both the test sets.
- Sixth, it was found that the proposed model is applicable to real-time quality monitoring thanks to its low computation complexity. This feature is especially important for cost-effective evaluation of streaming technologies.

### 3 SUBJECTIVE TEST FOR CUMULATIVE QUALITY

In this study, to measure the cumulative quality over time, each streaming session was converted into test sequences of different lengths. In the test, each subject viewed a random sequence and then rated the quality of the whole sequence. This approach is similar to that used in Reference [45], where each 15-s-long session was divided into three sequences of 5, 10, and 15 (seconds).

There are in total six 6-minute-long videos used in this study, denoted by Video#1, Video#2, Video#3, Video#4, Video#5, and Video#6, with features presented in Table 1. Their screenshots are illustrated in Figure 1. These videos were downloaded from Xiph.org Test Media and YouTube. Similarly to References [51, 66], audio tracks were removed from the source videos to eliminate the influence of acoustic information. The videos were then encoded using H.264/AVC (libx264) with a frame rate of 24 fps. In practice, service providers can use different adaptation sets on their video streaming platforms. Even, in the future, the setting of adaptation sets is expected to be adaptable to content characteristics of individual streamed videos [1]. However, most existing studies use only

Table 2. Average Bitrates of Versions

Version	Average bitrate (kbps)					
	Video#1	Video#2	Video#3	Video#4	Video#5	Video#6
1	146	187	187	179	455	570
2	196	239	244	310	794	1,034
3	310	333	353	382	1,010	1,304
4	455	482	528	548	1,397	1,823
5	717	717	813	675	1,764	2,295
6	1,118	1,097	1,263	791	2,017	2,647
7	1,751	1,743	2,005	977	2,549	3,330
8	2,802	2,910	3,362	1,303	3,209	4,382
9	4,538	4,993	6,089	1,613	3,930	5,500

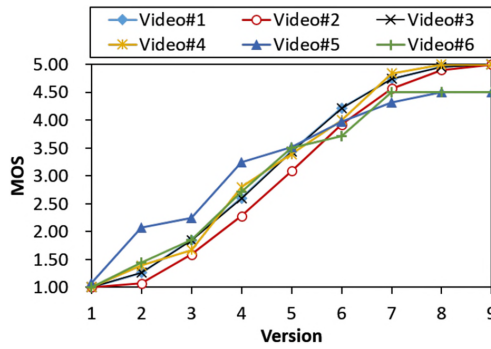


Fig. 2. Quality values of versions.

one adaptation set in their experiments [3, 11, 13]. In this study, we used two adaptation sets, each consisted of 9 versions with different QP values and/or resolutions. In particular, the nine versions in the first adaptation set have the same resolution of  $1280 \times 720$  and nine different QP values of 52, 48, 44, 40, 36, 32, 28, 24, and 20. This adaptation set was used to generate the streaming sessions of Video#1, Video#2, and Video#3. The nine versions in the second adaptation set are different in both resolution and QP. Specifically, the nine versions correspond to nine combinations of QP values and resolutions of  $\{24, 256 \times 144\}$ ,  $\{26, 426 \times 240\}$ ,  $\{24, 426 \times 240\}$ ,  $\{26, 640 \times 360\}$ ,  $\{24, 640 \times 360\}$ ,  $\{26, 854 \times 480\}$ ,  $\{24, 854 \times 480\}$ ,  $\{26, 1280 \times 720\}$ ,  $\{24, 1280 \times 720\}$ . This adaptation set was used to generate the streaming sessions of Video#4, Video#5, and Video#6. The average bitrates of the versions are shown in Table 2.

Figure 2 shows the quality values of the versions in MOS, which were calculated using an analytical function of encoding parameters proposed in Reference [34]. It can be seen that, because of different content characteristics, the quality of the same version is different across the videos. In addition, given a source video, the quality values of the versions are (roughly) evenly distributed over the rating scale from 1 to 5 MOS. It should be noted that, although version 1, which has very low quality, may be extremely annoying to users, it is still included in our experiment. The reason is that such low quality versions are currently used on popular video streaming platforms such as YouTube and Facebook (i.e., 144p or 240p quality versions). The aim is to avoid interruptions and so ensure smooth streaming, which is the primary objective of HAS [53]. In this study, every version was divided into short segments with the duration of 1 s.

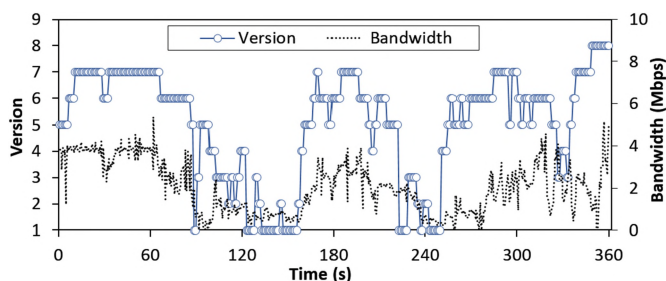


Fig. 3. An example of version variations in a streaming session.

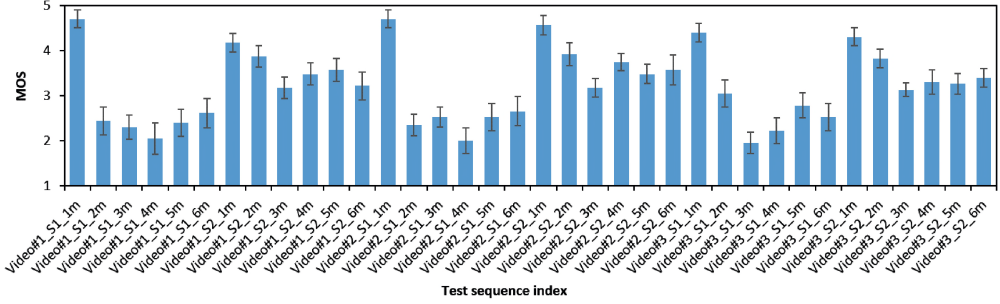
For each video, two full-length sessions of 6 minutes were generated by using the adaptation method of Reference [58] and two bandwidth traces from a mobile network [32]. The duration of 6 minutes was selected such that it is longer than the average video duration watched on YouTube, which is 5:01 minutes [33]. The used bandwidth traces have average throughputs varying from 1484.87 to 3432.33 kbps, and standard deviations from 867.01 to 1252.75 kbps. An example of version variations in a 6-minute session corresponding to a bandwidth trace is provided in Figure 3. In general, the selected versions tend to decrease following the bandwidth drops, and vice versa. Especially, besides smooth version switches (e.g., at the first 60 seconds), the used adaptation method also results in abrupt switches (e.g., at the 88th second) when the bandwidth falls dramatically. This enables our dataset to cover both smooth and abrupt switches in practice.

From each full-length session, six test sequences were extracted, from the timestamp 0 to the 1st, 2nd, 3rd, 4th, 5th, and 6th minutes. So, from the six original videos, there were in total 72 test sequences, with durations from 1 minute to 6 minutes. The total duration of all the test sequences is 252 minutes. Because a rating time that is longer than 1.5 hours may cause fatigue and boredom [44], the subjective test was divided into four parts that were conducted in different days. The duration of each part was approximately 1.5 hours, of which about 1 hour was spent for rating the test sequences. In the rating process, every 20 minutes, there was a break of 10 minutes. In order to avoid boredom, each subject took part in at most two test parts.

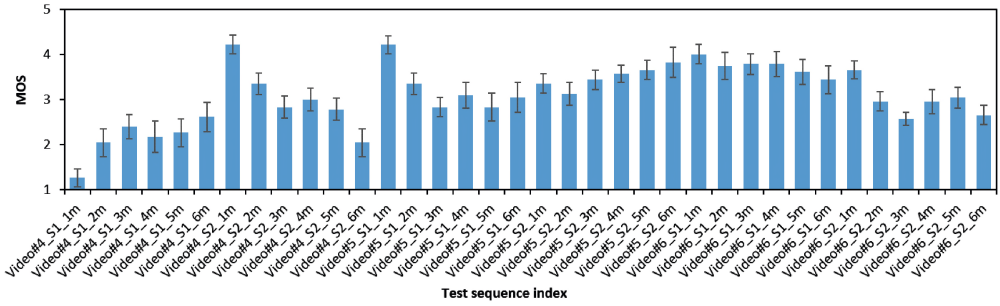
The subjective test was conducted using the absolute category rating method. Test conditions were designed following Recommendation ITU-T P.913 [44]. In the subject-training stage, the subjects got used to the procedure and the range of quality impairments. In the test, the sequences were randomly displayed on a black background using a LED screen with the size of 14 inches and the resolution of 1,366×768. An experimental interface was designed to play the individual test sequences and record the corresponding ratings. In particular, given a sequence displayed, a 1-s blank screen with 50% grey was presented at the end of the sequence. After that, each subject gave a score with the value ranging from 1 (worst) to 5 (best), which reflects his/her option of quality of the whole sequence. Following a 1-s blank screen with 50% grey, the rating process was repeated for the next sequence.

