Quitting Ratio-Based Bitrate Ladder Selection Mechanism for Adaptive Bitrate Video Streaming

Pierre Lebreton, *Member, IEEE*, and Kazuhisa Yamagishi

***Abstract*—To improve users' experience and decrease their likelihood of quitting watching videos, this paper addresses the question of how to encode the videos used in adaptive bitrate (ABR) video streaming. When addressing ABR video streaming, a lot of effort has been put into developing ABR control schemes. However, ways to appropriately encode videos also need to be defined. Unlike previous approaches that focus on coding quality, this paper considers the *user quitting ratio*. The *user quitting ratio* is the percentage of users still watching videos at a given time and enables us to address the consequences of quality and stimulus duration on the decision of a user to quit. Considering the value of the *user quitting ratio*, this paper describes a method that uses content analysis, as well as a network's historical throughput data, to define how video should be encoded to decrease the likelihood of users quitting watching. Unlike previous approaches, the method is independent of the ABR control scheme used by the video player, and the selected ladders perform equivalently across different players with different behaviors. Results of experiments based on real-world network traces demonstrate the usefulness of the proposed method.**

***Index Terms*—Adaptive bitrate video streaming, bitrate ladder, quality of experience, engagement, user quitting ratio.**

I. Introduction

THE usage of video streaming, one of the major applications on the Internet, has drastically increased as users are now consuming more and more videos on video on demand (VOD) services. Considering the large and continuous increase in the amount of content that is streamed, it becomes more and more important to encode video optimally to prevent wasting bandwidth or storage space on servers while still ensuring increased user engagement. User engagement is vitally important for service providers because metrics such as quitting ratio and viewing time have a direct link with their income, so these metrics have become key indicators for service providers. Therefore, these metrics need to be taken into account while defining coding conditions in order to lengthen viewing time and lower the quitting ratio.

Manuscript received 19 August 2022; revised 13 December 2022; accepted 11 January 2023. Date of publication 18 January 2023; date of current version 12 December 2023. The Associate Editor coordinating the review of this manuscript and approving it for publication was Dr. Zhi Wang. *(Corresponding author: Pierre Lebreton.)*

The authors are with NTT Network Service Systems Laboratories, NTT Corporation, Tokyo 180-8585, Japan (e-mail: lebreton.pierre.mz@hco.ntt.co.jp; kazuhisa.yamagishi.vf@hco.ntt.co.jp).

Digital Object Identifier 10.1109/TMM.2023.3237168

A. Adaptive Bitrate Video Streaming

Considering that the Internet connectivity of each user is different and varies over time, adaptive bitrate (ABR) video streaming is needed and is a key feature of modern video streaming services. With ABR video streaming, videos are stored on servers at various quality levels, which are called representations, and correspond to different bitrate value requirements. The set of all different bitrate values for each representation is referred to as a bitrate ladder. Each quality level uses a different encoding configuration such as different resolutions, frame rate, or quantization values, resulting in different bitrate requirements for each quality level. Then, as throughput varies, video players dynamically adjust the quality of the video played by selecting a quality level among the available representations that have an appropriate bitrate requirement considering the constraint given by the network.

The overall experience of the users then depends on both how the player behaves when requesting parts of the video (called chunks) and how the videos were encoded. If both aspects are properly handled, stalling will not occur and the highest possible video quality will be delivered. Therefore, to improve user experience, two ways can be considered: improving the player behavior, or improving the encoding of the videos. Regarding the player behavior (ABR control schemes), different schemes were compared by Yan et al. [1].

B. Challenges

In this work, the particular scenario of Video on Demand (VOD) is considered, and the problems to be addressed are how to encode videos and how this encoding affects the ability of ABR control schemes to deliver high quality video without stalling. Defining how to encode videos can be challenging as numerous parameters depend on contents that need to be tuned, and since the network performance varies across users, each user has different needs for how the videos should be encoded. Therefore, research should be performed on how to encode videos considering content properties, without needing to store too many representations of the same video on servers so as to preserve storage space. Finally, this should be done with the goal of reaching high user engagement.

C. Contributions

Although previous work has defined methods to identify the bitrate ladder on the basis of quality-related features, the relationship between quality and user's desire to use the services is not obvious. Therefore, in this work, a bitrate ladder estimation method is proposed that is based on the likelihood of user quitting [2].

Then, another contribution of this work is to take into account knowledge of network historical throughput data to define an ABR-control agnostic bitrate ladder estimation method. Indeed, while previous network dependent bitrate ladder estimation techniques have focused on a predefined ABR-control algorithm, decorrelating bitrate ladder and ABR control mechanism is important to allow both technologies to be improved independently. By doing so, the goal of the bitrate ladder estimation is then to provide the ABR control mechanism with different options that matches the network conditions, and let the ABR control mechanism select wisely what quality shall be used to deliver high-quality services. A large variety of ABR control mechanisms exists and is used across various video players, a contribution of this paper will then be studying the consistency of the performance of the selected bitrate ladder across various ABR control mechanisms.

Finally, the last contribution of the paper lies in its evaluation as the performance of the selected bitrate ladder are evaluated using quality and quitting estimation models that account for both coding degradation stalling events in a joined manner [2], [3].

D. Structure

The rest of this paper is organized as follows. Section II describes related work, and Section III introduces the proposed method for bitrate ladder estimation. Section IV describes the experimental environment to validate the results that are presented in Section V. Finally, Section VI discusses the results, and Section VII concludes this paper.

II. State of the Art

When encoding video for ABR video streaming, one of the first proposed approaches is to use pre-defined coding recipes. These can be found as tables provided by video streaming platforms [4], [5], [6] and give information on what bitrate values should be used to encode videos depending on the resolution, the frame rate, the color dynamic range (high dynamic range, HDR, vs. standard dynamic range, SDR), or the codec used to encode the videos. However, one challenge with this type of approach is that not all contents are equally difficult to encode and some contents require higher bitrates than others to achieve high quality. To address this challenge, a first approach is to cluster content into different classes of coding complexity (for example, easy, moderate, or difficult to encode), and then define a coding recipe for each of these categories. However, although this approach increases the coding efficiency more than a single pre-defined coding recipe, it does not address the core of the problem as each content has different coding complexity and would still result in unnecessary usage of storage space or lower quality. Because of these inefficiencies, this paper addresses the problem of per-title video coding.

A. Per-Title Video Coding

Per-title video coding consists of defining a bitrate ladder on a per-content basis. To achieve this, an approach based on multiple trial encoding has been proposed [7], [8]. In this framework, videos are encoded in different resolutions, and various bitrate values are evaluated using the video multi-method assessment fusion (VMAF) video quality estimation algorithm [9]. On the basis of these different trials of encoding and respective quality estimation, rate-distortion curves for each resolution can be drawn and enable us to identify for each bitrate value what resolution should be used to provide the highest quality to the users. The curve that provides the highest quality achievable across all resolutions (or any other parameters) for any bitrate is the convex hull. By using this approach, the dimensionality of the search space for an optimal bitrate ladder is then reduced from a multidimensional problem that includes the resolution, bitrate, frame rate, codec, etc. to a single dimension: the bitrate. This is possible as the dependency between all other factors can be derived from the convex hull.

Then, a direct extension of the framework of Aaron et al. [7] is to perform the analysis on a per-scene basis to account for the diversity of coding complexity within videos [10].

However, three main challenges remain. The first is that this type of framework is highly computationally intensive as it relies on encoding videos multiple times at various resolutions and bitrate values. The second is that although this framework enables the decrease of the dimensionality of the search space in only setting bitrate values, the questions of how many steps should be used in the ladder and what quality should be selected for each of these steps are still left open. Finally, the last open challenge is the interpretability of the criteria used for defining the ladder.

B. Low Complexity Per-Title Video Encoding

Researchers have focused on a computationally optimized method to estimate the convex hull. Katsenou et al. [11] described a method to predict the convex hull. With this approach, videos are encoded using constant quantization parameters (QP), and then on the basis of a pre-trained support vector regression model, QPs values are predicted where rate-distortion curves of different resolutions intersect. When using this approach, trial coding of the videos is still required, but the amount is drastically reduced.

Aiming to reduce computational complexity even further, Ling et al. [12] described a method to estimate the video coding complexity by categorizing content into classes of coding complexity. A classification algorithm using random forest, and crafted spatial-, temporal-, contrast-, and chrominance-based features, videos are categorized into a discrete number of classes of video coding complexity.

Bhat et al. [13], [14] described a similar framework to predict which resolution should be used for a given content at a given

bitrate. Rate control, spatial, temporal, and encoder pre-analysis features are used in a machine learning model.

Aiming for low latency live TV content which requires low complexity algorithms, Menon et al. [15] describe an alternative method using a video complexity analyzer on the basis of a discrete cosine transform (DCT) coefficients. Computed features are used for predicting the resolution that should be used for a given bitrate.

Finally Mux [16] described a long short-term memory (LSTM) recurrent neural network with Inception V3 features [17] to predict the rate-distortion category for various resolutions. However, due to its commercial nature, little information is known about the algorithm.

Based on these methods, it is possible at a very low computational cost to know what resolution (among other parameters) should be used for a given bitrate value, but it is still not clear what actual bitrate values should be used in the services. Because of this limitation, the contribution of this paper will be to describe how network throughput historical data can be leveraged for bitrate value selection and increasing user engagement.

C. Network Dependent Ladders

The convex hull enables the highest quality achievable for a given bitrate to be identified, but it is still necessary to know what bitrate/quality values should be used. A first approach can be based on just noticeable difference (JND) [18]. With this approach, bitrate values are selected such that there is Δ_Q JND between each coding condition, and results in having coding conditions uniformly distributed on the quality domain. However, the size of the interval Δ_Q still needs to be defined.

