Predicting User Quitting Ratio in Adaptive Bitrate Video Streaming

Pierre Lebreton and Kazuhisa Yamagishi

***Abstract*—To improve user engagement such as viewing time, this paper addresses the understanding and prediction of the *user quitting ratio* for users watching videos using adaptive bit rate video streaming. The *user quitting ratio* is defined as the percentage of users still watching videos at a given time. To perform this study, five subjective experiments involving up to 264 participants were conducted in a laboratory setting. Results indicated the effects of coding quality, initial buffering, and midway stalling on *user quitting ratio*. Then, a framework was defined to predict the *user quitting ratio* as a function of time. This framework achieves good prediction accuracy and can be used in multiple scenarios including when quality adaptation and stalling occur. Finally, it is suitable for monitoring applications where bitstream are encrypted and low processing cost is required.**

***Index Terms*—Adaptive bitrate video streaming, monitoring, quitting ratio, user engagement.**

I. Introduction

VIDEO streaming is one of the main applications on the Internet. To encourage users to use their services, video streaming service providers (e.g., Netflix, YouTube, Hulu, and Twitch) need to ensure that the experience of the users is adequate. This enables longer watching periods and is referred to a *high user engagement*. High user engagement also benefits businesses that engage with the video streaming services such as advertising agencies and enables a healthy ecosystem.

When watching videos, services can be degraded, and as the network bandwidth fluctuates, video quality varies and stalling occurs. These degradations can lead to low engagement from the users. For example, deep packet inspection of Wi-Fi traffic on a campus revealed that, across multiple video streaming services, 50% of sessions were shorter than half of the total video duration. Data from an Internet Service Provider (ISP) in the US shows similar results: videos were fully watched in only 21% of cases, and viewing sessions were less than 5 min in 70% of cases [1]. Akami [2] and Krishnan and Sitaraman [3] also showed this, as engagement was found to be closely related to quality. Therefore, quality and its impact on engagement need to be controlled.

Manuscript received March 27, 2020; revised September 14, 2020; accepted December 6, 2020. Date of publication December 14, 2020; date of current version December 9, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. João M Ascenso. *(Corresponding author: Pierre Lebreton.)*

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

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TMM.2020.3044452.

Digital Object Identifier 10.1109/TMM.2020.3044452

A lot of effort has been put into predicting the perceived quality of video services in Quality of Experience (QoE) studies [4]. This has been done in both academia and industry and has resulted in the successful development of models that can predict QoE [5]–[8]. Recently, user engagement has been considered as it is a further step to enable services to be monitored in terms of not perceived quality (QoE) but willingness to use the service, which is more closely related to businesses' incomes. This comes with new challenges as the prediction of engagement depends on QoE, but other factors such as context and contents are also involved. The outcome of this study is twofold: first, it aims to identify the complex relationship between quality-related features and user quitting actions with respect to temporal aspects. Second, this knowledge is applied to establish a new prediction model that can be used by video service providers to extend their monitoring [9], [10] and service enhancement [11], [12] activities with user-quitting-based engagement control.

To enable this, "actionable models" are needed [13]. Actionable models can be used for optimization purposes as they depend on factors that are well known, measurable, and available at a reasonable cost with respect to the application. Examples of applications include the control of quality adaptation and pre-fetching of video players [2], [11], [14], the optimization of server load [15], [16], the optimization of wireless infrastructure [1], and the content placement on content delivery networks [13], [17].

This paper describes a study in which features relevant to the *user quitting ratio* are identified, analyzed, and quantified with respect to temporal aspects. Then, these features are used to develop an actionable prediction model of the *user quitting ratio*. The model is no-reference parameter-based, enabling it to be used in monitoring applications where low complexity is needed and for which bitstream data is not available due to encryption.

The *user quitting ratio* is the percentage of users watching a given video as a function of time. Its main advantage over traditional approaches is that it does not restrict itself to predicting a single quitting time value but is a function of time. It provides a risk-based analysis of the quitting behavior. Moreover, studying temporal evolution of the quitting ratio has other benefits as it can be used to identify the main quality issues that need to be addressed first: the most critical events are those for which the quitting ratio increases largely, whereas areas with flat quitting ratios indicate sufficient quality. Therefore, this gives an overall view on the viewing session and enables a good understanding of the status of sessions and the consequences of actions taken

during optimization. This will enable services to make informed decisions on how to control the video quality while users are watching videos, which will decrease the risk of abandonment.

This paper is organized as follows. Section II details related work. Sections III and IV describe the subjective experiments used in this study and their results, respectively. Sections V and VI address the modeling of the user quitting ratio and performance evaluation. Sections VII and VIII discuss results and alternative models. Finally, Section IX concludes this paper.

II. Related Work

A. Measurements and Factors Affecting Engagement

1) Methodology: User engagement in video services is difficult to evaluate as a function of technical, content, and context-related factors. Different approaches have been used and were either based on laboratory experiments or from real-life services. The use of laboratory experiments enables the effect of precisely defined conditions to investigated, making modeling work easier. However, it comes with challenges as laboratory experiments are a very specific context that can affect results. One example is the "good participant" phenomenon, where participants behave in accordance with their understanding of the experimenter's expectation instead of behaving normally [18]. To address this problem, "deceptive studies," where the goal of the experiment is hidden, can be used [18]. Another potential issue is that participants may be afraid to change the experimental setup and do not dare to interact with it, which will result in fewer manipulations than in a real-life scenario [18]. One way to address this problem may to provide users with more familiar equipment or letting them use their own devices in crowdsourcing experiments. Another challenge can be "cognitive dissonance" [19], [20]. This relates to participants feeling stress and discomfort as they may consider different contradictory aspects while doing their task. For example, they may fill out a questionnaire while thinking about how their answers match their actions in previous steps of the test [19], [20]. Special attention is thus needed in designing questions and tasks. Last but not least, motive is another aspect as participants in laboratory tests are paid, and this may also affect the results [19].