There were in total 71 subjects with 43 male and 28 female taking part in the test. They range in age from 20 to 30. The total time of the test was approximately 160 hours. Screening analysis of the test results was performed following Recommendation ITU-T P.913 [44], and two subjects were rejected. After discarding these subjects' scores, each test sequence was rated by 23 valid subjects. The MOS of each sequence was computed as the average of the valid subjects' scores.

The MOSs of the test sequences are shown in Figure 4, where the error bars represent the 95% confidence intervals. Here each test sequence is denoted by a structure of  $\{VideoID\}_{SessionID}_{Duration}$ . In particular, *VideoID* allows to determine the source video used



(a) Video#1, Video#2, and Video#3



(b) Video#4, Video#5, and Video#6

Fig. 4. MOSs of test sequences and their 95% confidence intervals.

to generate that sequence (i.e., from Video#1 to Video#6). *SessionID* is to distinguish the two sessions of the same source video (i.e., *S1* or *S2*). *Duration* denotes the length of the sequence (i.e., from *1m* for 1 minute to *6m* for 6 minutes). For example, *Video#3\_S2\_3m* denotes the sequence extracted from the timestamp 0 to the 3rd minute of the second streaming session of Video#3. From Figure 4, we can see that the cumulative quality varies drastically during a session. In addition, the MOSs are in the range from 1.3 to 4.7. Also, the 95% confidence intervals are in the range from 0.09 to 0.35.

## 4 CUMULATIVE QUALITY MODEL

### 4.1 Overview

To build a cumulative quality model taking into account the impacts of multiple factors, the basic ideas of our solution are as follows.

- Quality variations over a long session are divided into long-term and short-term changes. Specifically, short-term changes refer to quality variations of neighboring segments, while long-term changes refer to quality variations between temporal intervals.
- To represent the impact of long-term changes, the concept of “sliding window” is used. Specifically, a window of  $K$  segments is moved along the session, segment by segment as illustrated in Figure 5. After each time, a window quality value is computed.
- To represent the impact of short-term changes within a window, an existing overall quality model is used. For this purpose, such a model is called *window quality model*. It should be noted that, besides short-term changes, a window quality model should additionally take into account the impacts of initial delay and interruptions appearing in the window, since they are also key factors affecting the human quality perception [4, 23].

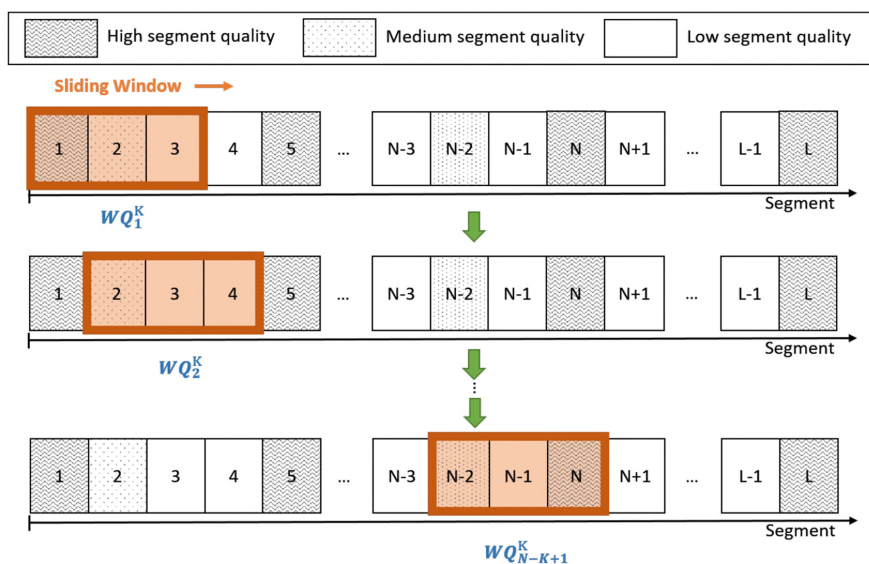


Fig. 5. An illustration of “sliding window” with size  $K = 3$ .

- The cumulative quality value at any time point is computed based on window quality values, taking into account the impacts of factors such as long-term changes and recency. Note that, at the first time points, when the watched video duration is (very) short (i.e., less than  $K$  segments), the corresponding cumulative quality values are directly computed from the window quality model.

In the next subsection, a window quality model that is our previous overall-quality model in Reference [62] (called Tran’s) will be presented. This model was found to be very effective in representing the impact of short-term changes, initial delay, and interruptions. In particular, its Pearson Correlation Coefficients (PCC) with MOSs is always higher than or equal to 0.90 as reported in References [61, 62].

## 4.2 Window Quality Model

A detailed description of Tran’s model for a window (or a short video) was presented and validated in Reference [62]. In this part, we just highlight the key points of that model. Given a window, it is assumed that each segment is represented by a quality value, which can be obtained by a subjective test or by an analytical function of the corresponding encoding parameters [60]. Although a subjective test could provide more reliable results, it is very costly, time-consuming, labour-intensive, and so difficult to be applied to real-time quality monitoring in practice. To overcome this issue, some previous studies have proposed analytical functions inputted by encoding parameters to predict segment quality. It is much simpler and easier to integrate such functions into real service platforms [34, 39, 56].

In this study, given a window, we first employed the analytical function proposed in Reference [34] to calculate segment quality values, since its performance was found to be very high in our previous study [60]. Then, the histogram of segment quality values, the histogram of quality switches, and the histogram of interruptions were calculated. Finally, a weighted sum was applied to these histograms to calculate the window quality value. Here the histograms of segment quality values and quality switches are to represent the short-term quality changes in the window.

In particular, the quality  $WQ_i^K$  of window  $\{i|i > 1\}$  is given by

$$WQ_i^K = \max \left( \sum_{n=1}^N \alpha_n F_n^Q - \sum_{i=1}^N \sum_{j=-M}^{-1} \beta_{i,j} F_{i,j}^V - \beta^{um} F^{um} - \sum_{l=1}^L \gamma_l F_l^I, 1 \right), \quad (1)$$

where  $\alpha_n, \beta_{i,j}, \beta^{um}$ , and  $\gamma_l$  are model parameters,  $N$  is the number of segment quality bins,  $F_n^Q$  is the frequency of segment quality bin  $B_n^Q$ ,  $N \times M + 1$  is the number of quality switching bins,  $F_{i,j}^V$  is the frequency of down-switching bin  $B_{i,j}^V$ ,  $F^{um}$  is the total frequency of up-switches and quality maintaining, and  $F_l^I$  is the frequency of interruption bin  $B_l^I$ .

Since the initial delay appears only once at the beginning of a session, its impact is just considered in the first window. Specifically, the first window quality  $WQ_1^K$  is computed by

$$WQ_1^K = \max \left( \sum_{n=1}^N \alpha_n F_n^Q - \sum_{i=1}^N \sum_{j=-M}^{-1} \beta_{i,j} F_{i,j}^V - \beta^{um} F^{um} - \sum_{l=1}^L \gamma_l F_l^I - \sigma \log(ID + \mu), 1 \right), \quad (2)$$

where  $ID$  denotes the duration of the initial delay,  $\sigma$  and  $\mu$  are model parameters.

As the model parameters  $\alpha_n, \beta_{i,j}, \beta^{um}, \gamma_l, \sigma$ , and  $\mu$  in Equations (1) and (2) have been already obtained and validated in References [61, 62], they were reused in this study. In general, a higher segment quality value has a bigger weight and so a more substantial contribution to the window quality. For quality switches, a larger switching amplitude has a higher weight and consequently a more adverse impact. Also, an interruption with the longer duration results in a more severe effect. It should be noted that the implementation of this model is also included in our public source code.

In the next subsection, effect analysis of the quality variations, primacy, and recency will first be presented. Then, based on the obtained results, a cumulative quality model will be proposed.

### 4.3 Proposed Quality Model

As mentioned, to identify the key components of a cumulative quality model, we carried out a statistic analysis of some window quality values. In particular, the first window quality value  $WQ_f^K$  and the last window quality value  $WQ_l^K$  were employed to represent the impacts of the primacy and recency, respectively. For the factor of long-term changes, three window quality statistics are considered, which are the average window quality  $WQ_{av}^K$ , the maximum window quality  $WQ_{ma}^K$ , and the minimum window quality  $WQ_{mi}^K$  of all windows until a given time point.