To address this problem, coding conditions should account for the capability of the internet access of end users. These types of ladders will then be referred to network dependent ladders.

Chen et al. [19] used a probabilistic approach with a focus on saving storage space while ensuring a minimum quality for the users. This method is based on historical throughput data and user's device resolution statistical data, which enabled the estimation of empirical throughput distribution and empirical user's device resolution distribution, respectively. A bitrate ladder is then defined by considering the likelihood that a given resolution and throughput are available.

Without prior knowledge of users' device resolution statistics, Reznik et al. [20], [21] described an analytical approach using non-linear constrained optimization to estimate the bitrate ladder.

In Toni et al. [22], the bitrate ladder is estimated using an integer linear program. Unlike other previous work, the authors also took into account the experience of the users on a per-session basis (estimated using the NTIA VQM quality model [23], [24]) and the overall content delivery network (CDN) budget. Constraints on maximizing session quality, load of CDNs, minimum quality delivered to users are defined and used to select ladders using integer programming.

However, when addressing network-dependent ladders, the importance of the video player behavior has only been weakly studied. Previous network-dependent bitrate ladders estimation methods only considered ABR control schemes that select chunks directly under the available throughput. This is problematic as a lot of effort has been put into developing advanced ABR control schemes [1], [25] that are not addressed in existing network-dependent ladder estimation methods. These advanced ABR schemes can predict future throughput and maximize quality while also minimizing quality variation [1], [26]. Such sophisticated approaches are necessary as the quality of experience (QoE) [27] of users degrades when numerous quality changes occur [28]. Therefore, a sophisticated ABR control model is needed that can account for the perceptual quality of the entire viewing session and drive the selection of chunks on the basis of perceptual aspects.

Lebreton and Yamagishi [29] described a bitrate ladder selection mechanism that accounts for both perceptual video coding quality (measured through VMAF [9]) and historical throughput data while being independent of the ABR control scheme.

In this paper, previous work is extended in order to take into account the effect of quality on the likelihood of user quitting videos considering content properties such as coding complexity and duration of scenes while also providing a network-dependent ladder that does not make assumptions about the behavior of the video player.

D. Towards Measuring Engagement

Although recent research has already shown a shift from choosing ladders on the basis of peak signal-to-noise ratio (PSNR) [19] to more advanced quality evaluation metrics such as NITA VQM [22], [23] or VMAF [9], [29], the relationship between the bitrate ladder and whether users will want to use the service or quit watching video has still not been addressed.

Considering that the ultimate goal of a service provider when defining a bitrate ladder is to maximize the engagement of the users (frequently measured using viewing time or the abandonment ratio) with respect to its operational cost, the consequences of the decisions made when defining the bitrate ladder on the engagement need to be understood. To do this, significant work on engagement understanding and modeling is introduced in the following.

First, note that work has shown that the concept of "acceptability" contains different levels, and a quality level that is acceptable may not necessarily be enjoyable [30], [31]. This led to the study of the transition between what is acceptable quality and what is not. Previous studies [32], [33], [34] have shown that acceptability could be estimated on the basis of technical parameters such as bitrate, frame rate, and resolution with non-linear regression. This model was further extended [30], [35], and it was shown that acceptability could also be related to high-level quality estimation on the basis of NTIA VQM [23] or VMAF [9].

However, quality is not the only parameter that needs to be considered when considering acceptability. Indeed, previous work has shown that there is also a temporal dependency on the acceptability, and what is acceptable for a short period may not be acceptable after a long one [36], [37], [38], [39]. Therefore, this highlights the need to consider both quality and stimulus

duration when studying acceptability and then when defining a bitrate ladder.

Then, another significant body of work worth mentioning is the studies of the relationship between viewing time and quality impairments such as initial loading delay [40], [41], [42] and midway-though stalling [43], [44]. These studies showed that the odds of users quitting increase exponentially as they need to wait for the video to play. The increase in user quitting because of initial loading is an important point, as when QoE is considered, results have shown that initial loading does not have such a high impact on QoE [45]. Therefore, these results show that when the likelihood of quitting is considered, other factors need to be considered in addition to those in traditional QoE evaluation.

Finally, as for coding quality, users were more likely to abandon video when quality changes occurred than when quality was more constant [36], [46], which highlights the need for advanced ABR control mechanisms. Therefore, all these results show that bitrate ladders should also be defined by taking into account features that relate to quitting behavior.

Considering the complexity of the relationship between abandonment and session quality, existing user quitting prediction models will be considered when defining coding conditions of bitrate ladders. While a regression tree-based model has been proposed [47], [48], this work will be based on the user quitting ratio model developed by Lebreton and Yamagishi [2] as this model provides closed form equations that are suitable for defining a quitting-based bitrate ladder and will ultimately decrease the likelihood of users quitting the service.

E. Contributions

Summarizing the contribution of this work, this paper presents a per-title bitrate ladder estimation algorithm that takes into account a network's historical throughput data to optimally select different quality levels to be used in the ladder.

Unlike in previous approaches, the bitrate ladder mechanism is independent of the ABR control scheme used by the video player. This enables the wide diversity of player behavior to be supported. Then, whereas previous methods addressed the bitrate ladder selection process on the basis of quality measures, it is proposed to define the bitrate ladder based on the likelihood of user quitting (the user quitting ratio). Doing so adds further sophistication as it is necessary to account for not only perceptual quality of the encoded video but also the duration of the content.

Finally, this paper also differs from previous work in the evaluation methodology. While previous methods only evaluated the appropriateness of the selected ladder in terms of the PSNR [19], number of quality changes [22], or average QoE using quality evaluation metrics such as VQM [22], in this work the overall session quality experienced by users is measured using metrics designed to evaluate the overall experience of the users accounting for both the effect of coding quality, quality adaptation and stalling in a joined up manner to address user engagement [2], [49].

Furthermore, this work will also evaluate the sessions in terms of the likelihood of users quitting because of quality, which provides an even higher level of understanding of users' viewing experience [2].

TABLE I
RANGE OF CODING CONDITIONS ACROSS DATASETS

Resolution	Frame-rate	Video Codec	Audio Codec
144p - 2160p	15 - 60	H.265/HEVC, H.264/AVC	AAC-LC
Video Bitrate	Audio Bitrate	Num. coding conditions	Num. PVSs
100-15000 kbps	32-384 kbps	48	134

III. PROPOSED METHOD

This section describes the methodology to identify the bitrate ladder so as to minimize the *user quitting ratio* [2] of VOD services.

A. Background on User Quitting Ratio Prediction

Considering that the proposed approach aims to decrease the user quitting ratio, this section will first introduce the computational process of the user quitting ratio model that was published in [2]. The goal of this section is to describe the key concepts about this method, detail the scope for which it has been defined and validated, and illustrate the suitability of its usage for the bitrate ladder estimation process. Then, the following section will explain its application to define a bitrate ladder.

The user quitting ratio model was designed on the basis of intensive subjective testing that involved 5 experiments and 264 participants to analyze and model quitting behavior. In these experiments, large ranges of video and audio quality values were tested. Table I list the range of resolution, frame-rate, codec, and bitrate for which the quitting model have been tested and validated. Different pairings for audio and video quality were tested, enabling their individual contribution as well as their interactions to be evaluated, and resulting in 48 coding conditions, and a total of 134 processed video signal (PVS).

The general procedure for the quitting evaluation experiments was to let participants watch videos without a predefined task. Participants were shown videos in a randomized order and were instructed to watch these videos in 6-10 minutes sessions. Participants were told that they could stop watching a video whenever they desired but quitting should only be based on quality. Therefore, the results give the quitting ratio in terms of acceptability of quality with respect to time [2].

Considering that the experiments did not include a quality evaluation task, quality of videos is estimated using a no-reference parameter-based audio visual quality estimation model [3], [49]. The model takes the video bitrate, the resolution, the frame rate, the audio bitrate, and stalling position and duration information and can predict audiovisual quality both on a per-second basis as well overall session quality. Detailed model equations can be found in [3], [49].

Fig. 1 depicts examples of the temporal evolution of the quitting ratio. The user quitting ratio is found to increase as a function

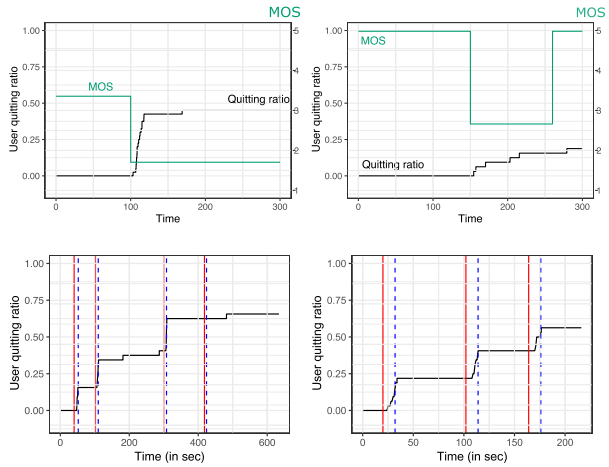


Fig. 1. Quitting likelihood as a function of time and MOS on two PVSs (top), and quitting likelihood in the presence of stalling (bottom). Stalling start and end marked respectively by a vertical continuous red line and a blue dashed line (data from [2]).

time with an increasing rate as quality (MOS) becomes low. Furthermore, stalling events were also identified as a major reason for user quitting.

On the basis of a thorough data investigation, an iterative model was introduced to address the temporal evolution of quitting probability. The model decomposes the viewing session into quality events. These events can either be the playback of a video at a constant quality level, which is referred to as a "segment of constant quality" in [2], or stalling events. On the basis of this decomposition of the viewing session into events, two sub-models were designed to address individually the effect of low coding quality and stalling events on quitting likelihood.