Alternatively to laboratory tests, a common way to study engagement is to analyze data collected from real-life services. This comes with the advantage of providing actual usage information enabling the observation of the consequences of technical parameters, interest in content, and effect of context on engagement. However, as conditions are not crafted, causality is difficult to identify, and results are limited to correlation analysis. This is even more challenging as real-life data does not enable us to easily disambiguate the effect of each parameter affecting engagement. Indeed, many factors vary across measurements: the users' preferences, the network conditions, the contents, the coding conditions, the context, etc. This makes the effects of different factors difficult to disambiguate. For example, Seufert *et al.* [21] showed that a direct correlation analysis fails to find a correlation between quality ratings with technical parameters such as throughput, stalling duration, number of stalling, initial delays, etc. even though a large amount of research in laboratory context has shown otherwise [22], [23]. This shows that when too many factors are changing at the same time, the effect of each factor taken individually becomes difficult to measure. To address this challenge, previous work has suggested using multiple analysis techniques (correlation, information gain, regression, etc.) as using only one analysis technique may lead to incorrect conclusions [13]. Confounding factors (context, contents, users, etc.), could also be either treated as a new feature when building models, or disambiguated by splitting the data. To address causality between network-related and contextual features with user engagement, the Quasi-Experimental Design (QED) was proposed [3]. QED consist of paring a measurement from a "treated" case with an "untreated" one that is "significantly identical". For example, a condition with stalling can be compared with another condition without stalling, while keeping other factors such as content, bitrate, frame rate, and network conditions constant. The pairing process enables the impact of confounding factors to be decreased and the effect of one factor to be tested. However, authors stress that it is difficult to eliminate all hidden factors as the pairing process and the identification of a "significantly identical" condition are difficult. Therefore, this should be considered as a way to strengthen correlation analysis but not as proof of causality [3].

These results show that both laboratory tests and data-driven analysis are challenging and have their own strengths and weaknesses. This work contributes to the state of the art in different ways. First, it addresses a relaxed problem: users quitting because of bad quality and not user engagement per se. This relaxation in terms of the research question enables us to conduct laboratory experiments, so we can repeat the same conditions over multiple participants. Doing so enables the cumulative distribution of viewing time to be computed on a per-video basis for well-defined quality impairments. This precise analysis of the cumulative distribution of viewing time will be further referred to as the *user quitting ratio*. Such detailed analysis has not been done to this extent in earlier work and has the benefit of enabling the investigation of quality-related features on user quitting actions for well-defined conditions, which led to the development of a comprehensive model of the *user quitting ratio*. Finally, it should also be stated that addressing the *user quitting ratio* has a benefit in terms of an evaluation technique, as it is a measurement over groups of participants that enables the noise coming from individuals to be decreased.

2) Effect of Quality Parameters: When watching videos, three main quality-related impairments can occur: initial loading delay, midway stalling, and low coding quality. In this section, previous work on the relationship between these impairments and user abandonment is reported.

First, the initial loading delay is discussed. Past studies showed that the initial loading delay has a limited impact on QoE [22]. However, when abandonment was considered, a strong effect was found. If the initial loading delay is less than 1 s, it has a limited effect, but if it is more than 2 s, users begin to quit [3]. Then, quitting further increases for every additional second by a rate of 5.8% [2], [3]. The relationship between initial delay duration and abandonment is thus exponential [2], [24],

[25]. However, the initial loading delay is not the only factor, and other aspects need to be considered, including the quality of the video that users are waiting for, the content, and the context of the viewing session. This has been shown in studies where users were more patient towards higher quality [26], and studies about rare sports events showed higher acceptance levels such as finding a 10-s initial delay acceptable [27]. This shows that the simplest case of the initial delay is already multi-dimensional.

Second, stalling midway through a video is discussed. In this case, abandonment is found to increase logarithmically with the stalling duration [26], [28]. Going from 12 to 24 s results in abandonment increasing by 30% [29]. A 20-s stalling event results in 50% abandonment, and a 40-s stalling event results in 80% abandonment [26]. Similarly to initial loading delay, stalling duration is not the only characteristic to be considered, and stalling frequency is another key factor. Indeed, if the stalling frequency is low, the stalling duration is the dominant factor affecting abandonment [26]. However, if the stalling frequency is high, the interval between two stalling events becomes the most relevant factor [30], [31]. This has been addressed in different ways in the literature. For example, for the same total stalling duration, abandonment was found to increase by a factor of 3 if the number of stalling events goes from 2 to 3 [30]. Another example is rebuffering ratio (total stalling duration divided by playback duration) studies that showed that increasing the buffering ratio results in increasing abandonment [27], [28], [31]–[33] (a 1% increase in the rebuffering ratio reduced viewing time by 67% [27], a 0.13 increase in stalling per minute decreased viewing time by 20 min [32], and an 1% increase in the rebuffering ratio results in a 3.33% reduction in viewing time [34]).