Suppose that the window is just moved to the  $N^{th}$  segment with  $N \geq K$ . By using the window quality model, the window quality value  $WQ_{N-K+1}^K$  is calculated. After that, the window quality statistics of  $WQ_f^K, WQ_l^K, WQ_{av}^K, WQ_{ma}^K$ , and  $WQ_{mi}^K$  are updated by the following equations:

$$WQ_f^K = WQ_1^K, \quad (3)$$

$$WQ_l^K = WQ_{N-K+1}^K, \quad (4)$$

$$WQ_{av}^K = \begin{cases} WQ_1^K, & \text{if } N = K \\ \frac{WQ_{av}^K \times (N-K) + WQ_{N-K+1}^K}{N-K+1}, & \text{otherwise,} \end{cases} \quad (5)$$

$$WQ_{mi}^K = \begin{cases} WQ_1^K, & \text{if } N = K \\ \min\{WQ_{mi}^K, WQ_{N-K+1}^K\}, & \text{otherwise,} \end{cases} \quad (6)$$

$$WQ_{ma}^K = \begin{cases} WQ_1^K, & \text{if } N = K \\ \max\{WQ_{ma}^K, WQ_{N-K+1}^K\}, & \text{otherwise} \end{cases} \quad (7)$$

Table 3 shows the obtained results from one-way ANOVA. To assess the effect size, partial Eta-squared values ( $\eta_p^2$ ) are also reported. In some previous experiments related to the human ability to

Table 3. Results of Effect Analysis of Window Quality Statistics

Window quality statistics		Window size $K$ (seconds)					
		10	20	30	40	50	60
$WQ_f^K$	$F$	4.868	1.594	8.589	2.088	7.321	1.478
	$p$	0.027	0.207	0.003	0.149	0.007	0.224
	$\eta_p^2$	0.003	0.001	0.005	0.001	0.004	0.001
$WQ_l^K$	$F$	2.111	0.149	6.959	6.687	18.977	16.063
	$p$	0.146	0.699	0.008	0.010	<0.001	<0.001
	$\eta_p^2$	0.001	0.000	0.004	0.004	0.010	0.008
$WQ_{av}^K$	$F$	4.103	9.359	0.207	1.404	11.613	44.283
	$p$	0.043	0.002	0.649	0.236	<0.001	<0.001
	$\eta_p^2$	0.002	0.005	0.000	0.001	0.006	0.023
$WQ_{mi}^K$	$F$	3.826	2.202	3.338	16.730	38.648	6.397
	$p$	0.051	0.138	0.068	<0.001	<0.001	0.012
	$\eta_p^2$	0.002	0.001	0.002	0.009	0.020	0.003
$WQ_{ma}^K$	$F$	12.075	0.896	1.971	16.366	19.644	6.958
	$p$	<0.001	0.344	0.161	<0.001	<0.001	0.008
	$\eta_p^2$	0.006	0.000	0.001	0.009	0.010	0.004

memorize items such as numbers, words, and syllables, the duration of human short-term memory was found to be in range from 15 to 30 s [35, 46]. Therefore, here, the window size  $K$  is set from 10 to 60 s with the step size  $S$  of 10 s. Obviously, the choice of a step size  $S$  has a tradeoff between accuracy and computation complexity. To determine a suitable step size, we investigated 20 different step sizes  $S$  from 1 to 20. According to the criterion of just noticeable difference, the impact of step size  $S$  on the quality difference between window sizes  $K$  is trivial when  $S < 10$ , but noticeable for  $S \geq 10$ . Therefore, the step size  $S$  was set to 10 in this article, since it is the smallest value that could provide significant quality differences between the investigated window sizes. Here, the window quality model is Tran's model that is presented in Section 4.2.

The  $p$  values in Table 3 indicate that, for all the considered window sizes, no significant effect was observed for  $WQ_f^K$  (i.e.,  $p > 0.001$ ). In contrast, significant results with small effects were obtained for  $WQ_l^K$  (i.e.,  $p < 0.001$  and  $\eta_p^2 < 0.06$ ) when the window size  $K$  is 50 or 60 s. Especially, the larger effect size was found for the window size of 50 s (i.e.,  $\eta_p^2 = 0.010$  vs.  $\eta_p^2 = 0.008$ ). This implies that the impact of the primacy on the cumulative quality can be neglected, while the impact of the recency has to be considered.

With regard to long-term changes, some significant effects with small sizes were also observed for  $WQ_{av}^K$ ,  $WQ_{mi}^K$ , and  $WQ_{ma}^K$  (i.e.,  $p < 0.001$  and  $\eta_p^2 < 0.06$ ). Particularly, the window size corresponding to the strongest effect size is 60 s for  $WQ_{av}^K$  (i.e.,  $\eta_p^2 = 0.023$ ), 50 s for  $WQ_{mi}^K$  (i.e.,  $\eta_p^2 = 0.020$ ), and 50 s for  $WQ_{ma}^K$  (i.e.,  $\eta_p^2 = 0.010$ ). This implies that the three window quality statistics of the average, minimum, and maximum quality should be considered.

To sum up, the results suggest that  $WQ_l^K$ ,  $WQ_{av}^K$ ,  $WQ_{mi}^K$ , and  $WQ_{ma}^K$  should be key components of a cumulative quality model. Based on these observations, we propose a cumulative quality model with three different pooling modes of window quality values as follows. It should be noted that, while the first mode (called Selected) and the second mode (called Broad) include different window sizes, window quality statistics in the so-called Fixed mode have the same window size.

**4.3.1 Selected Mode.** In the Selected mode, we use a weighted sum of four window quality statistics, namely  $WQ_l^{50}$ ,  $WQ_{av}^{60}$ ,  $WQ_{mi}^{50}$ , and  $WQ_{ma}^{50}$ , that have the strongest effect sizes as analyzed above. Specifically, the cumulative quality value  $CQM$  is calculated by

$$CQM = w_1 \cdot WQ_l^{50} + w_2 \cdot WQ_{av}^{60} + w_3 \cdot WQ_{mi}^{50} + w_4 \cdot WQ_{ma}^{50}, \quad (8)$$

where  $w_1, w_2, w_3$ , and  $w_4$  are the corresponding weights of  $WQ_l^{50}$ ,  $WQ_{av}^{60}$ ,  $WQ_{mi}^{50}$ , and  $WQ_{ma}^{50}$  components, respectively.

**4.3.2 Broad Mode.** The so-call Broad mode consists of all the window quality statistics whose effects are significant (i.e.,  $p > 0.001$ ). In particular, the cumulative quality value  $CQM$  is computed by the following equation:

$$CQM = a_1 \cdot WQ_l^{50} + a_2 \cdot WQ_l^{60} + a_3 \cdot WQ_{av}^{50} + a_4 \cdot WQ_{av}^{60} + a_5 \cdot WQ_{mi}^{40} + a_6 \cdot WQ_{mi}^{50} + a_7 \cdot WQ_{ma}^{10} + a_8 \cdot WQ_{ma}^{40} + a_9 \cdot WQ_{ma}^{50}, \quad (9)$$

where  $a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8$ , and  $a_9$  are the corresponding weights of  $WQ_l^{50}, WQ_l^{60}, WQ_{av}^{50}, WQ_{av}^{60}, WQ_{mi}^{40}, WQ_{mi}^{50}, WQ_{ma}^{10}, WQ_{ma}^{40}$ , and  $WQ_{ma}^{50}$  components, respectively.

**4.3.3 Fixed Mode.** The Fixed mode that contains all the five window quality statistics with the same window size is given by

$$CQM = b_1 \cdot WQ_f^K + b_2 \cdot WQ_l^K + b_3 \cdot WQ_{av}^K + b_4 \cdot WQ_{mi}^K + b_5 \cdot WQ_{ma}^K, \quad (10)$$

where  $b_1, b_2, b_3, b_4$ , and  $b_5$  are the corresponding weights of  $WQ_f^K, WQ_l^K, WQ_{av}^K, WQ_{mi}^K$ , and  $WQ_{ma}^K$  components, respectively.

In the next section, we will investigate the performance of the proposed model using the three pooling modes and different window quality models, in comparison with some existing models.

## 5 MODEL EVALUATION AND ANALYSIS

### 5.1 Evaluation Methodology

This section is divided in four evaluations, each aiming at an important question. In the first evaluation, we will investigate what is the best window quality model for the proposed model. The second evaluation aims to determine which pooling mode is most effective for cumulative quality prediction. The third one is carried out to see if existing overall quality models can predict the cumulative quality, especially in long sessions. The last focuses on the performance of the proposed model in overall quality prediction, compared to existing models.

There are in total 10 existing models employed in this study, which are denoted by Tran's [61], Guo's [19], Vriendt's [63], Yin's [69], P.1203 [29, 37, 41, 48], SQI [13], KSQI [9, 10], Eswara's [14, 15], Rehman's [45], and preCQM [59]. Note that the preCQM model is one proposed in our preliminary work [59]. In the preCQM model,  $WQ_{ma}^K$ , which is one of the key components in the proposed model, is not included. In addition, all the components (or the window quality statistics) have the same window size. Meanwhile, different window sizes are selected for different components in the proposed model.

Among these models, only the Rehman's and preCQM models were proposed for cumulative quality prediction. Eswara's model was devoted to continuous quality prediction and the other models were originally built for overall quality prediction. Similarly to References [12, 13], to evaluate the performance of existing models, except the P.1203, SQI, KSQI, and Eswara's models, we re-implemented the models using the parameter settings stated in the corresponding publications. The reason is that the implementations of these models are not publicly available. For the remaining models, we used the corresponding implementations publicized by the original authors [10, 14, 29, 37, 48].