In terms of performance, cross-validation was applied on the five databases by performing training on some databases and validating on the others. When the evaluation is performed per-segment of constant quality, the proposed model shows an average Pearson correlation coefficient (PCC) of 0.938 and an average root mean square error (RMSE) of 0.0734 (7%). As for quitting from stalling, the model was validated using cross-validation across databases as well and showed an average PCC of 0.896 and an average RMSE of 0.109 (11%) on validation data, which both show high performance. For more details on training and validation performance, interested readers are invited to consider the original paper [2].

Considering these performance results, it is proposed to leverage the acquired knowledge on the relationship between quality and quitting to define bitrate ladders.

B. User Quitting Ratio-Based Ladder

Fig. 2 provides an overview of the proposed bitrate ladder estimation framework. The general flow of the framework is as follows: the video is split into scenes to address changes in coding complexity within a video (Section III-B1), identifies the convex hull to address each scene coding complexity (Section III-B2), converts the convex hull from quality domain to quitting likelihood (Section III-B3), use network historical throughput

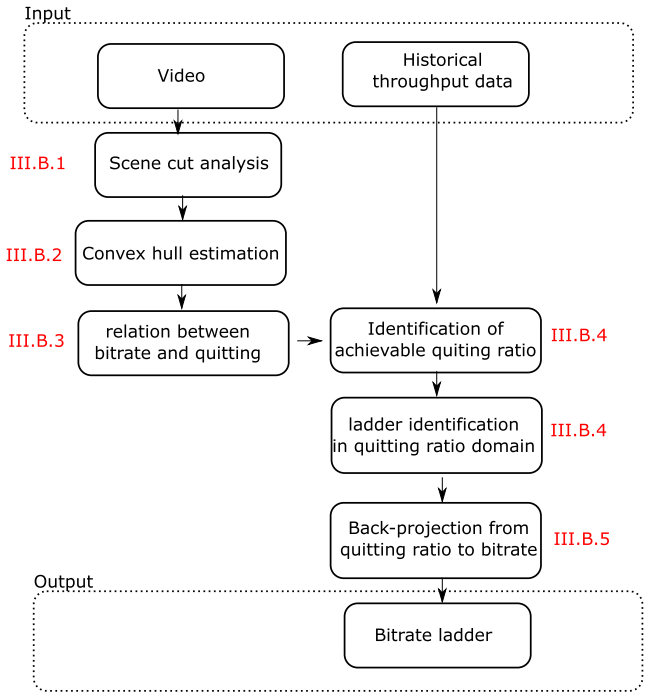


Fig. 2. Bitrate ladder estimation framework block diagram.

data to successively convert throughput into achievable quality and achievable quitting ratio. Identifies the steps in the bitrate ladder in the quitting likelihood domain (Section III-B4). Finally, back-project ladder steps from the quitting likelihood domain into bitrate values and leverage the convex hull to identify what resolution should be used with the identified bitrate values (Section III-B5).

The following explains the method in greater depth.

1) Decomposition Into Scenes: First, the proposed method analyzes the entire video to be encoded and decomposes it into scenes (see Fig. 3(a)). This decomposition has two main objectives. The first is to consider the fact that each scene of a video has a different coding complexity, so an optimal ladder should be defined on a per-scene basis [10]. The second is to consider the duration of each scene. Scene duration is an important aspect as there exists a temporal effect on the acceptability of a given quality level [2], [39]. Since quality is defined on a per-scene basis, scene duration should also be considered when defining its coding quality. Regarding the process of scene boundary detection on its own, various automatic scene boundary detection mechanisms have been proposed. In this work, a possible approach is the method proposed by Chen et al. [50], which uses self-supervised and supervised learning to identify scene cuts. However, other high-performing techniques could also be considered. The process of scene cut detection being not a contribution of this work will not be described in this paper. The interested reader can refer to the work of Chen et al. [50] to learn more about these types of algorithms.

2) Coding Complexity Analysis (Convex Hull Estimation): Once the entire video is divided into scenes, a bitrate ladder is defined on a per-scene basis. Fig. 3(b) describes the coding complexity analysis process that is based on [7]. With this approach,

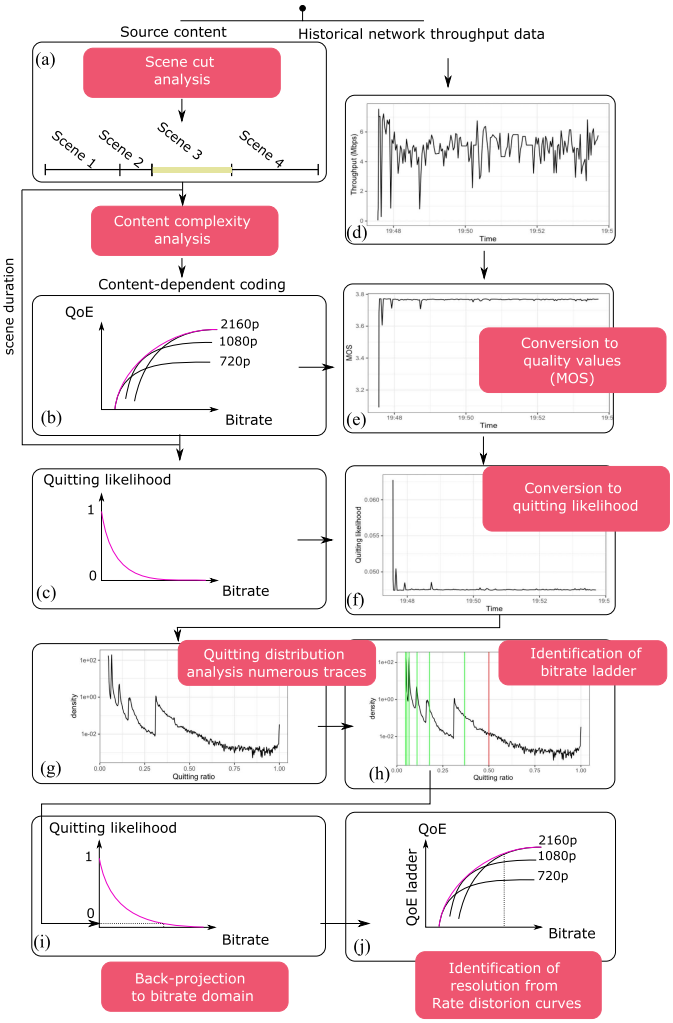


Fig. 3. Detailed overview on the bitrate ladder estimation framework.

the scene under consideration is encoded at various resolutions, and bitrate values enable the rate-distortion curves to be identified for different resolutions. In this work, videos are encoded using ffmpeg's x265 codec and constant QPs ranging from 12 to 48 in increments of 4, resulting in 10 different encodes per resolution. Doing the encodes using constant QP enables an approximate uniform sampling in terms of quality and the identification of characteristics of the rate-distortion curves. Then, the quality of the trial encodes is evaluated using a content-dependent quality estimation, VMAF [9] in this work. VMAF being a "full reference" metric, it compares the encoded video with that same video before encoding and will provide video quality estimates in the range of 0 to 100 (with 0 and 100 being the lowest and highest quality).

On the basis of the quality evaluation of the multiple trial encodes, the convex hull is defined. The convex hull is represented as a pink curve in Fig. 3(b) and represents the highest quality achievable across all resolutions for a given bitrate.

$$V_h = \frac{F_c}{1 + e^{-K_c \times log_{10} b_v + G_c}} \tag{1}$$

At the end of this process, the highest quality achievable across all resolutions for each bitrate value is obtained. A model to relate the bitrate and highest quality achievable across all resolutions (e.g., the convex hull) is defined by a sigmoid function as shown in (1). In this equation, F_c, G_c, K_c are model parameters. These parameters are obtained on a per-scene basis using non-linear regression by fitting data points on the relationship between the bitrate and highest quality values. Finally, V_h and b_v are respectively the convex hull and the video bitrate.

3) Quitting Likelihood Estimation: The next step consists of taking into account the relationship between quality estimates obtained using VMAF and likelihood of users quitting watching video. To do this, the quitting likelihood prediction model for low quality coding conditions introduced in Section III-A is used.

When defining coding conditions, it is only the coding-related model that is used and not the full model that also account for stalling events. This is motivated by the fact that during the encoding process, the ABR control mechanism is expected to prevent stalling events occurring as they are known to have a major impact on user's experience. Stalling is avoided by decreasing coding quality, so the impact of lower quality on user quitting behavior needs to be addressed.

The coding quality-related quitting model defines quitting as a negative exponential function defined in (2). In this equation, Q is the quitting likelihood after a duration d. λ_s is a parameter that accounts for the audio and video quality. Its expression can be found in (3). In this equation, M_v, M_a are the video and audio quality. ΔM_v enables quality adaptation to be handled and is the magnitude of the last quality change. Q_p is the percentage of users that have already quit when the constant quality condition begins. C_{1-6} are model parameters.

$$Q = 1 - e^{-\frac{d}{\lambda_s}} \tag{2}$$

$$\lambda_s = max\,(\epsilon, C_1 + C_2 \times M_v + C_3 \times M_a + C_4 \times M_v \times M_a + C_5 \times \Delta M_v + C_6 \times Q_p) \tag{3}$$

Equation (3) enables different factors to be accounted for including quality adaptation and the existing quitting ratio before the beginning of the segment. Since at the time of encoding videos, the effect of video quality on likelihood to complete watching videos (of duration d) is addressed, quality changes do not apply. Therefore, λ_s is altered into (4). Note that the case of constant coding quality was addressed during designing and is a supported case [2].

$$\lambda_s = max\,(\epsilon, C_1 + C_2 \times M_v + C_3 \times M_a + C_4 \times M_v \times M_a) \tag{4}$$

In this case, three parameters remain (d, M_v and M_a) and four model coefficients (C_{1-4}). The model coefficients C_{1-4} are obtained from Lebreton and Yamagishi [2], and d is the scene duration obtained from the scene cut analysis performed in the preliminary step. Then, M_v and M_a are the parameters to be defined so as to optimize the bitrate ladder in terms of quitting. Considering the video represents most of the data that needs to be transmitted, the following will focus on the estimation of the bitrate ladder for the video, and audio quality level will be

assumed to be at a high level ($M_a = 5$). Note that asymmetric audio and video quality were tested in [2], and is a supported scenario for the user quitting ratio prediction model.