These studies provide knowledge on the effect of stalling on abandonment. However, several limitations need to be considered. Indeed, these analyses are based on a mixture of multiple conditions and do not address the specificities of a per-viewing session analysis. For example, a 10-s stalling event occurring 30 s after the video starts will be perceived differently from a 10-s stalling event occurring after 1 h of uninterrupted playback (although average duration and stalling frequency are identical). This was neglected in previous work, and further work is needed to address per-viewing session analysis.

Third, video coding quality is discussed. Video coding quality affects viewing time, but this was not clearly characterized. Increasing the bitrate does not necessarily increase viewing time [1], [13], [27], [30], [34], [35], and average bitrate is lowly correlated with the viewing time [34]. On the other hand, quality variation has a high impact. If quality drops, users will quit, but even if quality increases, the abandonment rate also increases compared with a constant quality [30], [31]. Therefore, the quality's variation has the highest correlation with viewing time [30], [31], [35]. Then, similarly to stalling, the relationship between quality and abandonment is multi-dimensional, and factors such as context are important. Indeed, depending on the type of service (Video On Demand (VOD) vs. live streaming), the relationship between bitrate and abandonment differs, and the average bitrate is significant for live streaming but not for VOD [13], [32].

From these studies, several points require further investigation. First, the limited effect of bitrate on abandonment in the case of VOD is unexpected as perceived image quality has been a heavily researched topic. Such a result can be explained by the fact that previous quality-based analysis is limited to bitrate analysis, and bitrate is only one component of the overall quality equation along with resolution, frame rate, and content complexity. Therefore, we contribute to the state of the art by relating quitting behavior with quality measurements and not only bitrate-based analysis. Second, although previous work has identified the effect of quality variation frequency on abandonment, these studies did not clarify the effect of the quality before and after the change as too many factors were changing at the same time to draw conclusions. However, this work can clarify this effect.

Finally, note that in a viewing session, all types of quality impairments can occur (initial loading delay, stalling midway through, and low coding quality), so their relative impacts on abandonment need to be understood. Overall, midway stalling is identified as having the largest effect on abandonment of services [31]. Midway stalling was found to result in an abandonment rate six times higher than a similarly long initial loading delay [30], and a single stalling event was found to have three times the impact of bitrate changes in YouTube videos [30]. The stalling ratio is more highly correlated with viewing time than either the average bitrate [27], [34], [36] or the number of bitrate changes [31]. However, depending on context, the importance of the bitrate may be strengthened [13], [34].

These results give an understanding of the weight between impairments, but further quantification is still needed as the consequences of low coding quality and midway stalling vary depending on their severity. In addition, the temporal aspect between impairment should also be addressed. For example, a low quality condition following a stalling event will differ more in terms of quitting ratio than a low coding quality condition only. These interactions were not previously studied in terms of abandonment but can be studied in this work thanks to the per-viewing session analysis.

3) Effect of Contents and Channels: Contents are an important aspect to consider in engagement studies. Content dependency is visible while encoding videos as bitrate requirements differ across contents. Moreover, frame rate reduction and stalling have different visibilities and also result in different magnitudes of abandonment [37]. In addition, the interest of users in content results in different actions [37] and tolerance towards technical issues [3], [38]. This can be seen as certain types of content such as animation can have a longer viewing time than other types of contents [37], [38]. To study the user engagement with regards to content, popularity on the video streaming platform was found to be relevant. Wu *et al.* [39] defined the concept of "relative engagement," which is based on the ranking of content on a streaming platform. Popularity of videos was found to be logarithmically related to migration between service providers [40] and videos' completion ratio [41].

Beyond popularity, content properties such as content duration also impact abandonment [1], [30], [31], [33], [35], [39], [42], and a linear relationship was found between the content

duration and completion ratio [31], [39], [41]. By categorizing videos as short (less than 30 min) or long, differences in abandonment could be found [2], [3], [35]. Content duration was also found to affect the effect of stalling on abandonment [2], [30], [31], and for a similar stalling duration, abandonment is increased by a factor of two in short videos compared with long ones [2].

This work extends past analysis as it investigates the temporal evolution of the user quitting ratio for contents with different durations. Considering temporal aspect instead of only the completion ratio is important as it enables disambiguation of the effect of content duration and quality-related factors.

4) Effect of Context: Across the influencing factors described in this paper, context is an important confounder. The viewing environment (e.g., the type of Internet access or device) results in different expectations [2], [3], [27]. An initial delay of 3 s can result in an abandonment of 13% when using cable or 15% when using optic fiber [3]. Moreover, the day of viewing is also important as viewing time was found to be 32% lower on weekdays than weekends [16], [43]. This shows that external factors such as the time available to the user also affect user decisions. Finally, attributes such as the age of the users are also relevant [13]. In this work, context is constant and corresponds to a laboratory environment. This ensures stability of the results across test conditions and experiments, which enables the effect of quality factors on the *user quitting ratio* to be tested. However, this is a specific context, and it should be specified that the results are defined in that context.

B. Modeling

Different types of modeling can be found in the literature. Conceptual models of user engagement have been proposed [20], [44] and provide thoughts about how users engage with content. However, cannot be practically implemented.