In general, given a streaming session, each reference model first calculates segment quality values using a quality metric. Particularly, the SQI, KSQI, and Eswara’s models use Video Multi-Method Assessment Fusion (VMAF), which is a metric developed by Netflix [27, 28]. Meanwhile, the others employ MOS that can be calculated using some analytical functions of encoding parameters [34, 39, 40]. Next, some statistics such as the average and minimum of segment quality values are derived. Finally, analytical functions or machine learning algorithms are applied to compute the predicted scores. Note that, in Eswara’s model, we used the mean pooling of continuous quality values to obtain the predicted cumulative and overall quality scores, which was also used in the original publication [15]. In addition, following Recommendation ITU-T P.1401 [42], a first order linear regression between the predicted scores and MOSs was performed for each model to compensate for possible variances between subjective tests. The obtained coefficients of slope and intercept will be stated in the following subsections.

For the performance evaluations in cumulative quality prediction, we randomly selected three videos among the six source videos used in our dataset. The set of all the 36 sequences generated from these three videos was used as a training set. The 36 remaining sequences constituted a test set. The selection was repeated  $\binom{6}{3} = 20$  times, resulting in 20 unique pairs of training and test sets. The training set was used to obtain the model parameters by curve fitting. The test set was to evaluate the performance of the models. Note that, to obtain MOSs of segments, we used the analytical function proposed in Reference [34] as described in Section 4.2. In addition, because there has been no open cumulative quality dataset so far, the evaluation was conducted using only our dataset. The obtained results will be analyzed in Sections 5.2 and 5.3.

Although many overall quality datasets are publicly available [3, 6, 9, 16, 18, 48], most of them contain only very short sessions (i.e., less than 1 minute) [9, 11–13] or short sessions (i.e., less than 3 minutes) [3, 16, 18]. Meanwhile, the average video duration watched on YouTube is up to 5:01 minutes [33] as mentioned in Section 3. In addition, a large number of datasets include only either quality variations or interruptions [6, 11, 13, 18]. To the best of our knowledge, only P.1203 dataset in [48] includes both short and long sessions with appearances of not only quality variations but also initial delay and interruptions, which are key factors affecting the human perceived quality [4]. In particular, this dataset includes two test sets, denoted *VL04* and *VL13*. The *VL04* set consists of sixty 1-minute-long sessions, and the *VL13* set contains fifteen 4-minute-long sessions. Note that, in these sets, MOSs of segments have been already provided (denoted *O.34* in the original publication [48]), which were calculated by the P.1203 model at input mode 3. A detailed discussion on the obtained results will be presented in Section 5.4.

In order to measure the performance of the models, we used two metrics of PCC and Root-Mean-Squared Error (RMSE). For cumulative quality prediction, the PCC and RMSE values reported below were calculated over the 20 test sets of our dataset. Meanwhile, for overall quality prediction, the PCC and RMSE values were derived over the two test sets of *VL04* and *VL13* in the P.1203 dataset. Since the capability of real-time processing is an especially important feature for cumulative quality models, we also measured the computation complexity of the models. In this study, the computation complexity was measured as the average time required to obtain a cumulative quality value per 1-s-long segment. The measurement was conducted on a computer with Intel Core i3-2120 processor at 3.30 GHz and 8 GB RAM.

## 5.2 Performance Analysis of CQM Model in Cumulative Quality Prediction

In this subsection, we first investigate the performance of the proposed model using different window quality models. Our goal is to find the best window quality model for the proposed model. Then, to determine the best pooling mode, a performance comparison between the three pooling

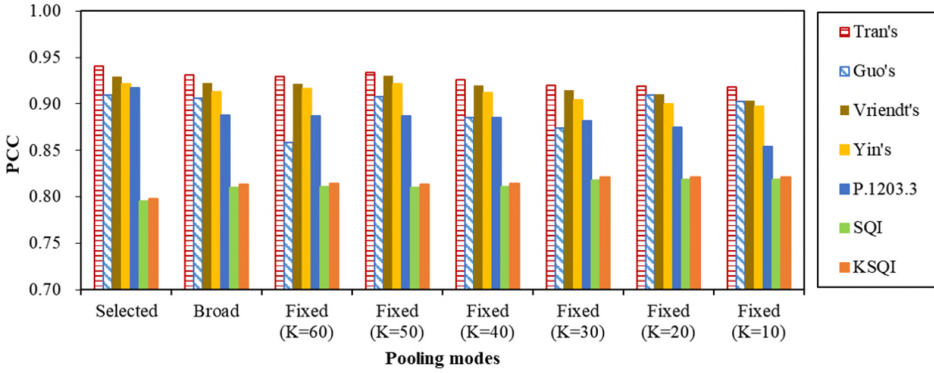


Fig. 6. Average performance of the proposed model using different window quality models over 20 test sets.

Table 4. Average and Median Performances of CQM Model Using Different Window Quality Models with the Selected Mode

Window quality model	Average performance				Median performance			
	Training sets		Test sets		Training sets		Test sets	
	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE
Tran's [61]	0.94	0.26	0.94	0.28	0.94	0.26	0.94	0.29
Guo's [19]	0.91	0.30	0.91	0.31	0.91	0.31	0.91	0.32
Vriendt's [63]	0.93	0.27	0.93	0.28	0.94	0.28	0.94	0.29
Yin's [69]	0.92	0.29	0.92	0.29	0.93	0.30	0.93	0.30
P.1203 [41]	0.92	0.30	0.92	0.30	0.92	0.30	0.92	0.30
SQI [13]	0.80	0.47	0.80	0.47	0.79	0.47	0.79	0.47
KSQI [9]	0.80	0.47	0.80	0.47	0.79	0.47	0.79	0.47

modes is carried out. Finally, for quantitative analysis on the contributions of the components in the proposed model, the model parameters are determined and discussed.

**5.2.1 Window Quality Model.** In this part, a performance evaluation of the CQM model using different window quality models will be presented. In particular, the seven overall quality models of Tran's, Guo's, Vriendt's, Yin's, P.1203, SQI, and KSQI are employed to obtain window quality values. Note that these models all take into account the impact of short-term changes. Further, note that Eswara's is a continuous quality model and Rehman's and preCQM are cumulative quality models, which were not used here, but are only used later for comparison purpose.

Figure 6 depicts the average performance of the proposed model with the three pooling modes using the different window quality models over the 20 test sets. Note that, similar to Section 4, the window size in the Fixed model is also set from 10 to 60 s with the step size of 10 s. It can be seen that, regardless of the pooling modes, the performance of the CQM model is generally good with all the window quality models (i.e.,  $PCC \geq 0.84$ ), except SQI and KSQI.

Especially, for all the pooling modes, the use of Tran's model as a window quality model always provides the best prediction performance. Specifically, the average PCC values are in range from 0.91 to 0.94 and the average RMSE values are from 0.28 to 0.36. Table 4 shows the average and median performances of the CQM model with the Selected mode using the different window quality models. It can be seen that, when using the window quality model of Tran's, the proposed model achieves very high performance in both average and median. In particular, the average PCC and

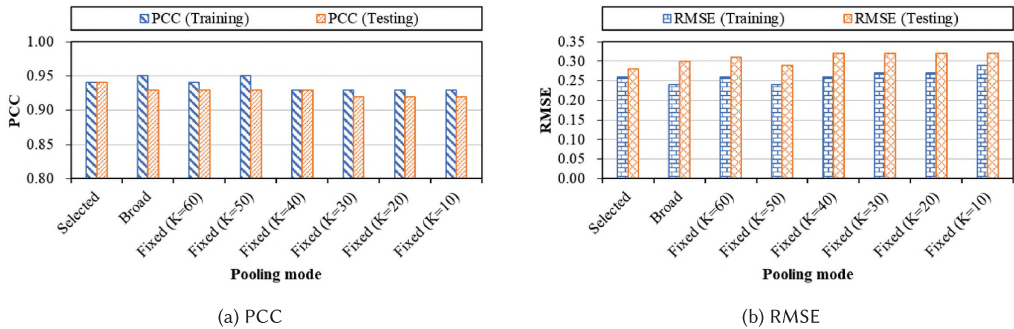


Fig. 7. Average performance of the CQM model using different pooling modes with the window quality model of Tran’s.

RMSE values are 0.94 and 0.26 for the training sets, and 0.94 and 0.28 for the test sets. In addition, the median PCC and RMSE values are 0.94 and 0.26 for the training sets, and 0.94 and 0.29 for the test sets. The main reason is that, for modeling the impact of short-term changes, Tran’s model utilizes the histograms of segment quality values and quality switches (as mentioned in Section 4.2), which are shown to be more effective than the statistics used in the remaining models [60]. From the results, it is suggested that Tran’s model should be used to calculate window quality values in the proposed model.