To relate the video bitrate with the video quality M_v, the model of the convex hull defined in (1) is used. Considering that estimates of quality for the different trial encoding were based on VMAF, quality scores were provided in the range $[0, 100]$. However, considering that the user quitting prediction model defined in (2) that is used to predict quitting ratio requires video quality estimates in the range of $[1, 5]$, VMAF scores estimated using the convex hull are mapped in the range of $[1, 5]$ using a linear function as defined in (5). In this equation, $A_1 = 0.02827$ and $A_2 = 1.693$ are model coefficients that are constant across all scenes and all content. These parameters were trained with data from subjective experiments described in Lebreton and Yamagishi [2], and a linear regression is made to match scores obtained by VMAF with subjective quality scores obtained in the range of $[1, 5]$. This model shows a Pearson correlation of 0.91 and a root mean squared error (RMSE) of 0.44 on a 5-point scale (VMAF being run its version v0.6.1).

$$M_v = A_1 \times V_h + A_2 \tag{5}$$

On the basis of (1), 2 and 5, the relationship between video bitrate, scene duration, and quitting ratio can be estimated on a per-scene basis, enabling understanding of the relationship between bitrate and likelihood of users quitting early during the scene under investigation.

4) Leveraging Network Historical Data for Bitrate Values Selection: On the basis of the model of the likelihood of users quitting video as a function of the bitrate, the next step of the framework consists of using this knowledge for defining the bitrate values to be used in the bitrate ladder. Fig. 3(d) shows a time series of throughput values of one viewing session. These throughput values provide indications on the bandwidth that was available to the client. Available throughput is important information as it provides insight on what is the highest quality that can be achieved at a given time. Fig. 3(e) and (f) show respectively how the available throughput information translates to the user in terms of maximum achievable quality and the resulting likelihood of users quitting the video.

This analysis is then extended to a large number of sessions across numerous users, and the distribution of user quitting ratio values is depicted in Fig. 3(g).

$$r = \sum_t Q_t - argmin_{l \in L} |Q_t - l|^2 \tag{6}$$

On the basis of all estimates of quitting ratio values, the identification of the ladder is expressed as the problem of identifying appropriate quantization steps. The general idea behind this approach is that the bitrate ladders will be tailored to the bandwidth available to the users. If bitrate values higher than the available bandwidth are used, then the quality level cannot be reached or will lead to frequent stalling. If bitrate values lower than the available bandwidth are used, then constant lower quality, or frequently changing quality is delivered to the users, which also results in lower overall quality for the users. Therefore, using user bandwidth information when defining coding conditions enables the quality to be increased for the users by providing quality levels that they are more likely to use. Considering that available throughput varies over time, the quality and induced quitting ratio will change as well. However, as throughput varies, video quality can only be adjusted in a discrete manner that corresponds to the different steps of the bitrate ladder. Therefore, throughput values in between different steps of the bitrate ladders occur that correspond to missed opportunities for higher quality delivered to the users. A network-dependent ladder aims to decrease these missed opportunities for higher quality by defining steps in the ladder such that the differences between the available throughput and bitrate values in the ladder are as low as possible. This problem is then a matter of decreasing the quantization error of the available throughput into discrete values that correspond to the bitrate ladder. However, considering that video bitrate and quality have a logarithmic relationship, providing a bitrate ladder close to the available throughput may result in wasted resources as an increase in video bitrate may not result in a perceivable increase in quality. Therefore, instead of quantizing available throughput, achievable quality and achievable quitting ratio should be quantized.

On the basis of this principle, the ladder is then identified by applying the conversion from available throughput to potential quitting values using the equations previously introduced. Then, finding the ladder consists of identifying a number of N discrete values that can closely quantize continuous values taken by the quitting ratio estimates computed from the historical network throughput data. The value of N, the number of steps in the ladder, is a parameter of the framework and is to be defined by the service's operator. This number of steps can be defined on the basis of common practices of the operator. However, as will be shown in the results section, opportunities for saving storage space can be created by choosing different values of N.

Once the value of N is chosen, the process of identifying quantization steps can be formalized as in (6). In this equation, $L = l_1, l_2, \ldots, L_N$ corresponds to the N discrete steps of the ladder in the quitting domain. $argmin_{l \in L} |Q_t - l|^2$ provides the closest value l from the actual quitting measurement Q_t. Therefore, $Q_t - argmin_{l \in L} |Q_t - l|^2$ is the residual error between the quitting measurement Q_t and its quantized value l. Finally, quantization error is summed across all measurements over all sessions and all users providing the overall quantization error introduced by using a ladder with discrete steps. The goal of the ladder estimation is then to identify L that contains N discrete values, so as to minimize r. A possible approach to solve this problem is to use the k-means algorithm with $k = N$. Once k-means converged, the N values l from L corresponds to the center of the clusters identified by the k-means. An example of such classification results is shown in Fig. 3(h).

5) Recovering the Bitrate Ladder: Once the quantization levels of quitting ratio values are identified (L), corresponding bitrate values can be obtained by consecutively inverting (2), 5 and 1. This provides the required quality values to ensure the quitting values in L (see Fig. 3(i)), the corresponding quality scores, and the bitrate values (see Fig. 3(j)), respectively.

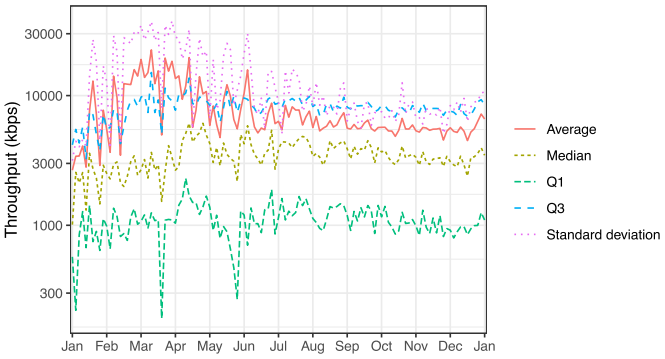


Fig. 4. Analysis of throughput measurements in 2019. Average and median are the average throughput and median throughput across all sessions on a 3-day interval basis. Q1 and Q3 refers respectively to the first and third quartile.

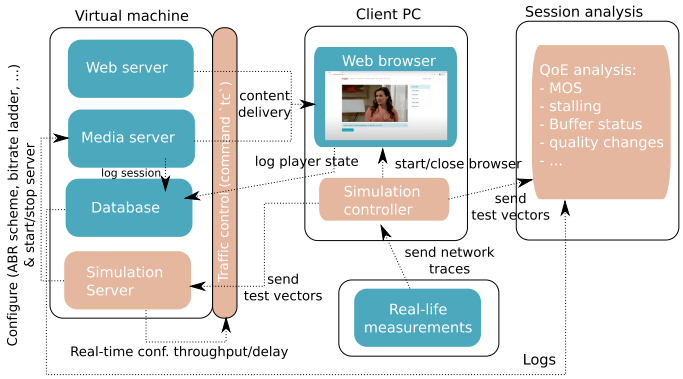


Fig. 5. Real-world network emulation framework. Items highlighted in blue are part of the Puffer framework [1], items in orange were designed for this work.

With bitrate values obtained, resolution values that should be used along these bitrate values are obtained from the coding complexity analysis performed during the identification of the convex hull. Resolution is then selected by choosing which resolution best matches the convex hull for the each identified bitrate value, which finally leads to the bitrate ladder.

IV. Real-World Network Emulation Framework and Settings

This section describes the approach used to evaluate the proposed bitrate ladder selection method.

A. Real-World Experimental Data

To perform experiments that reflect realistic scenarios, this work is based on data collected by the platform "Puffer" [1], [51]. Puffer is a video streaming service hosted by Stanford University that enables users living in the United States to watch live television programs of affiliated U.S. TV stations. The general principle of the platform is as follows. The platform developers receive TV using radio antennae, and affiliated TV programs are captured. These programs are encoded in real time using a high-performance server, and content is then re-broadcasted via the Puffer website. Users are then able to connect to the Puffer web application and watch TV using this service.

Puffer gives information on viewing data on a per-session basis on information such as throughput measurement and round trip time (RTT) that will be used to perform real-world like experiments. Finally, the platform also introduces four different ABR control schemes are compared. One is a buffer-based only method [52], two are control-theory oriented ABR control schemes (Model Predictive Control (MPC) and Robust MPC), and the last is a learning-based model called Fugu [1]. All the data is available to download and is updated daily [51]. In this work, data from January 2019 - January 2020 will be used.

Fig. 4 provides an overview on the distribution of throughput values for the considered time interval. The throughput across sessions varied largely from January-June 2019, whereas throughput values were more stable in the second half of the year.