Several models with a well-defined implementation have been proposed. Among them, behavior models are commonly based on queuing models [15], [45] or finite state machines [16]. In the latter case, a first-order Markov chain does not perform sufficiently [16], and a semi-Markov chain should be used instead [16]. Although behavior models are of interest, these cannot be applied in this work as user actions are constrained in our studies.

When addressing viewing time, a popular approach is to employ regression trees. This is motivated by the fact they are expressive enough to capture complex relationships between quality and contextual-related features. In addition, these provide good accuracy while still being comprehensive [1], [13]. Models based on matrix factorization have also been proposed to predict viewing time by considering user preferences [46]. These models are of interest, but as this work aims at estimating the user quitting ratio, it is not obvious how these methods can be directly applied.

Addressing viewing time distribution, most types of video viewing time distribution can be modeled by a skew-normal distribution [41]. A general approach consists of using a combination of two distributions: one for modeling the initial screening phase, and the other for modeling the main viewing session. For the initial screening phase, a Weibull distribution is frequently chosen [26], and the main viewing session can be either based on a generalized Pareto distribution [16], [47] or a Log-Normal distribution [47], [48]. The viewing time distribution is content dependent, so previous work proposed training the model on a per-content category basis [37]. In a more generic case, the viewing time distribution can be considered as the general problem of estimating the probability density function [49] or a survival problem [50]. These works have the limitation that analyses were always performed over multiple viewing sessions mixing various types of impairments with different magnitudes, positions in time, and frequencies. This resulted in distributions that have conceptual differences in characteristics from the user quitting ratio. Therefore, these results cannot be directly applied, and further work is needed.

In terms of modeling, this work contributes to the state of the art as it describes a comprehensive model of the user quitting ratio. This extends previous work, as it estimates the user quitting ratio on a per-video basis with specific impairment and does not limit itself to averages over sessions with multiple quality conditions as was done in past studies on viewing time distribution. Going into such detail enables a quitting-based model to be established that can be used to monitor services and actively control the quality on a per-viewing session basis as users watch videos. To achieve this, the model is no-reference and parameter-based, enabling it to be used with low processing cost and with encrypted video bitstreams as required by monitoring applications.

C. Contributions

The evaluation of user abandonment presents numerous challenges. This subsection summarizes key points addressed in this paper.

- First, methodology-wise, quality-induced abandonment is investigated in laboratory experiments. This enables this work to test the exact same conditions over multiple participants, enabling the cumulative distribution of viewing time to be studied on a per-video basis for precise quality settings.
- Previous quality-based analysis were limited to bitrate analysis. Therefore, this work relates abandonment to QoE.
- Previous work identified the effect of quality variation frequency on abandonment but weakly characterized it. Therefore, this study investigates the effect of quality change magnitude as well as quality before and after the change on the abandonment.
- Previous work compared the impact of low coding quality and stalling on abandonment. However, the link to the severity of each impairment was missing and is now tested on the basis of intensive testing.
- The temporal evolution of abandonment is studied and enables disambiguation between the effect of content duration and abandonment due to quality-related factors.
- Finally, a model of the user quitting ratio is proposed. This model has the key novel property that it estimates the user

quitting ratio on a per-viewing session basis. The model can then reflect on the individual experience of the users instead of providing averages across groups.

III. Subjective Experiment

This work is based on multiple subjective experiments. Considering that user quitting ratio is addressed, videos from 3 to 10 min are used. Working with medium-length videos raises some challenges, so to maintain the total duration of laboratory tests, the number of conditions in a single experiment needs to be limited. Therefore, five different experiments (A, B, C, D, and E) were conducted in 2018 and 2019 and involved different participants.

Thirty-two participants took part in experiment A and 40 in experiment B. In experiments C, D, and E, two groups of 32 participants were formed and watched half of the conditions (hence 64 participants took part in those experiments). This resulted in a total of 264 participants. Finally, special care was taken to ensure exact gender balance, and participants' aged ranged from 18 to 30 years old, with a mean of 21.

In the following, the experiments are detailed in depth.

A. Test Conditions

The goal of this study is to understand and predict the user quitting ratio in adaptive bitrate video streaming. Therefore, conditions were selected to offer a wide variety of quality changes patterns: constant/increasing/decreasing quality, oscillations, and large/smooth quality changes that could include (or not) stalling events with various durations and positions (including initial buffering). These conditions were chosen on the basis of statistics about real-life services and past research into QoE and engagement that showed that quality changes and stalling are important factors (see Section II-A2). The test design of test conditions includes constant quality conditions with various quality levels enabling the effect of coding quality on the user quitting ratio to be modeled. This also serves as a baseline for conditions with quality changes and/or stalling. These conditions enable us to test the effect of quality changes and account for the quality before and after changing on the user quitting ratio. This has been missing in previous work. Different positions for stalling and quality adaptation are also tested as these were not addressed in previous work on user abandonment. Finally, interactions between quality events: low coding quality after stalling vs. only low coding quality is also tested, as it was missing in past studies. All these conditions enable a better understanding of the user quitting ratio and comprehensive model development.