For the window quality models of SQI and KSQI, their performances are significantly lower than those of the other models as presented in Figure 6. Particularly, their average PCC values are lower than or equal to 0.82. Note that similar PCC values of these two models were also reported in Reference [9] (i.e., PCC = 0.76 for the SQI model and PCC = 0.79 for the KSQI model). One possible reason is that these models were originally built based on very short sessions (i.e., 8- and 10-s-long sessions). Therefore, when calculating window quality values with the large sizes  $K$  of 50 and 60, they do not perform very well. Also, compared to small window sizes (i.e.,  $K \leq 30$ ), the Fixed mode using the two models with large window sizes (i.e.,  $K \geq 40$ ) has consequently lower performance as shown in Figure 6. However, all the cases result in quite low performances, suggesting the SQI and KSQI models are not very effective to calculate window quality values in the proposed model.

Since the Tran’s model provides the best performance, it is used as the window quality model in the rest of this article.

**5.2.2 Pooling Mode.** In this part, we investigate the performance of the proposed model using the different pooling modes. The obtained results are shown in Figure 7. It can be seen that the average performances of all the modes are quite high (i.e., PCC  $\geq 0.91$  and RMSE  $\leq 0.36$  for the test sets). Obviously, the Broad mode has the highest PCC and the lowest RMSE for the training sets as it includes much more parameters than the others. However, for the test sets, the best performance is derived by the Selected mode (i.e., PCC = 0.94 and RMSE = 0.28). One possible reason of the lower performance of the Broad mode could be an over-fitting phenomenon because of its large number of parameters [8]. Regarding the Fixed mode, its performance with any window size  $K$  is always lower than that of the Selected mode for the test sets. This implies that, in comparison to using the same window size, the use of different window sizes for different window quality statistics is more effective in cumulative quality prediction.

Regarding the computation complexity of each pooling mode, it mainly depends on (1) the number of different window sizes (denoted  $N_{ws}$ ) and (2) the number of window quality statistics (denoted  $N_{wqs}$ ) employed in that mode. A higher value of either  $N_{ws}$  or  $N_{wqs}$  results in a

Table 5. Computation Complexity of the Proposed Model Using Different Pooling Modes of Window Quality Values

Pooling mode		$N_{wqs}$	$N_{ws}$	Computation complexity (ms)	
				In serial	In parallel
Selected		4	2	0.31	0.16
Broad		9	4	0.60	0.16
Fixed	K = 60	5	1	0.16	
	K = 50			0.16	
	K = 40			0.15	
	K = 30			0.17	
	K = 20			0.15	
	K = 10			0.15	

larger computation complexity. Note that, although the impact of  $N_{ws}$  is more severe, it can be eliminated by parallel processing of different window sizes.

Table 5 shows the computation complexity of the three modes in serial and parallel processing when using the same window quality model of Tran's. It can be seen that, while the computation complexity of all the modes is similar for parallel processing (i.e., about 0.16 ms), there are significant differences reported for serial processing. Particularly, the Broad mode has approximately two times as much computation complexity as the Selected mode has (i.e., 0.60 vs. 0.31). The main reason is that  $N_{ws}$  of the Broad mode is about two times higher than that of the Selected mode. In a similar way, the computation complexity of the Selected mode is also more than half of the Fixed mode. However, it can be seen that the complexity of all the modes is lower than 1 ms for both serial and parallel processing. Therefore, the predicted cumulative quality can be updated after every segment as the window slides forward. In other words, all the modes are applicable to real-time quality monitoring.

From the above discussions, we can conclude that the Selected mode, which includes the four window quality statistics with the different window sizes, is really efficient and effective for cumulative quality prediction, especially in parallel processing. Therefore, it will be used in the rest of this article.

**5.2.3 Analysis of Model Parameters.** In this subsection, we first determine the model parameters by averaging the individual parameters obtained using the 20 training sets. Next, based on these parameters, quantitative analysis on the contributions of the components are provided.

In particular, the cumulative quality model is given by

$$CQM = w_1 \cdot WQ_l^{50} + w_2 \cdot WQ_{av}^{60} + w_3 \cdot WQ_{mi}^{50} + w_4 \cdot WQ_{ma}^{50}, \quad (11)$$

$$= 0.31 \cdot WQ_l^{50} + 0.37 \cdot WQ_{av}^{60} + 0.31 \cdot WQ_{mi}^{50} + 0.01 \cdot WQ_{ma}^{50}. \quad (12)$$

The positive numerical values of the weights  $w_1$ ,  $w_2$ ,  $w_3$ , and  $w_4$  reconfirm the observations in Section 4 that  $WQ_l^K$ ,  $WQ_{av}^K$ ,  $WQ_{mi}^K$ , and  $WQ_{ma}^K$  are key components of the cumulative quality model. Also, the impacts of the quality variations and recency are significant on the cumulative quality of a session. In addition, it can be seen that  $w_2$  is highest while  $w_4$  is lowest. So the impact of the average window quality is strongest, and the impact of the maximum window quality is weakest.

It is interesting to note that these results are in agreement with the peak-end rule [24]. The peak-end rule says that users judge an experience largely at its lowest peak and at its end. Here the peaks (lowest and highest) of a session are the minimum window quality  $WQ_{mi}^{50}$  and the maximum

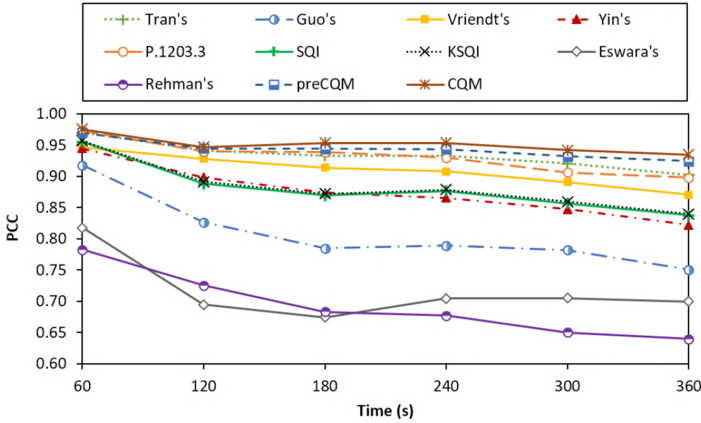


Fig. 8. Average performance of models with different sequence lengths.

window quality  $WQ_{ma}^{50}$ . Also, the end of a session is the last window quality  $WQ_l^{50}$ . It can be seen that the sum of  $w_1$  and  $w_3$  is 0.62 while the sum of the others is 0.38. Therefore, the human cumulative quality perception also mainly depends on the lowest peak (or the minimum window quality) and the end (or the last window quality). In comparison between the (lowest and highest) peaks, the lowest one has more substantial impact as  $w_3$  is much higher than  $w_4$ . The possible explanation is that users tend to pay more attention to the worst discomfort [24]. In the case of speech quality, a large impact of the minimum quality on QoE was also shown [25]. Also, Reference [54] showed that a good temporal pooling method is taking the average over the whole session, implying that  $WQ_{av}^{60}$  is a key influence factor. Thus all the key factors of the proposed model are inline with the findings in previous studies. Yet, the CQM model is the first one that integrates these factors into a single model for predicting the cumulative quality of HAS sessions.

### 5.3 Model Comparison in Cumulative Quality Prediction

In this subsection, we compare the CQM model and the ten existing models in terms of the performance and the computation complexity in cumulative quality prediction. Figure 8 shows the average performances of the models with different sequence lengths. We can see that, when the sequence length is 1 minute, the PCC values of Tran's, Guo's, Vriendt's, Yin's, P.1203, SQI, and KSQI models are high (i.e.,  $PCC \geq 0.92$ ). This suggests that these models can predict well the overall quality of a short session, and thus most of them can be used as window quality models with good performance as discussed in Section 5.2.1.

However, when the sequence length increases, the PCC values of the models tend to decrease. Among the models, the PCC of the CQM model is highest for all the sequence lengths, implying that CQM is the best model for cumulative quality prediction of streaming sessions. Meanwhile, the performances of Esvara's and Rehman's models are lowest. A possible explanation is that Esvara's model was actually developed for continuous quality prediction, but not cumulative quality prediction. The mean pooling strategy of continuous quality values used in this model may be not an effective way to predict the cumulative quality values. For Rehman's model, it was originally designed for cumulative quality prediction of only very short sessions with a duration of 5–15 s. Thus it is not really suitable for longer sessions (i.e., 1–6 minutes). In addition, there is no simultaneous consideration for long-term changes and recency in Tran's, Guo's, Vriendt's, Yin's, SQI, KSQI, Esvara's, and Rehman's models, so their performances are all lower than that of the CQM model. It turns out that the simple preCQM model's performance is only a little worse than that

Table 6. Average Performances and Computation Complexity of the Models in Cumulative Quality Prediction

Model	Coefficients		Performance (Test set)		Computation complexity (ms)
	Slope	Intercept	PCC	RMSE	
Tran's [61]	1.24	-1.27	0.91	0.31	0.23
Guo's [19]	1.01	-0.25	0.76	0.48	0.01
Vriendt's [63]	1.02	-0.41	0.89	0.35	0.02
Yin's [69]	1.07	-0.79	0.85	0.41	0.07
P.1203 [41]	1.04	-0.93	0.91	0.32	1817.25
SQI [13]	2.48	-8.17	0.84	0.40	2.57
KSQI [9]	2.47	-8.12	0.84	0.39	4.31
Eswara's [15]	0.69	0.72	0.72	0.70	0.36
Rehman's [45]	25.11	-26.68	0.63	0.56	0.06
preCQM [59]	—	—	0.93	0.33	0.16
<b>CQM</b>	—	—	<b>0.94</b>	<b>0.28</b>	<b>0.16</b>

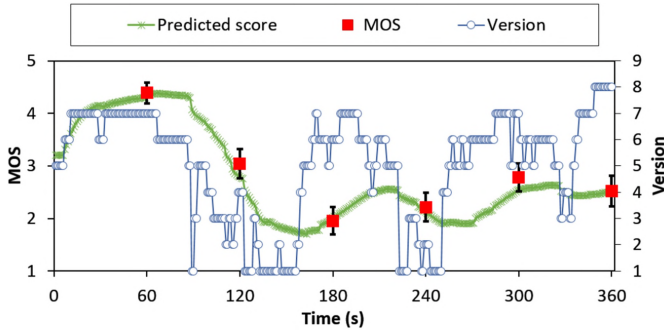


Fig. 9. An example of the cumulative quality values of a streaming session.

of the CQM model. It is mainly because the impact of maximum window quality is actually small in the obtained CQM model.