B. Real-World Network Emulation Platform

Fig. 5 depicts the framework used in this study to evaluate the bitrate ladder selection processes. This framework is divided into a server and client side. On the server side, a web server, a media server, and a database can be found. The web server provides a web interface to the users so they can log into the service, select the content they desire to watch, and load a video player. Then, the media server ensures the delivery of the video content to the video player on the client side. Clients are based on the HTML5 "video" tag. To enable the ABR control schemes to be used, statistics on the video playback sessions are collected at different points of the transmission chain. On the server side, information such as delivery rate, round trip time, number of packet sent, chunk sizes, and SSIM of those chunks are collected. In addition to these statistics, player-related information is also collected thanks to self-reports from the players. Indeed, during the video playback, the video players periodically (or after specific events) send the status of the buffer filling rate, the cumulative stalling duration, acknowledgment of received packets, and respective timestamps. Then, by joining session identification codes, the server is able to precisely understand the status of every client. This information is then used for two different purposes. The first is to enable various ABR control mechanisms to be used. The second is to enable precise monitoring of the viewing sessions of real-world streaming sessions. The level of details in the logs enables in-depth analysis of the differences in quality between different ABR control mechanisms or different bitrate ladders and also aspects to be considered such as the quality of each chunks, the different quality adaptation that may occur, temporal evolution of buffer filling rate, frequency and duration of stalling events, etc.

In this work, different ABR control schemes are considered: a buffer-based only algorithm [52] and two MPC algorithms [26]. MPC-algorithms predict future throughput by using past throughput measurements and use them to optimize QoE while also minimizing quality changes. To perform experiments, a "simulation controller" software was implemented. It takes as input measurement from real-world services of the Puffer dataset [1], [51] as well as parameters on the configuration of the server, such as the bitrate ladder per content and the

ABR control scheme to be used. The software handles the configuration of the media server and will also start a service that controls the throughput and latency of the connection between the server and clients (using the command "tc" on Linux). This throughput control is based on historical log data [1], [51] and enables realistic network conditions to be reproduced. By using this approach, systematic testing can be applied with multiple bitrate ladder, ABR scheme, and network conditions. Finally, the streaming session is evaluated thanks to the logging performed on the server side that includes video bitrate, resolution, and frame rate as a function of time as well as stalling information (when and how long stalling occurs) enabling the video quality estimation model [49] and user quitting ratio model [2] to be used and measure the overall experience of the users both in terms of QoE [27] and likelihood of quitting at the end of the streaming session. It should be further stressed that in these evaluations, models account for both coding quality and stalling events. Therefore, although when defining coding conditions, quitting ratio model only based on coding quality, in the performance evaluation both coding quality and stalling events are considered.

Note that leaving the handling of quality adaptation logic to the server is different from commonly used video players that handle the quality adaptation logic themselves, and "dash.js" could have also been considered for conducting the experiments. However, the reason for using the Puffer framework as the basis of this work was that it is currently in use in real-world services and was then extensively tested. Other advantages of this framework are the availability of multiple ABR control mechanisms and the ability to switch from one ABR model to another by only updating configuration files, making it easy to validate across different ABR models. Finally, logging of playback statistics was already provided and automatically stored in a database allowing in-depth analysis of the user's viewing experience. Of course, all these features could have been implemented using dash.js, but the Puffer framework enables already all these features.

C. Experimental Settings

To address content with various coding complexities, 379 different source content videos were collected. These source content videos depicted various types of content including animation, talk shows, documentaries, and scenery and came from professionally recorded contents from internal and open databases [53]. Videos with a resolution of 1920×1080 were considered, and if the resolution was higher than 1920×1080, videos were downscaled to 1920×1080. Fig. 6 depicts an analysis of all source content videos in terms of spatial information (SI) and temporal information (TI) values [54] computed on a per-video basis across all scenes, as well the cumulative distribution function of compressibility defined in Robitza et al. [55]. The latter metric analyzing the area under the curve of rate-distortion curves (with quality values evaluated using VMAF). The estimated cumulative distribution function increasing linearly shows a rather uniform distribution of compressibility values for the source content that are used in the dataset. These

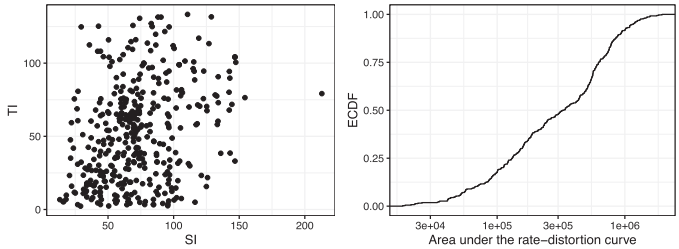


Fig. 6. Distribution of SI and TI values [54] (left). And estimated cumulative distribution function of compressibility measure [55] (right).

different analyses show that the selected source content covered a large variety of content types. All source content videos were then encoded using ffmpeg (version 3.4.8) in H.265 using the codec x265. Videos were encoded using the presets "slower" using constant rate factor values ranging from 12 to 48 in increments of 4. Quality of the different trial coding was evaluated using VMAF [9], and rate-distortion curves were identified. On the basis of these rate-distortion curves, source content videos were classified into 15 classes of coding complexity. The number of 15 classes was chosen to be a compromise between having a large number of content coding complexity-categories so as to test the performance of the bitrate ladder estimation against various coding complexity while still being low enough to minimize the time required for performing all experiments as one set of simulation runs is needed for each category.

Once classes of source reference circuits (SRCs) are identified, the following step is to estimate the bitrate ladder. For each class of coding complexity, the model described in (1) is used to relate quality and bitrate on a per-content category basis by training the model to use rate-distortion measurements across all the source content from the same class. The resulting fitted coefficients are then used to define bitrate ladder for content having different types of coding complexity.

As for the historical throughput data used to define the ladder, data from January 2020 is used as training data while data from January-December 2019 is used for validation. This approach was performed as temporal stability of performance needs to be validated over a long time span. Therefore, either January 2019 or January 2020 should be used for training and establishing the bitrate ladder leaving a 1-year time span for validation. January 2020 was selected, as Puffer grew in popularity during 2019, so January 2020 data offers a larger number of traces to be used for training than January 2019 data.

Finally, for each SRC category a ladder of six different quality levels is estimated. This number of quality levels was chosen as it corresponds to the current number of quality levels delivered by YouTube with six quality levels up to HD (144p, 240p, 360p, 480p, 720p, 1080p) and two more over HD (1440p and 2160p). The source content being limited to HD (for the sake of finding diverse high-quality SRCs), choosing six quality levels appeared a reasonable choice. However, in the results section, other numbers of quality levels are discussed.

Videos were then encoded using a two-pass constant bitrate encoding with the preset "slower," and a group of picture (GOP) size of two seconds.

V. Results

This section addresses the performance of the proposed bitrate ladder selection method. In this section, all results described were obtained on validation data. Performance figures are provided in terms of overall QoE [3] and quitting ratio [2] that are measured with models that account for coding quality and stalling events in a joined manner.

A. Compared Ladders

Seven different bitrate ladders are compared: k-means, constant VMAF-step, constant bitrate steps, JNDs, 1000 kbps steps, 100 kbps steps, and Apple static ladder. k-means refers to the proposed quitting-based bitrate ladder selection algorithm with $k = 6$. Then, two bitrate ladders derived from the proposed algorithm are proposed: constant VMAF-step and constant bitrate steps. These two ladders use the same maximum and minimum bitrates identified by the proposed k- means method that enable the identification of a proper range of bitrate values to be used. However, instead of designing intermediate steps on the basis of k-means, it is proposed to set the fourth intermediate bitrate values on the basis of a constant increase in quality or a constant increase in bitrate. Then, another ladder is referred to as JNDs, and define a bitrate ladder were each quality levels are separated by a constant quality step in terms of VMAF score. Compared to "constant VMAF-step," the ladders focus on defining conditions in terms of constant quality steps without the constraint of using the network historical data to set the range of quality that should be used. Such an approach is commonly used in the literature for per-title video coding [11], [12], [15], as work have focused on identifying the convex hull and define steps in the bitrate ladder by uniformly sample the quality space. Then, the two following baselines use constant increases of 1 Mbps between 1 to 12 Mbps or use 100 kbps between 100 kbps to 1 Mbps and 1-Mbps steps between 1 and 12 Mbps. These ladders (1000 kbps steps and 100 kbps steps) are content and network agnostic and enable the display of the added value of considering these features while defining a bitrate ladder and are referred to as static ladders. Finally, the last ladder is based on Apple's HLS recommendations [5] and is a static ladder. These ladders are then estimated for each class of content identified in Section IV-C.

B. Performance Across Ladder and ABR Control Schemes

Fig. 7 shows the cumulative distribution function of quitting ratio [2] and quality [3] for the different bitrate ladder and ABR control schemes. In the evaluation work described here, the quitting model is used in its complete form, identical to the model described in Lebreton and Yamagishi [2]. Therefore, the quitting ratio values are based on coding quality, stalling, initial loading, quality adaptation, and content duration providing a measure of the overall user's experience and its consequences in terms of abandonment. Considering that approximately 600 simulations were performed to cover a large variety of network conditions for each combination of bitrate ladder and ABR control schemes, quitting ratio values can vary from low to high as in some cases available throughput was not sufficient to transmit high-quality

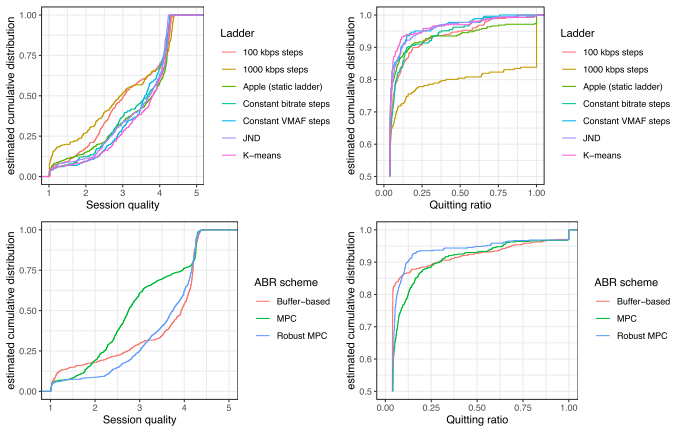


Fig. 7. Distribution of quitting ratio values across all source content. Results are shown per bitrate ladder and per ABR control schemes.