To test the effect of coding quality on the user quitting ratio, various coding conditions are defined and listed in Table I. Encoding was done using H.265/HEVC (high efficiency video coding) for the coding conditions Q1- Q36, and H.264/AVC (advanced video coding) for the coding conditions Q37-Q48. These codecs were chosen as H.265 is frequently used for 4 K videos and H.264 is still heavily used for high-definition (HD) videos. Therefore, to cover a wide variety of scenarios, both codecs were considered. Videos were encoded using FFmpeg (versions 4.0.2-4.2.0) with the x265 and x264 codecs, respectively. A two-pass encoding with the preset "slower" was used. As for audio, two channels and a sampling rate of 48 kHz were used. The audio codec was always AAC-LC (low complexity advanced audio coding), and the encoding was performed using libfdk_aac from FFmpeg. As adaptive bitrate video streaming is addressed, multiple combinations of resolution, bitrate, and frame rate were selected. As making changes across multiple scales (resolution, bitrate, frame rate, and audio bitrate) makes it difficult to compare quality across conditions, estimates of the video, audio, and audiovisual quality using the "NTT audiovisual quality estimation model" [7], [51] are reported. This is a parameter-based no-reference audiovisual quality estimation model. The predictions are defined on a scale from 1 to 5, 1 being the lowest quality and 5 the highest quality. As for audio, special care was taken so that audio quality was never so low as to annoy the participants.

Test conditions are reported in Table I. In this table, each entry corresponds to a Processed Video Signal (PVS) of a given test. The details about the quality adaptations are given by a list of pairs$(Q[148], T)$ that describes consecutive quality changes. $Q[148]$ is the quality level that characterizes a coding condition in Table I, and T refers to the duration of this condition in seconds. As for stalling events, these are marked by $(stall, T)$ with T being the duration in seconds. Stalling events were simulated by freezing the video and by adding a dynamic loading wheel [52].

Video duration varied from 3 to 10 min. The videos (source reference circuit, SRCs) were 3840x2160 4K-UHD videos with a frame rate of 60 frames per second (FPS). The videos showed a large variety of content corresponding to common TV shows in Japan. Scenes included scenery, festivals, documentaries, sports, interviews, etc. These videos were recorded by video professionals using professional grade cameras. Complexity-wise, the videos had low to high amounts of details as well as varying temporal complexity. Table III provides information on the spatial information (SI) and temporal information (TI) as defined in ITU-T Recommendation P.910 [53]. Both temporal average and maximum values of SI and TI are given. Finally, correspondence between SRCs and processing can be found in Table II.

B. Experimental Setup and Test Methodology

The subjective experiments were conducted on smartphones using a video player designed to record when users click on the stop playback button and hence record viewing time. A 5.5-inch Sony Xperia XZ Premium with a resolution of 3840×2160 was used. Participants listened to the audio using headphones. Special care was taken so that they listened to the audio at -21 dB. The viewing distance was set to 5-7 H (with H being the height of the screen). The experimental room was a laboratory environment that fulfills the standards for video quality tests [53] (gray room, controlled ambient light, acoustic treatments, etc.). The illumination was set to 20 lx, which corresponds to a dark room.

Considering that the quitting ratio is studied, participants were instructed to watch the videos and told that they could stop watching whenever their desired. After quitting, they were not allowed to resume watching. Participants were not able to skip part of the video. Finally, it is important to note that participants

TABLE I
LIST OF QUALITY LEVELS (L). h IS THE RESOLUTION GIVEN BY THE NUMBER OF PIXEL IN HEIGHT OF THE VIDEOS WITH A 16:9 FORMAT, b_a AND b_v ARE THE CONSTANT BITRATE VALUES IN KBPS OF THE AUDIO AND VIDEO BITSTREAM, AND r IS THE FRAME RATE. MOS_v, MOS_a, AND MOS ARE RESPECTIVELY VIDEO, AUDIO, AND AUDIOVISUAL QUALITY ESTIMATES [51]

L	h	b_v	r	b_a	MOS_v	MOS_a	MOS	L	h	b_v	r	b_a	MOS_v	MOS_a	MOS
HEVC															
Q1	2160	15000	60	128	4.71	4.91	5.00	Q19	720	250	30	32	2.28	4.17	2.55
Q2	2160	8000	30	128	4.48	4.91	4.95	Q20	720	160	30	32	1.94	4.17	2.25
Q3	2160	4000	30	64	4.17	4.74	4.56	Q21	480	8000	30	48	3.90	4.58	4.23
Q4	2160	800	30	32	2.86	4.17	3.09	Q22	480	3000	30	32	3.68	4.17	3.83
Q5	2160	200	30	32	1.74	4.17	2.07	Q23	480	900	30	32	3.10	4.17	3.30
Q6	1080	7000	60	192	4.49	4.94	4.98	Q24	480	640	30	128	2.87	4.91	3.37
Q7	1080	4000	30	128	4.29	4.91	4.76	Q25	480	330	30	32	2.37	4.17	2.64
Q8	1080	3000	30	64	4.17	4.74	4.56	Q26	480	200	30	32	2.02	4.17	2.32
Q9	1080	800	30	48	3.27	4.58	3.63	Q27	360	3000	30	48	3.30	4.58	3.66
Q10	1080	600	30	32	3.02	4.17	3.23	Q28	360	450	30	128	2.39	4.91	2.91
Q11	1080	300	30	32	2.39	4.17	2.66	Q29	360	350	30	32	2.23	4.17	2.51
Q12	1080	150	30	32	1.86	4.17	2.18	Q30	360	200	30	32	1.88	4.17	2.20
Q13	720	10000	60	128	4.38	4.91	4.85	Q31	240	4000	30	128	2.67	4.91	3.18
Q14	720	4000	30	48	4.13	4.58	4.44	Q32	240	800	15	32	2.14	4.17	2.43
Q15	720	1000	60	384	3.36	4.96	3.87	Q33	240	50	15	32	1.22	4.17	1.60
Q16	720	1000	30	128	3.42	4.91	3.91	Q34	144	4000	30	128	1.86	4.91	2.39
Q17	720	1000	30	48	3.42	4.58	3.77	Q35	144	100	30	64	1.20	4.74	1.70
Q18	720	400	30	32	2.67	4.17	2.91	Q36	144	100	30	32	1.20	4.17	1.57
AVC															
Q37	1080	12000	60	384	4.69	4.96	5.00	Q43	480	500	30	384	2.35	4.96	2.89
Q38	1080	8000	30	384	4.51	4.96	5.00	Q44	360	1000	30	384	2.90	4.96	3.43
Q39	720	7500	60	384	4.42	4.96	4.92	Q45	360	400	30	384	2.00	4.96	2.54
Q40	720	5000	30	384	4.27	4.96	4.78	Q46	240	300	30	384	1.55	4.96	2.10
Q41	720	1500	30	384	3.69	4.96	4.20	Q47	480	300	30	384	1.79	4.96	2.33
Q42	480	2500	30	384	3.76	4.96	4.27	Q48	360	200	30	384	1.43	4.96	1.99