Table 6 summarizes the average performances and the computation complexity of the models. Here, the PCC and RMSE are averaged over the 20 test sets containing sequences of different lengths. We can see that the results of performances are similar to those in Figure 8. In particular, the performance of the CQM model is highest and the performances of Eswara's and Rehman's model are lowest.

Regarding the computation complexity, it can be seen that the CQM model takes less than 1ms to obtain a cumulative quality value, and so it is applicable to real-time quality monitoring as presented in Section 5.2.2. It should be noted that here the computation complexity of the CQM model is in case where different window sizes are computed in parallel.

For the P.1203 model, its computation complexity is considerably higher than the others. In particular, the P.1203 model takes an average of 1.81 s to calculate a cumulative quality value. Meanwhile, the remaining models have an average processing time less than 5 ms per cumulative quality value.

Table 7. Performance of the Models in Overall Quality Prediction for *VL04* and *VL13* Sets

Test set	VL04 (1-minute)				VL13 (4-minute)			
	Coefficients		Performance		Coefficients		Performance	
	<i>Slope</i>	<i>Intercept</i>	<i>PCC</i>	<i>RMSE</i>	<i>Slope</i>	<i>Intercept</i>	<i>PCC</i>	<i>RMSE</i>
Tran’s [61]	0.79	0.82	0.90	0.39	1.17	-0.60	0.90	0.46
Guo’s [19]	0.75	0.67	0.72	0.62	0.94	0.22	0.75	0.68
Vriendt’s [63]	0.80	0.50	0.71	0.63	1.10	-0.44	0.80	0.62
Yin’s [69]	0.55	1.61	0.80	0.54	0.65	1.35	0.86	0.53
P.1203 [41]	1.04	0.08	0.88	0.42	1.18	-0.51	0.92	0.40
SQI [13]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
KSQI [9]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Eswara’s [15]	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rehman’s [45]	4.00	-1.94	0.36	0.87	4.00	-1.59	0.29	1.03
preCQM [59]	0.94	1.06	0.90	0.40	1.23	0.16	0.90	0.44
<b>CQM</b>	<b>0.79</b>	<b>0.83</b>	<b>0.90</b>	<b>0.39</b>	<b>1.22</b>	<b>-0.66</b>	<b>0.92</b>	<b>0.40</b>

To better understand the cumulative quality, Figure 9 shows the MOSs and the predicted scores by the CQM model corresponding to the adaptation result in Figure 3. We can see that the predicted scores closely follow to the MOSs. In addition, the cumulative quality fluctuates strongly during the session. This means that evaluating the overall quality at the end of a streaming session is obviously not enough to fully understand the quality of the video streaming service. So, cumulative quality over time is of crucial importance in adaptive streaming.

#### 5.4 Model Comparison in Overall Quality Prediction

In this subsection, the focus is on the performance of the models for overall quality prediction. The obtained results are shown in Table 7. Note that the SQI, KSQI, and Eswara’s models are not evaluated in this section because of the lack of input data (i.e., VMAF values of segments). We can see that the behaviors of the models are similar to those obtained in cumulative quality prediction (as presented in Table 6). In particular, for both of the *VL04* and *VL13* sets, the performance of the Rehman’s model is lowest (i.e.,  $PCC \leq 0.36$  and  $RMSE \geq 0.87$ ) while the performance of the CQM model is highest (i.e.,  $PCC \geq 0.90$  and  $RMSE \leq 0.40$ ).

Interestingly, although the CQM model does not directly include the impacts of initial delay and interruptions, it still performs very well for both the test sets containing these two factors. This is because their impacts are actually counted in the window quality model. Therefore, thanks to taking into account the impacts of various factors, namely quality variations (i.e., both long-term and short-term changes), recency, initial delay, and interruption, the proposed model could perform well and outperform the reference models for overall quality prediction.

In particular, with the *VL04* set including short sessions, Tran’s and CQM models achieve the highest performance that is slightly higher than those of the preCQM and P.1203 models. Specifically, the PCC and RMSE values are respectively 0.90 and 0.39 for Tran’s and CQM models, 0.90 and 0.40 for the preCQM model, and 0.88 and 0.42 for the P.1203 model. Meanwhile, the performances of the other models are quite low (i.e.,  $PCC \leq 0.80$  and  $RMSE \geq 0.87$ ).

For the *VL13* set with long sessions, the similar conclusions can also be drawn. Particularly, the performances of Tran’s, P.1203, preCQM, and CQM models are markedly higher compared with the others (i.e.,  $PCC \geq 0.90$  and  $RMSE \leq 0.46$ ). Among these four models, Tran’s model has the lowest performance (i.e.,  $PCC = 0.90$  and  $RMSE = 0.46$ ) while the P.1203 and CQM models achieve the highest performance (i.e.,  $PCC = 0.92$  and  $RMSE = 0.40$ ). The reason may be that Tran’s model

was originally designed for short sessions of 1 minute. Meanwhile, the others were built for various lengths (i.e., 1–5 minutes for the P.1203 model and 1–6 minutes for the preCQM and CQM models).

### 5.5 Remarks

Based on the above results and discussions, some remarks on cumulative quality prediction can be summarized as follows.

- Regarding the impacts of factors, the recency and quality variations (i.e., long-term changes) were found to have significant effects on the cumulative quality of streaming sessions. Meanwhile, the influence of the primacy can be neglected.
- To reflect the impacts of the recency and quality variations, the four window quality statistics, namely the last, average, maximum, and minimum window quality, were found to be essential in a cumulative quality model. Especially, it was found that the human cumulative quality perception mainly depends on the minimum and last window quality, which is in agreement with the peak-end rule. In comparison with the maximum window quality, the minimum window quality has more substantial impact. This result suggests that, compared to the best comfort, users tend to pay more attention to the worst discomfort.
- With respect to the window quality model, most of the investigated models performed very well. The highest performance was achieved when using the window quality model of Tran's.
- In comparison to using the same window size, it is more effective to employ different window sizes for different window quality statistics. Also, it was suggested to use the window sizes of 50 and 60 in cumulative quality models.
- As for the three pooling modes proposed in this article, they all derived quite high prediction performances. Especially, the Selected mode, which achieved the highest performance, was found to be both effective and efficient in cumulative quality prediction.
- In regard to the computation complexity of the proposed model, it mainly depends on the number of different used window sizes. But this can be effectively supported by parallel processing. For both serial and parallel processing, all the pooling modes in the proposed model take less than 1 ms to update the cumulative quality after every segment. Therefore the proposed model (with any mode) is applicable to real-time quality monitoring.
- Based on the experiment results, the CQM model was found to be very effective and outperform the 10 reference models for both cumulative and overall quality predictions.

## 6 CONCLUSIONS AND FUTURE WORK

In this article, we have presented a model for predicting the cumulative quality of adaptive video streaming. The proposed model was developed based on the concept of a “sliding window” over a streaming session, where each window is characterized by a quality value.

First, a subjective test was specifically designed and conducted for measuring the cumulative quality. Second, through statistical analysis, it was found that the impacts of the quality variations and recency are significant. We integrated the significant key components, namely, the last window quality, the average window quality, the minimum window quality, and the maximum window quality, into a new cumulative quality model CQM, which is able to accurately predict the cumulative quality of streaming sessions. The advantage of the proposed CQM model is its simplicity, while being inline with other well known effects from literature, namely, the applicability of simple temporal pooling plus the peak-end rule.

The CQM model was compared with ten existing models, where it could outperform the other models in predicting both the cumulative and overall quality. Moreover, the proposed model is

applicable to real-time quality monitoring thanks to its low computation complexity. This feature is especially important for cost-effective evaluation of streaming technologies, e.g., for real-time quality monitoring of video streams. In the future, the model will be used to assess the quality of different adaptive streaming techniques. Also, we will develop novel quality adaptation strategies, which are based on the CQM model.