TABLE II
Compared Bitrate Ladders. 100 k, 1000 k, Cst.B, ASL, JND, Cst.V and K-Means Refers Respectively to the Ladders "100 Kbps Steps," "1000 Kbps Steps," "constant Bitrate Steps," the "apple Static Ladder," the JND-Based Bitrate Ladder, "constant VMAF Steps," and the Proposed Quitting-Based Approach

Name	Ladder type	Opt. Criteria	Bitrate ladder
100k	Static	-	100 kbps 12 Mbps, steps 100 kbps under 1 Mbps then steps of 1000 kbps
1000k	Static	-	1000 kbps 12 Mbps, steps of 1000 kbps
Cst.B	Bitrate Range is Optimized	Network	k-means (k=6) on quitting ratio values is used to provides five classes, and five corresponding bitrate values. Lowest and highest bitrates define range to be used, intermediate three values are linearly sampled in the bitrate domain between identified min/max.
ASL	Static	-	Ladder based on Apple's HLS recommendation [5]
JND	Quality range fully covered. Constant quality steps.	Content (based on Quality)	Bitrate ladder where each of the 8 coding conditions is distant by constant steps of 12 in the VMAF quality scale.
Cst.V	Bitrate Range is Optimized	Network & Content (based on Quality)	k-means (k=6) on quitting ratio values is used to provides five classes, and five corresponding bitrate values. Lowest and highest bitrates define range to be used, intermediate three values are linearly sampled in terms of quality (VMAF) between identified min/max quality.
k-means	Ladder is Fully Optimized	Network & Content (based on Quitting Ratio)	k-means (k=6) on quitting ratio values provides five classes, and five corresponding bitrate values.

videos. Therefore, cumulative distribution function of session quality and quitting ratio values are reported to describe overall performance across all tested conditions.

In this figure, it can be seen that content and network agnostic approaches (100 kbps steps and 1000 kbps steps) performed worse than other approaches even though they involved a much larger number of encodes than the other ladders. The statics ladders "constant bitrate steps" and "Apple static ladders" appear to be consistently outperformed by "Constant VMAF steps," which is outperformed by the proposed approach, "k- means". A non-parametric Kruskal-Wallis Test was applied and showed that there are significant differences in quitting ratios across different bitrate ladders (chi-squared $= 57.407$, p-value < 0.001). Table III reports the results of a multiple comparison test after a Kruskal-Wallis post-hoc test as well as average quitting and quality values per ladder. Results show that in terms of quitting ratio,

TABLE III
STATISTICAL DIFFERENCES BETWEEN DIFFERENT BITRATE LADDERS ON VALIDATION DATA IN TERMS OF USER QUITTING RATIO AND QUALITY. IN THE TABLE PAIRS (a, b) INDICATES WHETHER THERE EXISTS A SIGNIFICATIVE DIFFERENCE IN TERMS OF QUITTING RATIO (A), OR IN TERMS OF QUALITY (B). IN THIS PAIR, A SIGN "+" OR "-" INDICATE A SIGNIFICANT DIFFERENCE OR NOT. LADDER NAMING CONVENTIONS ARE CONSISTENT WITH TABLE II

	100k	1000k	Cst.B	ASL	JND	Cst.V	k-means
100k	(-,-)	(-,-)	(-,-)	(-,-)	(-,-)	(+,+)	(+,+)
1000k	(-,-)	(-,-)	(-,-)	(-,-)	(+,-)	(+,+)	(+,+)
Cst.B	(-,-)	(-,-)	(-,-)	(-,-)	(+,-)	(+,-)	(+,-)
ASL	(-,-)	(-,-)	(-,-)	(-,-)	(+,-)	(+,-)	(+,-)
JND	(-,-)	(+,-)	(+,-)	(+,-)	(-,-)	(-,-)	(-,-)
Cst.V	(+,+)	(+,+)	(+,-)	(+,-)	(-,-)	(-,-)	(-,-)
k-means	(+,+)	(+,+)	(+,-)	(+,-)	(-,-)	(-,-)	(-,-)
ladder	100k	1000k	Cst.B	ASL	JND	Cst.V	k-means
avg. Q.R.	0.09630	0.23470	0.09126	0.1012	0.074243	0.070336	0.067855
std. Q.R.	0.1519	0.3625	0.1398	0.19204	0.11977	0.1264	0.1046
avg. qual.	3.045	2.904	3.301	3.310	3.319	3.360	3.437
std. qual.	1.044	1.220	0.978	1.058	0.9609	0.8949	0.9250

TABLE IV
AVERAGE IMPROVEMENT IN TERMS OF USER QUITTING RATIO FROM CONSIDERING NETWORK AND QUALITY, AND QUITTING-RELATED FEATURES OVER STATIC LADDERS. NOTATION ARE CONSISTENT WITH TABLE II

Type	Improvement over static ladder	
	Quitting ratio	Quality
Ladder based on network statistics (Cst.B)	57.8%	6.7%
Ladder based on Quality (JND)	194%	7.5%
Ladder based on network & Quality (Cst.V)	204%	8.9%
Ladder based on network & Quitting (k-means)	212%	11.3%

ladders "100 k," "1000 k," "Constant bitrate steps," and "Apple static ladders," are statistically equivalent and are outperformed significantly by "k-means," and "Constant VMAF steps". These results show the importance of taking into account coding complexity information while defining a bitrate ladder. This is further stressed by Table IV, which reports the improvement of using network-based, quality-based, and quitting-based ladders over static ladders (100 k, 1000 k, Apple static ladder). Table IV also shows that tuning the range of bitrate values considered in the ladder on the basis of network statistics would enable the quitting ratio to be decreased by approximately 60%. Considering coding quality provides a great improvement, and result in decreasing quitting by half compared to content-agnostic ladders. Combining the use of network statistics and coding complexity allows decreasing quitting even further. Finally, optimizing intermediate steps in terms of quitting using the proposed $k - means$ method provided the highest performance among the tested conditions. It can also be observed that the range of performance improvement differs between the quitting and quality scales as quitting take into account duration of the viewing session ((2)).

Following up on this analysis, a Kruskal-Wallis Test was applied to compare the quitting ratio values across the different ABR control schemes. Results show that ladders performed significantly differently (chi-squared $= 134.07$, p-value < 0.001). A post-hoc analysis shows that there was a significant difference between the performance of the buffer-based algorithm and the MPC-based ABR control schemes. However, no significant difference was observed between the performances of MPC and Robust MPC ABR control schemes. These results show the importance of considering different ABR control schemes when evaluating a bitrate ladder selection method.

TABLE V
STATISTICAL DIFFERENCES BETWEEN LADDERS IN TERMS OF USER QUITTING RATIO ON A PER-ABR CONTROL SCHEME BASIS. IN THE TABLE (A,B,C) INDICATES THE SIGNIFICANCE OF DIFFERENCES BETWEEN BITRATE LADDERS, WITH "A" IS THE BUFFER-BASED, "B" IS THE MPC, AND "C" IS THE ROBUST MPC ALGORITHM. NOTATION ARE CONSISTENT WITH TABLE II

	100k	1000k	Cst.B	ASL	JND	Cst.V	k-means
100k	(-,-,-)	(-,-,-)	(+,-,-)	(+,-,-)	(-,-,-)	(+,+,-)	(+,-,-)
1000k	(-,-,-)	(-,-,-)	(+,+,-)	(+,+,-)	(-,+,-)	(+,+,-)	(+,+,-)
Cst.B	(-,-,-)	(-,-,-)	(-,-,-)	(-,-,-)	(-,+,-)	(+,+,-)	(+,-,-)
JND	(-,-,-)	(-,+,-)	(-,+,-)	(-,-,-)	(-,-,-)	(-,-,-)	(-,-,-)
ASL	(-,-,-)	(-,-,-)	(-,-,-)	(-,-,-)	(-,-,-)	(+,+,-)	(+,-,-)
Cst.V	(+,+,-)	(+,+,-)	(+,+,-)	(+,+,-)	(-,-,-)	(-,-,-)	(-,-,-)
k-means	(+,-,-)	(+,+,-)	(+,-,-)	(+,-,-)	(-,-,-)	(-,-,-)	(-,-,-)

TABLE VI
COMPARISON OF PERFORMANCE IN TERMS OF USER QUITTING RATIO ACROSS ABR-SCHEMES

	Buffer-based	MPC	Robust MPC
Avg.	0.1128	0.1229	0.09676
Var.	0.2208	0.2076	0.18713

Considering that the quitting ratio depends on both bitrate ladders and ABR control schemes, it is of interest to study the extent to which an ABR control schemes performance depends on the bitrate ladder. To this end, different ladders are compared for each ABR control scheme. Results are reported in Table V. From this analysis, it is interesting to note that depending on the ABR control scheme, differences in performance between ladders vary. Indeed, the control scheme "Robust MPC" is able to perform consistently across all ladders, whereas the buffer-based algorithm is the most sensitive to the bitrate ladder among the considered algorithms. The control scheme "MPC," on the other hand, is less sensitive to the bitrate ladder than the buffer-based algorithm, but differences in performance can still be observed. This is important as it shows there are two main approaches to improve the experience of the users and decrease the likelihood of quitting (the use of ladders that depends on network statistics, or the use of more sophisticated ABR control schemes), and the weakness of one of these two can be overcome by using a more optimized other.