Encoding command line for HEVC:

ffmpeg -i [input.avi] -filter:v scale=-1:[h] -r [r] -c:v libx265 -b:v [b_v]k -preset slower -x265-params pass=1:keyint=[2*r]:stats=log -pix_fmt yuv420p
-c:a libfdk_aac -b:a [b_a]k -f mp4 /dev/null

ffmpeg -i [input.avi] -filter:v scale=-1:[h] -r [r] -c:v libx265 -b:v [b_v]k -preset slower -x265-params pass=2:keyint=[2*r]:stats=log -pix_fmt yuv420p
-c:a libfdk_aac -b:a [b_a]k [output.mp4]

Encoding command line for AVC:

ffmpeg -i [input.avi] -filter:v scale=-1:[h] -r [r] -c:v libx264 -b:v [b_v]k -preset slower -profile:v high10 -level 5.2 -pass 1 -x264opts merange=64:me=umh:b-pyramid=strict:slices=1:b-adapt=0:bframes=3:scenecut=-1:threads=16:keyint=[2*r] -passlogfile log -pix_fmt yuv420p -c:a libfdk_aac
-b:a [b_a]k -f mp4 /dev/null

ffmpeg -i [input] -filter:v scale=-1:[h] -r [r] -c:v libx264 -b:v [b_v]k -preset slower -profile:v high10 -level 5.2 -pass 2 -x264opts merange=64:me=umh:b-pyramid=strict:slices=1:b-adapt=0:bframes=3:scenecut=-1:threads=16:keyint=[2*r] -passlogfile log -pix_fmt yuv420p -c:a libfdk_aac
-b:a [b_a]k [output.mp4]

were asked to base their decision only on quality-related reasons and not content-related ones. Therefore, the results give the quitting ratio in terms of acceptability of quality with respect to time and not engagement per se.

Before taking the test, participants needed to pass vision tests: visual acuity (with correction glasses if needed) and color vision. If they passed, participants could take part in the tests and were provided with written instruction about their task. The experiment started with a training phase in which participants watched six 3-min videos over three sessions. Then, the main experiments started. Each experiment was split into several sessions. In each session, if PVSs were 3 min long, two PVSs were included in the session. If the PVS were more than 3 min, only one PVS was shown. Between sessions, participants were given a rest. Finally, PVSs were randomized across participants to avoid any effect of presentation order.

IV. EXPERIMENTAL RESULTS

A. Overview on the Results

Figure 1 depicts characteristic examples of the temporal evolution of the *user quitting ratio*. Considering that no experiments included a quality evaluation task, quality was estimated using the "NTT quality estimation model" [7], [51], and predictions for each quality level are listed in Table I. In the following, these estimates are referred to as mean opinion score (MOS). In addition, the figure reports on stalling events. The beginning and end of a stalling event are marked respectively by a continuous vertical red line and a dashed blue line. The plots report on the user quitting ratio that corresponds to the percentage of users who left the PVS at a given point in time. It can be seen that when the MOS is low, users quit the videos (Figure 1(a)). On the other hand, when the MOS is high, users keep watching (Figure 1(c)). If stalling events occur, users quit the videos (Figure 1(b)), and if a large drop in MOS occurs, users will quickly react to the loss of quality and quit the video (Figure 1(d)).

B. Quitting Ratio and Coding Quality

In this section, the effect of coding quality on the quitting ratio is addressed. Initial buffering and midway stalling are part of Section IV-C.