## REFERENCES

- [1] Anne Aaron, Zhi Li, Megha Manohara, Jan De Cock, and David Ronca. 2015. Per-title encode optimization. Retrieved from February 1, 2018 from <https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2>.
- [2] C. G. Bampis, Z. Li, I. Katsavounidis, and A. C. Bovik. 2018. Recurrent and dynamic models for predicting streaming video quality of experience. *IEEE Trans. Image Process.* 27, 7 (Jul. 2018), 3316–3331. DOI: <https://doi.org/10.1109/TIP.2018.2815842>
- [3] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik. 2017. Study of temporal effects on subjective video quality of experience. *IEEE Trans. Image Process.* 26, 11 (Nov. 2017), 5217–5231. DOI: <https://doi.org/10.1109/TIP.2017.2729891>
- [4] N. Barman and M. G. Martini. 2019. QoE modeling for HTTP adaptive video streaming—A survey and open challenges. *IEEE Access* 7 (Mar. 2019), 30831–30859.
- [5] A. Bentaleb, B. Taani, A. C. Begen, C. Timmerer, and R. Zimmermann. 2019. A survey on bitrate adaptation schemes for streaming media over HTTP. *IEEE Commun. Surv. Tutor.* 21, 1 (Firstquarter 2019), 562–585.
- [6] Chao Chen, Lark Kwon Choi, Gustavo De Veciana, Constantine Caramanis, Robert W. Heath, and Alan C. Bovik. 2014. Modeling the time-varying subjective quality of HTTP video streams with rate adaptations. *IEEE Trans. Image Process.* 23, 5 (May 2014), 2206–2221.
- [7] Giuseppe Cofano, Luca De Cicco, Thomas Zinner, Anh Nguyen-Ngoc, Phuoc Tran-Gia, and Saverio Mascolo. 2017. Design and performance evaluation of network-assisted control strategies for HTTP adaptive streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 13, 3s, Article 42 (Jun. 2017), 24 pages. DOI: <https://doi.org/10.1145/3092836>
- [8] Tom Dietterich. 1995. Overfitting and undercomputing in machine learning. *Comput. Surv.* 27, 3 (Sept. 1995), 326–327. DOI: <https://doi.org/10.1145/212094.212114>
- [9] Z. Duanmu, W. Liu, D. Chen, Z. Li, Z. Wang, Y. Wang, and W. Gao. 2019. A knowledge-driven quality-of-experience model for adaptive streaming videos. arXiv:1911.07944 <https://arxiv.org/abs/1911.07944>.
- [10] Z. Duanmu, W. Liu, D. Chen, Z. Li, Z. Wang, Y. Wang, and W. Gao. 2019. A knowledge-driven quality-of-experience model for adaptive streaming videos. Retrieved June 1, 2020 from <https://github.com/zduanmu/ksqi>.
- [11] Z. Duanmu, K. Ma, and Z. Wang. 2018. Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach. *IEEE Trans. Image Process.* 27, 12 (Dec. 2018), 6135–6146. DOI: <https://doi.org/10.1109/TIP.2018.2855403>
- [12] Z. Duanmu, A. Rehman, and Z. Wang. 2018. A quality-of-experience database for adaptive video streaming. *IEEE Trans. Broadcast.* 64, 2 (Jun. 2018), 474–487. DOI: <https://doi.org/10.1109/TBC.2018.2822870>
- [13] Z. Duanmu, K. Zeng, K. Ma, A. Rehman, and Z. Wang. 2017. A quality-of-experience index for streaming video. *IEEE J. Select. Top. Sign. Process.* 11, 1 (Feb. 2017), 154–166. DOI: <https://doi.org/10.1109/JSTSP.2016.2608329>
- [14] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. 2019. Streaming video QoE modeling and prediction: A long short-term memory approach. Retrieved June 1, 2020 from [https://github.com/lfovialstm\\_qoe](https://github.com/lfovialstm_qoe).
- [15] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. 2020. Streaming video QoE modeling and prediction: A long short-term memory approach. *IEEE Trans. Circ. Syst. Vid. Technol.* 30, 3 (Mar. 2020), 661–673.
- [16] N. Eswara, K. Manasa, A. Kommineni, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. 2018. A continuous qoe evaluation framework for video streaming over HTTP. *IEEE Trans. Circ. Syst. Vid. Technol.* 28, 11 (Nov. 2018), 3236–3250. DOI: <https://doi.org/10.1109/TCSVT.2017.2742601>
- [17] D. Ghadiyaram, J. Pan, and A. C. Bovik. 2018. Learning a continuous-time streaming video QoE model. *IEEE Trans. Image Process.* 27, 5 (May 2018), 2257–2271.
- [18] D. Ghadiyaram, J. Pan, and A. C. Bovik. 2019. A subjective and objective study of stalling events in mobile streaming videos. *IEEE Trans. Circ. Syst. Vid. Technol.* 29, 1 (Jan. 2019), 183–197.
- [19] Zhili Guo, Yao Wang, and Xiaoping Zhu. 2015. Assessing the visual effect of non-periodic temporal variation of quantization stepsize in compressed video. In *Proceedings of the IEEE International Conference on Image Processing (ICIP'15)*. 3121–3125.
- [20] Tobias Hößfeld, Raimund Schatz, Ernst Biersack, and Louis Plissonneau. 2013. Internet video delivery in YouTube: From traffic measurements to quality of experience. *Data Traffic Monitor. Anal.* 7754 (2013), 264–301.

- [21] Tobias Hoßfeld, Michael Seufert, Christian Sieber, and Thomas Zinner. 2014. 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'14)*. 111–116.
- [22] T. Huang, C. Zhou, X. Yao, R. Zhang, C. Wu, and L. Sun. 2020. Quality-aware neural adaptive video streaming with lifelong imitation learning. *IEEE Journal on Selected Areas in Communications* 38, 10 (2020), 2324–2342. DOI: [10.1109/JSAC.2020.300363](https://doi.org/10.1109/JSAC.2020.300363)
- [23] P. Juluri, V. Tamarapalli, and D. Medhi. 2016. Measurement of quality of experience of video-on-demand services: A survey. *IEEE Commun. Surv. Tutor.* 18, 1 (Feb. 2016), 401–418.
- [24] Daniel Kahneman, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier. 1993. When more pain is preferred to less: Adding a better end. *Psychol. Sci.* 4, 6 (Nov. 1993), 401–405.
- [25] Friedemann Köster, Gabriel Mittag, and Sebastian Möller. 2017. Modeling the overall quality of experience on the basis of underlying quality dimensions. In *Proceedings of the 9th International Conference Quality Multimedia Experience*. 1–6.
- [26] Patrick Le Callet, Sebastian Möller, and Andrew Perkis (eds.). 2013. *Qualinet White Paper on Definitions of Quality of Experience*. Technical Report. Version 1.2.
- [27] Zhi Li, Christos Bampis, Julie Novak, Anne Aaron, Kyle Swanson, Anush Moorthy, and J. D. Cock. 2018. VMAF: The journey continues. Retrieved June 1, 2020 from <https://netflixtechblog.com/vmaf-the-journey-continues-44b51ee9ed12>.
- [28] Zhi Li, Christos Bampis, Julie Novak, Anne Aaron, Kyle Swanson, Anush Moorthy, and J. D. Cock. 2019. VMAF—Video multi-method assessment fusion. Retrieved June 1, 2020 from <https://github.com/Netflix/vmaf>.
- [29] David Lindegren, Werner Robitza, Marie-Neige Garcia, Steve Göring, Alexander Raake, Peter List, Bernhard Feiten, Ulf Wüstenhagen, Jörgen Gustafsson, Gunnar Heikkilä, Junaid Shaikh, and Simon Broom. 2018. ITU-T Rec. P.1203 Standalone Implementation. Retrieved July 1, 2020 from <https://github.com/itu-p1203/itu-p1203/>.
- [30] Yao Liu, Sujit Dey, Fatih Ulupinar, Michael Luby, and Yinian Mao. 2015. Deriving and validating user experience model for DASH video streaming. *IEEE Trans. Broadcast.* 61, 4 (Dec. 2015), 651–665.
- [31] M. Hammad Mazhar and M. Zubair Shafiq. 2018. Real-time video quality of experience monitoring for HTTPS and QUIC. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'18)*. 1331–1339.
- [32] Christopher Müller, Stefan Lederer, and Christian Timmerer. 2012. An evaluation of dynamic adaptive streaming over HTTP in vehicular environments. In *Proceedings of the 4th Workshop on Mobile Video*. 37–42.
- [33] Hyunwoo Nam, Kyung-Hwa Kim, and Henning Schulzrinne. 2016. QoE matters more than QoS: Why people stop watching cat videos. In *Proceedings of the 35th Annual IEEE International Conference on Computer Communications*. 1–9.
- [34] Y. Ou, Y. Xue, and Y. Wang. 2014. Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions. *IEEE Trans. Image Process.* 23, 6 (Jun. 2014), 2473–2486.
- [35] Lloyd Peterson and Margaret Jean Peterson. 1959. Short-term retention of individual verbal items. *J. Exp. Psychol.* 58, 3 (Sep. 1959), 193–198.
- [36] Stefano Petrangeli, Jeroen Famaey, Maxim Claeys, Steven Latré, and Filip De Turck. 2015. QoE-driven rate adaptation heuristic for fair adaptive video streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 12, 2, Article 28 (Oct. 2015), 24 pages. DOI: <https://doi.org/10.1145/2818361>
- [37] Alexander Raake, Marie-Neige Garcia, Werner Robitza, Peter List, Steve Göring, and Bernhard Feiten. 2017. A bitstream-based, scalable video-quality model for HTTP adaptive streaming: ITU-T P.1203.1. In *Proceedings of the Ninth International Conference on Quality of Multimedia Experience (QoMEX'17)*. 1–6.
- [38] Recommendation ITU-R BT.500-13. 2012. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union* (2012).
- [39] Recommendation ITU-T P.1203.1. 2017. Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport—Video quality estimation module. *International Telecommunication Union* (2017).
- [40] Recommendation ITU-T P.1203.2. 2017. Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport—Audio quality estimation module. *International Telecommunication Union* (2017).
- [41] Recommendation ITU-T P.1203.3. 2017. Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport—Quality integration module. *International Telecommunication Union* (2017).
- [42] Recommendation ITU-T P.1401. 2012. Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models. *International Telecommunication Union* (2012).
- [43] Recommendation ITU-T P.880. 2004. Methods for objective and subjective assessment of quality: Continuous evaluation of time varying speech quality. *International Telecommunication Union* (2004).