Table VI provides information on the average and variance of the quitting ratio for all ABR control schemes across all ladders and streaming sessions. It can be seen that Robust MPC [26] provided the lowest average quitting ratio followed by the Buffer-Based method [52] and the MPC algorithm [26]. Note that the challenge of the dataset used in this work is that throughput can largely fluctuate as a function of time (see Fig. 4). Therefore, considering that the original MPC algorithm [26] uses harmonic means for throughput prediction, this resulted in lower performance. The algorithm Robust MPC accounting for throughput prediction error into the ABR control mechanism provides a more conservative chunk selection that outperforms MPC. The Buffer-based algorithm ranked second in terms of average quitting ratio across all ladders, which is consistent with the work of Yan [1], which showed that this buffer-based technique [52] already provides compelling results. Finally, in terms of variance of quitting ratio values, note that MPC approaches provided more stable results across all streaming sessions.

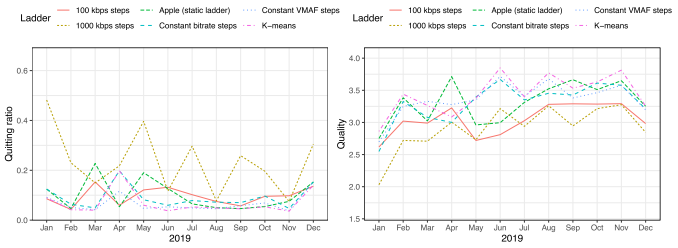


Fig. 8. Temporal evolution of user quitting ratio (left) and quality (right) as a function of time.

C. Temporal Analysis

Considering that throughput varies over time, the question of whether a network-dependent ladder can still perform well after a large period of time needs to be raised. To this aim, Fig. 8 depicts the temporal evolution of the average quitting ratio as a function of time. In this figure, the quitting ratio values across all ABR control schemes are averaged, and results show that the ladder "1000 kbps" performs consistently lower than the other approaches. The ladder "100 kbps," on the other hand, enables lower bitrate values to be used and higher performance in these scenarios. Moreover, the proposed "k- means" approach provided consistently lower quitting ratio values than other ladders for most months. April and May 2019 are exceptions where "100 kbps" and "Constant VMAF steps" outperformed "k-means." However, the confidence intervals are significantly larger for these two specific months than for the other months, and differences in performance are not statistically significant in these cases.

To validate the temporal stability of ladder performance, a non-parametric Kruskal-Wallis Test is applied to compare quitting ratio values per month for each ladder. Results show that there are no statistical differences in quitting ratio values across months for "k-means" (chi-squared = 14.537, p-value = 0.2047), "1000 kbps" (chi-squared = 12.662, p-value = 0.316), and "constant bitrate steps" (chi- squared = 12.175, p-value = 0.3506). However, "100 kbps" and "constant VMAF steps" showed significant differences across months (chi-squared = 20.963, p-value = 0.03376) and (chi-squared = 21.347, p-value = 0.02995), respectively. These results demonstrate that the proposed approach of defining a ladder on the basis of historical data can provide stable performance over time.

D. Dependency to the Number of Quality Levels

The results described in this paper are for a bitrate ladder composed of six quality levels. To test the effect of the number of quality levels in the bitrate ladder on the user quitting ratio, experiments were run with ladders ranging from three to six quality levels. Results show that going from six quality levels to five increased the quitting ratio by 1%. Going from six quality levels to four also resulted in a 1% increase in user quitting ratio. Finally, going from six quality levels to three resulted in a 3% increase in the user quitting ratio. A non-parametric Kruskal-Wallis Test was applied to test the significance of differences and shows that ladders performed statistically equivalently (chi-squared = 1.6572, p-value = 0.6465). A per-ABR control scheme analysis was also performed, and no statistically significant differences in terms of quitting ratio could be observed across ladders. This result is of interest as it shows that using network throughput statistics in addition to content coding complexity analysis enables performance to be maintained even with a low number of quality levels in the bitrate ladder. However, it should also be stressed that this result is obtained by performing evaluations in numerous sessions that cover a large diversity of temporal variation of throughput. Therefore, these results show that high quality service can be provided with a limited number of quality levels for the majority of users. On the other hand, note that among the large number of conditions, some cases still occur for which a higher number of quality levels in the ladder is still beneficial.

VI. Discussion

In this work, a bitrate ladder estimation method was proposed to decrease the user quitting ratio. Results have shown that there are different ways to improve the experience of the users. The analysis of performance across ladders and ABR control schemes has shown that simple ABR control schemes be greatly improved by using better-optimized ladders, and advanced ABR control schemes such as Robust MPC [26] are less sensitive to the choice of the bitrate ladder than simpler approaches [52].

One possible source of concern for a network-dependent ladder is that the selected bitrate ladder may only be valid for a limited period. Results based on the Puffer open dataset [1] have shown that stable performance over one year could be observed. Performance was also consistent across ABR control schemes. In a real-world scenario, availability of content needs to be maintained over multiple years. Considering the advancement of network access performance, the bitrate ladder may need to be re-estimated after multiple years. However, the proposed approach still has the benefit to define how to encode newly released and most popular contents, which may cover most of the load of the service provider. Therefore, ensuring that new and popular contents are properly encoded to improve user experience would provide high value to the users.

Results have also shown that stable performance could be obtained across a large variety of types of source contents, and thanks to the network and content-dependency, it was also possible to use a limited number of quality levels in the bitrate ladder to cover the need of most users. These results show that opportunities exist to save storage space by considering different types of encoding depending on network throughput and could be applied to tune the performance to specifically address different geographical regions.

It should be mentioned that the proposed approach in its current state is computationally intensive as it relies on a large number of trials encodes so as to define the bitrate ladders. In the current work, the main focus of the research has been towards taking into account the network's historical throughput data as well as choosing the ladder configuration based on the

user quitting ratio. Therefore, the most accurate content complexity estimation technique was used, but more computationally efficient approaches such as [11], [12], [13], [14], [15] could be considered as well.

Finally, future work will consider using the proposed framework in real-world scenarios. Indeed, results presented in this paper have been based on emulation using large variety of real-world network traces, but it may be of interest to test further this work by using a video streaming platform to which users shall connect and watch videos in real-world conditions. Setting up such platform would allow demonstrating the benefit of our approach outside of the controlled laboratory environment setup, and will then be part of future work.

VII. Conclusion

In this work, a bitrate ladder estimation method for adaptive bitrate (ABR) video streaming was proposed. The proposed method enables the coding condition to be defined to decrease the likelihood of users quitting videos by considering both content coding complexity and the network's historical throughput data. The method is independent of the strategy used by the video player to request chunks and performs consistently across them. It was shown that while better performing ABR control schemes can decreases the odds of users quitting, considering an optimized bitrate ladder can be also be used as an alternative solution. Moreover, it was shown that ladders can perform stably over long periods and may also perform consistently with fewer quality levels. This can be thought as a way to improve user experience by considering geographical regions and subsequent internet access availability. Future research will include the analysis of video coding complexity as this process can be computationally intensive.

References

[1] F. Y. Yan et al., "Learning in situ: A randomized experiment in video streaming," in *Proc. 17th USENIX Symp. Networked Syst. Des. Implementation*, 2020, pp. 495–512.

[2] P. Lebreton and K. Yamagishi, "Predicting user quitting ratio in adaptive bitrate video streaming," *IEEE Trans. Multimedia*, vol. 23, pp. 4526–4540, 2021.

[3] P. Lebreton and K. Yamagishi, "Transferring adaptive bit rate streaming quality models from H.264/HD to H.265/4 K UHD," *IEICE Trans. Commun.*, vol. E 102-B, no. 12, pp. 2226–2242, 2019.

[4] Google, "Recommended upload encoding settings," Accessed: Jun. 27, 2022. [Online]. Available: https://support.google.com/youtube/answer/1722171?hl=en

[5] Apple, "HLS Authoring Specification for Apple Devices," Accessed: Jun. 27, 2022. [Online]. Available: https://developer.apple.com/documentation/http_live_streaming/hls_authoring_specification_for_apple_devices

[6] Twitch, "Broadcasting Guidelines," Accessed: Jun. 27, 2022. [Online]. Available: https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2

[7] A. Aaron, Z. Li, M. Manohara, J. D. Cock, and D. Ronca, "Per-title encode optimization," *Netflix Technol. Blog*, 2015. [Online]. Available: https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2

[8] J. D. Cock, Z. Li, M. Manohara, and A. Aaron, "Complexity-based consistent-quality encoding in the cloud," in *Proc. IEEE Int. Conf. Image Process.*, 2016, pp. 1484–1488.

[9] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," *Netflix Technol. Blog*, 2022. Accessed: Jun. 27. [Online]. Available: https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652

[10] M. Manohara, A. Moorthy, J. D. Cock, I. Katsavounidis, and A. Aaron, "Optimized shot-based encodes: Now Streaming!," *Netflix Technology Blog*, Accessed: Jun. 27, 2022.

[11] A. V. Katsenou, J. Sole, and D. R. Bull, "Content-gnostic bitrate ladder prediction for adaptive video streaming," in *Proc. IEEE Picture Coding Symp.*, 2019, pp. 1–5.

[12] S. Ling, Y. Baveye, P. L. Callet, J. Skinner, and I. Katsavounidis, "Towards perceptually-optimized compression of user generated content (UGC): Prediction of UGC rate-distortion category," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2020, pp. 1–6.

[13] M. Bhat, J. M. Thiesse, and P. L. Callet, "A case study of machine learning classifiers for real-time adaptive resolution prediction in video coding," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2020, pp. 1–6.