First, the precision for the audiovisual quality model used in this work [7], [51] is required. In this model, audio and video

TABLE II
LIST OF TEST CONDITIONS. DUR. IS THE CONTENT DURATION IN SECONDS

PVS	SRC	Dur.	Details on quality adaptations	PVS	SRC	Dur.	Details on quality adaptations
Experiment A				Experiment A (following part of the list)			
01	01	180	(Q28, 60), (stall, 24), (Q28, 120)	20	09	180	(Q16, 20), (Q35, 60), (Q24, 100)
02	02	180	(Q7, 30), (Q2, 30), (Q7, 30), (Q2, 30), (Q7, 30), (Q2, 30)	21	10	180	(Q24, 120), (stall, 24), (Q24, 60)
03	03	180	(Q24, 60), (stall, 12), (Q24, 120)	22	11	180	(Q2, 50), (Q28, 30), (Q16, 30), (Q7, 30), (Q24, 30), (Q16, 10)
04	04	180	(Q28, 120), (stall, 12), (Q28, 60)	Exp. B			
05	05	180	(Q16, 60), (stall, 24), (Q16, 120)	23	01	300	(Q2, 300)
06	06	180	(Q16, 120), (stall, 12), (Q16, 60)	24	02	300	(Q2, 50), (stall, 24), (Q35, 30), (Q16, 30), (Q7, 20), (stall, 24), (Q35, 50), (Q16, 60), (stall, 24), (Q24, 60)
07	07	180	(Q28, 180)	25	03	300	(Q24, 180), (Q7, 120)
08	08	180	(Q24, 60), (Q16, 60), (Q2, 60)	26	04	300	(Q16, 300)
09	09	180	(Q16, 180)	27	05	300	(Q35, 50), (Q28, 50), (Q24, 50), (stall, 36), (Q28, 50), (Q24, 50), (Q16, 50)
10	10	180	(Q2, 180)	28	06	300	(Q24, 100), (stall, 36), (Q24, 200)
11	11	180	(Q24, 180)	29	07	300	(Q2, 200), (Q16, 100)
12	01	180	(Q2, 60), (Q16, 60), (Q24, 60)	30	08	300	(Q35, 50), (Q28, 50), (Q24, 50), (Q16, 50), (Q7, 50), (Q2, 50)
13	02	180	(Q35, 60), (Q28, 60), (Q24, 60)	31	09	300	(Q2, 50), (Q28, 30), (Q16, 30), (Q7, 20), (Q24, 50), (Q16, 60), (Q24, 60)
14	03	180	(Q24, 60), (Q28, 60), (Q35, 60)	32	10	300	(Q7, 180), (Q16, 120)
15	04	180	(Q16, 60), (stall, 8), (Q28, 60), (Q24, 60)	33	11	300	(Q35, 50), (Q28, 50), (Q24, 50), (stall, 36), (Q35, 50), (Q28, 50), (Q24, 50)
16	05	180	(Q16, 60), (stall, 8), (Q28, 30), (Q24, 30), (stall, 12), (Q28, 60)	34	12	300	(Q24, 100), (stall, 24), (Q24, 200)
17	06	180	(Q24, 60), (stall, 12), (Q24, 60), (stall, 12), (Q24, 60)	35	16	300	(Q16, 300)
18	07	180	(Q24, 60), (Q35, 60), (stall, 12), (Q28, 30), (stall, 6), (Q35, 30)	36	17	300	(Q24, 100), (Q35, 200)
19	08	180	(Q28, 60), (Q24, 60), (Q35, 40), (Q24, 20)				
Experiment C							
37	01	600	(Q4, 90), (Q9, 90), (Q17, 90), (Q22, 60), (Q17, 60), (Q5, 90), (Q31, 60), (Q34, 60)	50	01	600	(Q14, 40), (stall, 12), (Q14, 50), (stall, 8), (Q14, 50), (Q25, 30), (Q23, 30), (Q29, 30), (Q22, 50), (stall, 8), (Q23, 50), (Q17, 30), (Q23, 30), (stall, 6), (Q25, 30), (Q32, 30), (Q27, 30), (Q23, 30), (Q14, 50), (Q6, 40)
38	02	600	(Q14, 30), (Q18, 30), (Q11, 90), (Q18, 30), (Q20, 30), (Q14, 30), (Q23, 30), (Q12, 70), (Q18, 30), (Q17, 30), (Q19, 90), (Q10, 40), (Q1, 40), (Q17, 30)	51	02	600	(Q18, 40), (Q18, 50), (Q18, 50), (Q25, 30), (Q23, 30), (Q29, 30), (Q29, 50), (Q23, 50), (Q17, 30), (Q29, 30), (Q25, 30), (Q32, 30), (Q27, 30), (Q23, 30), (Q14, 50), (Q6, 40)
39	08	180	(Q19, 50), (Q17, 40), (Q10, 30), (Q3, 30), (Q17, 30)	52	08	180	(Q14, 20), (stall, 6), (Q14, 70), (stall, 6), (Q14, 50), (stall, 6), (Q14, 40)
40	09	180	(Q1, 180)	53	09	180	(Q13, 180)
41	10	180	(Q29, 40), (Q23, 30), (Q17, 30), (Q30, 40), (Q23, 40)	54	10	180	(Q14, 20), (stall, 12), (Q14, 70), (stall, 12), (Q14, 50), (stall, 12), (Q14, 40)
42	11	180	(Q10, 90), (Q17, 50), (Q3, 40)	55	11	180	(Q19, 20), (stall, 6), (Q19, 70), (stall, 6), (Q19, 50), (stall, 6), (Q19, 40)