- [44] Recommendation ITU-T P.913. 2014. Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment. *International Telecommunication Union* (2014).
- [45] A. Rehman and Z. Wang. 2013. Perceptual experience of time-varying video quality. In *Proceedings of the 2013 5th International Workshop on Quality of Multimedia Experience (QoMEX'13)*. 218–223.
- [46] Russell Revlin. 2012. *Cognition: Theory and Practice*. Macmillan.
- [47] Werner Robitza, Marie-Neige Garcia, and Alexander Raake. 2017. A modular HTTP adaptive streaming QoE model-Candidate for ITU-T P. 1203 (“P. NATS”). In *Proceedings of the 9th International Conference on Quality of Multimedia Experience (QoMEX'17)*. 1–6.
- [48] Werner Robitza, Steve Göring, Alexander Raake, David Lindgren, Gunnar Heikkilä, Jörgen Gustafsson, Peter List, Bernhard Feiten, Ulf Wüstenhagen, Marie-Neige Garcia, Kazuhisa Yamagishi, and Simon Broom. 2018. HTTP adaptive streaming QoE estimation with ITU-T Rec. P.1203—Open databases and software. In *Proceedings of the 9th ACM Multimedia Systems Conference*. 466–471. DOI: <https://doi.org/10.1145/3204949.3208124>
- [49] Demóstenes Zegarra Rodríguez, Renata Lopes Rosa, Eduardo Costa Alfaia, Julia Issy Abrahão, and Graça Bressan. 2016. Video quality metric for streaming service using DASH standard. *IEEE Trans. Broadcast.* 62, 3 (Sept. 2016), 628–639.
- [50] K. Seshadrinathan and A. C. Bovik. 2011. Temporal hysteresis model of time varying subjective video quality. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'11)*. 1153–1156. DOI: <https://doi.org/10.1109/ICASSP.2011.5946613>
- [51] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. 2010. Study of subjective and objective quality assessment of video. *IEEE Trans. Image Process.* 19, 6 (Jun. 2010), 1427–1441.
- [52] M. Seufert, Pedro Casas, Nikolas Wehner, Gang Li, and Li Kuang. 2019. Stream-based machine learning for real-time QoE analysis of encrypted video streaming traffic. In *Proceedings of the 3rd International Workshop on Quality of Experience Management (QoE-Management'19)*. 76–81.
- [53] Michael Seufert, Sebastian Egger, Martin Slanina, Thomas Zinner, Tobias Hofffeld, and Phuoc Tran-Gia. 2015. A survey on quality of experience of HTTP adaptive streaming. *IEEE Commun. Surv. Tutor.* 17, 1 (2015), 469–492.
- [54] M. Seufert, M. Slanina, S. Egger, and M. Kottkamp. 2013. “To pool or not to pool”: A comparison of temporal pooling methods for HTTP adaptive video streaming. In *Proceedings of the 5th International Conference Quality Multimedia Experience*. 52–57.
- [55] Kamal Deep Singh, Yassine Hadjadj-Aoul, and Gerardo Rubino. 2012. Quality of experience estimation for adaptive HTTP/TCP video streaming using H. 264/AVC. In *Proceedings of the 2012 IEEE Consumer Communications and Networking Conference (CCNC'12)*. 127–131.
- [56] M. Takagi, H. Fujii, and A. Shimizu. 2014. Optimized spatial and temporal resolution based on subjective quality estimation without encoding. In *Proceedings of the 2014 IEEE Visual Communications and Image Processing Conference*. 33–36.
- [57] Samira Tavakoli, Sebastian Egger, Michael Seufert, Raimund Schatz, Kjell Brunnström, and Narciso García. 2016. Perceptual quality of HTTP adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowd-sourced subjective studies. *IEEE J. Select. Areas Commun.* 34, 8 (Aug. 2016), 2141–2153.
- [58] Truong Cong Thang, Hung T. Le, Hoc X Nguyen, Anh T. Pham, Jung Won Kang, and Yong Man Ro. 2013. Adaptive video streaming over HTTP with dynamic resource estimation. *J. Commun. Netw.* 15, 6 (Dec. 2013), 635–644.
- [59] H. T. T. Tran, Nam Pham Ngoc, Tobias Hofffeld, and Truong Cong Thang. 2018. A cumulative quality model for HTTP adaptive streaming. In *Proceedings of the 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX'18)*. 1–6.
- [60] H. T. T. Tran, Nam Pham Ngoc, Yong Ju Jung, Anh T. Pham, and Truong Cong Thang. 2017. A histogram-based quality model for HTTP adaptive streaming. *IEICE Trans. Fundam. Electr. Commun. Comput. Sci.* E100.A, 2 (Feb. 2017), 555–564.
- [61] H. T. T. Tran, N. P. Ngoc, A. T. Pham, and T. C. Thang. 2016. A multi-factor QoE model for adaptive streaming over mobile networks. In *Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps'16)*. 1–6.
- [62] Huyen T. T. Tran, Nam Pham Ngoc, and Truong Cong Thang. 2020. A study on impacts of multiple factors on video quality of experience. arxiv:2006.12697. Retrieved from <https://arxiv.org/abs/2006.12697>.
- [63] J. De Vriendt, D. De Vleeschauwer, and D. Robinson. 2013. Model for estimating QoE of video delivered using HTTP adaptive streaming. In *Proceedings of the IFIP/IEEE International Symposium on Integrated Network Management (IM'13)*. 1288–1293.
- [64] Chen Wang, Jianfeng Guan, Tongtong Feng, Neng Zhang, and Tengfei Cao. 2019. BitLat: Bitrate-adaptivity and latency-awareness algorithm for live video streaming. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2642–2646. DOI: <https://doi.org/10.1145/3343031.3356069>

- [65] S. Wassermann, M. Seufert, P. Casas, L. Gang, and K. Li. 2019. Let me decrypt your beauty: Real-time prediction of video resolution and bitrate for encrypted video streaming. In *Proceedings of the Network Traffic Measurement and Analysis Conference (TMA'19)*. 199–200.
- [66] M. Xu, C. Li, Z. Chen, Z. Wang, and Z. Guan. 2019. Assessing visual quality of omnidirectional videos. *IEEE Trans. Circ. Syst. Vid. Technol.* 29, 12 (Dec. 2019), 3516–3530.
- [67] K. Yamagishi and T. Hayashi. 2017. Parametric quality-estimation model for adaptive-bitrate-streaming services. *IEEE Trans. Multimedia* 19, 7 (Jul. 2017), 1545–1557.
- [68] Hema Kumar Yarnagula, Parikshit Juluri, Sheyda Kiani Mehr, Venkatesh Tamarapalli, and Deep Medhi. 2019. QoE for mobile clients with segment-aware rate adaptation algorithm (SARA) for DASH video streaming. *ACM Trans. Multimedia Comput. Commun. Appl.* 15, 2, Article 36 (Jun. 2019), 23 pages. DOI: <https://doi.org/10.1145/3311749>
- [69] Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. *ACM SIGCOMM Comput. Commun. Rev.* 45, 4 (Aug. 2015), 325–338.
- [70] L. Yu, T. Tillo, and J. Xiao. 2017. QoE-driven dynamic adaptive video streaming strategy with future information. *IEEE Trans. Broadcast.* 63, 3 (Sept. 2017), 523–534.
- [71] T. Zhao, Q. Liu, and C. W. Chen. 2017. QoE in video transmission: A user experience-driven strategy. *IEEE Commun. Surv. Tutor.* 19, 1 (Firstquarter 2017), 285–302.