[14] M. Bhat, J. M. Thiesse, and P. L. Callet, "Combining video quality metrics to select perceptually accurate resolution in a wide quality range: A case study," in *Proc. IEEE Int. Conf. Image Process.*, 2021, pp. 2164–2168.

[15] V. V. Menon, H. Amirpour, M. Ghanbari, and C. Timmerer, "OPTE: Online per-title encoding for live video streaming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2022, pp. 1865–1869.

[16] Mux, "Better quality through machine learning," Accessed: Jun. 27, 2022 [Online]. Available: https://mux.com/per-title-encoding

[17] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[18] D. Laming, "Weber's law," in *Inside Psychology: A Science Over 50 Years*. London, U.K.: Oxford Univ. Press, 2009, pp. 179–191.

[19] C. Chen, Y. Lin, S. Benting, and A. Kokaram, "Optimized transcoding for large scale adaptive streaming using playback statistics," in *Proc. IEEE 25th Int. Conf. Image Process.*, 2018, pp. 3269–3273.

[20] Y. A. Reznik, K. O. Lillevold, A. Jagannath, J. Greer, and J. Corley, "Optimal design of encoding profiles for ABR streaming," in *Proc. 23rd Packet Video Workshop*, 2018, pp. 43–47.

[21] Y. A. Reznik, X. Li, K. O. Lillevold, A. Jagannath, and J. Greer, "Optimal multi-codec adaptive bitrate streaming," in *Proc. IEEE Int. Conf. Multimedia Expo Workshop*, 2019, pp. 348–353.

[22] L. Toni et al., "Optimal selection of adaptive streaming representations," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 11, pp. 1–26, 2015.

[23] S. Wolf and M. Pinson, "Video quality measurement techniques," in National Telecommunications and Information Administration, Report 02-392, 2002.

[24] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," ITU-T, 2008.

[25] J. Chen, H. Milner, I. Stoica, and J. Zhan, "Benchmark of bitrate adaptation in video streaming," *J. Data Inf. Qual.*, vol. 13, no. 4, pp. 1–24, 2021. [Online]. Available: https://doi.org/10.1145/3468063

[26] X. Yin, A. Jindal, V. Sekar, and B. Sinopoli, "A control-theoretic approach for dynamic adaptive video streaming over HTTP," in *Proc. Conf. ACM SIGCOMM*, 2015, pp. 325–338.

[27] "Qualinet white paper on definitions of quality of experience," *Eur. Netw. Qual. Exp. Multimedia Syst. Serv.*, P. L. Callet, S. Möller, and A. Perkis, Eds., Lausanne, Switzerland, Mar. 2013.

[28] M. Seufert et al., "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surv. Tut.*, vol. 17, no. 1, pp. 469–492, Jan.–Apr. 2015.

[29] P. Lebreton and K. Yamagishi, "Network and content-dependent bitrate ladder estimation for adaptive bitrate video streaming," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 4205–4209.

[30] J. Li, L. Krasula, Y. Baveye, Z. Li, and P. L. Callet, "AccAnn: A new subjective assessment methodology for measuring acceptability and annoyance of quality of experience," *IEEE Trans. on Multimedia*, vol. 21, no. 10, pp. 2589–2602, Oct. 2019.

[31] W. Song and D. W. Tjondronegoro, "Acceptability-based QoE models for mobile video," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 738–750, Apr. 2014.

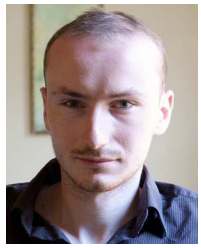
[32] M. Sasse and H. Knoche, "Quality in context-an ecological approach to assessing QoS for mobile TV," in *Proc. 2nd ISCA/DEGA Workshop Perceptual Qual. Syst.*, 2006, pp. 11–20.

[33] P. Spachos, W. Li, M. Chignell, L. Zucherman, and J. Jiang, "Acceptability and quality of experience in over the top video," in *Proc. IEEE ICC 2015 - Workshop Qual. Experience-based Manage. Future Internet Appl. Serv.*, 2015, pp. 1693–1698.

[34] R. Apteker, J. Fisher, V. Kisimov, and H. Neishlos, "Video acceptability and frame rate," *IEEE MultiMedia*, vol. 2, no. 3, pp. 32–40, 1995.

[35] T. C. M. de Koning, P. Veldhoven, H. Knoche, and R. E. Kooij, "Of MOS and men: Bridging the gap between objective and subjective quality measurements in mobile TV," *Proc. SPIE*, vol. 6507, pp. 196–206, 2007.

[36] H. Nam, H. Schulzrinne, and K. Kim, "Youslow: What influences user abandonment behavior for internet video?," Columbia University, 2016.
[37] S. Wu, M.-A. Rizoiu, and L. Xie, "Beyond views: Measuring and predicting engagement in online videos," in *Proc. AAAI Int. Conf. Weblogs Soc. Media*, 2018, pp. 434–443.
[38] Y. Chen, B. Zhang, Y. Liu, and W. Zhu, "Measurement and modeling of video watching time in a large-scale internet video-on-demand system," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 2087–2098, Dec. 2013.
[39] P. Lebreton and K. Yamagishi, "Study on viewing completion ratio of video streaming," in *Proc. IEEE 22nd Int. Workshop Multimedia Signal Process.*, 2020, pp. 1–6.
[40] Akamai, "Maximizing audience engagement: How online video performance impacts viewer behavior," *White Paper*, 2012. [Online]. Available: https://content.akamai.com/PG2114-Audience-Engagement-WP.html
[41] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior : Inferring causality using quasi-experimental designs," *ACM Trans. Netw.*, vol. 21, no. 6, pp. 2001–2014, 2013.
[42] M. Zhibin, A. Raake, W. Robitza, and N. Zhangyan, "Training test results for G.QUIT and model structure," *International Telecommunication Union, Study Group 12, ITUI Contribution SG12-C370R2*, 2019.
[43] S. Takahashi, K. Yamagishi, P. Lebreton, and J. Okamoto, "Impact of quality factors on users' viewing behaviors in adaptive bitrate streaming services," in *Proc. IEEE 11 th Int. Conf. Qual. Multimedia Experience*, 2019, pp. 1–6.
[44] X. Tan, Y. Guo, M. Orgun, L. Xue, and Y. Chen, "An engagement model based on user interest and QoS in video streaming systems," *Wireless Commun. Mobile Comput.*, vol. 2018, pp. 1–11, 2018.
[45] M.-N. Garcia et al., "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. IEEE 6th Int. Workshop Qual. Multimedia Exp.*, 2015, pp. 141–146.
[46] H. Nam, K. Kim, and H. Schulzrinne, "QoE matters more than QoS: Why people stop watching cat videos," in *Proc. IEEE 35th Annu. IEEE Int. Conf. Computer Commun.*, 2016, pp. 1–9.
[47] M. Z. Shafiq et al., "Understanding the impact of network dynamics on mobile video user engagement," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 42, pp. 367–379, 2014.
[48] A. Balachandran et al., "Developing a predictive model of quality of experience for internet video," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, pp. 339–350, 2013.
[49] K. Yamagishi and T. Hayashi, "Parametric quality-estimation model for adaptive-bitrate-streaming services," *IEEE Trans. Multimedia*, vol. 19, no. 7, pp. 1545–1557, Jul. 2017.
[50] S. Chen et al., "Shot conservative self-supervised learning for scene boundary detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 9796–9805.
[51] F. Y. Yan et al., "Puffer experimental results," Accessed: Jun. 27, 2022. [Online]. Available: https://puffer.stanford.edu/results
[52] T.-Y. Huang, R. Johari, N. McKeown, M. Trunnell, and M. Watson, "A buffer-based approach to rate adaptation: Evidence from a large video streaming service," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, pp. 187–198, 2014.
[53] Institute for Telecommunication Sciences, "The consumer digital video library,". Accessed: Jun. 27, 2022. [Online]. Available: https://www.cdvl.org
[54] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," *ITU-T*, 2008.
[55] W. Robitza, R. R. R. Rao, S. Göring, and A. Raake, "Impact of spatial and temporal information on video quality and compressibility," in *Proc. IEEE 13th Int. Conf. Qual. Multimedia Experience*, 2021, pp. 65–68.



Pierre Lebreton (Member, IEEE) received the Engineering degree in computer science from Polytech' Nantes, Nantes, France, in 2009. In 2010, he joined the Group Assessment of IP-based Applications, Berlin Institute of Technology, Berlin, Germany, where he studied toward his Ph.D. on 3D video QoE. After graduating, he joined the Group of Audio Visual Technology, TU-Ilmenau, Germany, in 2015, and the Group of Networked Sensing and Control, Zhejiang University, Hangzhou, China, in 2016. His research interests include various topics including aesthetic appeal, large scale video quality monitoring, and bike sharing systems. In 2017, he joined NTT Laboratories, where he currently focuses on quality and user-engagement prediction for video streaming applications.

Kazuhisa Yamagishi received the B.E. degree in electrical engineering from the Tokyo University of Science, Tokyo, Japan, in 2001, and the M.E. and Ph.D. degrees in electronics, information, and communication engineering from Waseda University, Shinjuku City, Japan, in 2003 and 2013. Since joining NTT Laboratories in 2003, he has been engaged with the development of objective quality estimation models for multi-media telecommunications. From 2010 to 2011, he was a Visiting Researcher with Arizona State University, Tempe, AZ, USA. He was the recipient of the Young Investigators Award in Japan in 2007, the Telecommunication Advancement Foundation Award in Japan in 2008, the ITU-AJ Encouragement Award in 2017, and the TTC Award for distinguished service in 2018.