43	06	300	(Q22, 60), (Q32, 40), (Q22, 50), (Q23, 30), (Q17, 30), (Q11, 40), (Q18, 50)	56	06	300	(Q29, 60), (stall, 12), (Q29, 20), (Q32, 20), (stall, 12), (Q29, 30), (Q32, 30), (stall, 12), (Q32, 30), (Q23, 30), (stall, 4), (Q25, 30), (Q23, 20), (Q25, 30)
44	07	300	(Q23, 30), (Q33, 30), (Q17, 60), (Q25, 40), (Q22, 60), (Q14, 60), (Q30, 20)	57	07	300	(Q23, 50), (Q17, 40), (Q6, 40), (stall, 12), (Q14, 40), (Q23, 50), (Q14, 40), (Q6, 40)
45	12	180	(Q17, 50), (Q25, 30), (Q23, 40), (Q36, 60)	58	12	180	(Q17, 50), (Q13, 50), (Q1, 80)
46	13	180	(Q6, 180)	59	13	180	(Q1, 40), (stall, 4), (Q1, 20), (stall, 4), (Q8, 30), (stall, 4), (Q17, 20), (stall, 4), (Q23, 10), (stall, 6), (Q29, 30), (stall, 4), (Q29, 30)
47	03	420	(Q17, 20), (Q25, 30), (Q17, 70), (Q23, 30), (Q13, 30), (Q17, 30), (Q25, 90), (Q23, 30), (Q13, 30), (Q33, 20), (Q23, 40)	60	03	420	(Q6, 50), (stall, 6), (Q8, 30), (stall, 6), (Q6, 40), (stall, 6), (Q14, 30), (stall, 6), (Q23, 40), (stall, 6), (Q8, 30), (Q14, 40), (Q8, 40), (stall, 4), (Q23, 40), (Q14, 40), (Q8, 40)
48	04	420	(Q8, 30), (Q32, 30), (Q17, 40), (Q23, 30), (Q32, 30), (Q17, 30), (Q23, 20), (Q32, 30), (Q17, 30), (Q32, 30), (Q17, 30), (Q32, 90)	61	04	420	(Q8, 40), (Q14, 40), (Q23, 40), (stall, 4), (Q8, 40), (Q14, 40), (Q8, 30), (stall, 6), (Q14, 40), (stall, 6), (Q23, 30), (stall, 6), (Q6, 40), (stall, 6), (Q8, 30), (stall, 6), (Q6, 50)
49	05	420	(Q13, 30), (Q23, 30), (Q19, 90), (Q21, 30), (Q23, 30), (Q19, 30), (Q23, 30), (Q14, 30), (Q19, 20), (Q23, 30), (Q19, 10), (Q14, 30), (Q6, 30)	62	05	420	(Q8, 40), (Q14, 40), (Q29, 40), (Q18, 40), (Q14, 40), (Q25, 30), (stall, 2), (Q25, 40), (Q23, 30), (Q18, 40), (stall, 2), (Q17, 30), (Q8, 50)
Experiment D							
63	01	420	(stall, 12), (Q8, 30), (Q32, 30), (Q17, 40), (Q23, 30), (Q32, 30), (Q17, 30), (Q23, 20), (Q32, 30), (Q41, 30), (Q32, 30), (Q17, 30), (Q32, 90)	81	01	420	(stall, 6), (Q17, 50), (Q19, 50), (stall, 12), (Q17, 50), (Q29, 60), (Q1, 50), (stall, 8), (Q41, 50), (Q40, 50), (stall, 12), (Q17, 60)
64	02	180	(Q36, 60), (Q32, 60), (Q29, 60)	82	02	180	(stall, 24), (Q17, 30), (Q32, 30), (Q17, 30), (Q32, 30), (Q17, 30), (Q32, 30)
65	03	180	(Q29, 60), (Q32, 60), (Q36, 60)	83	10	180	(Q29, 30), (Q32, 30), (Q29, 30), (Q32, 30), (Q29, 30), (Q32, 30)
66	04	180	(Q29, 60), (Q32, 60), (Q29, 60)	84	11	180	(Q14, 180)
67	05	300	(Q1, 150), (Q19, 110), (Q17, 40)	85	05	300	(Q1, 30), (Q19, 120), (Q1, 150)
68	06	300	(Q1, 150), (Q23, 110), (Q1, 40)	86	06	300	(Q1, 30), (Q23, 120), (Q1, 150)
69	07	300	(Q1, 150), (Q17, 110), (Q6, 40)	87	07	300	(Q6, 30), (Q11, 120), (Q6, 150)
70	08	300	(Q6, 150), (Q11, 110), (Q6, 40)	88	08	300	(Q39, 100), (Q38, 200)
71	09	300	(Q6, 150), (Q10, 110), (Q6, 40)	89	09	300	(Q38, 50), (Q40, 50), (Q41, 50), (stall, 36), (Q38, 50), (Q40, 50), (Q41, 50)
72	10	180	(Q5, 180)	90	03	180	(Q17, 80), (stall, 12), (Q19, 100)
73	11	180	(Q11, 180)	91	04	180	(Q17, 40), (stall, 24), (Q19, 140)
74	12	180	(Q12, 180)	92	12	180	(Q17, 40), (stall, 12), (Q32, 140)
75	13	180	(Q19, 180)	93	13	180	(Q17, 80), (stall, 24), (Q32, 100)
76	14	180	(Q23, 180)	94	14	180	(Q1, 40), (stall, 4), (Q8, 30), (stall, 4), (Q17, 20), (Q23, 30), (stall, 6), (Q29, 30), (stall, 4), (Q29, 30)
77	15	180	(Q29, 180)	95	15	180	(Q17, 40), (stall, 12), (Q17, 60), (stall, 24), (Q17, 80)
78	16	180	(Q32, 180)	96	16	180	(Q17, 40), (stall, 24), (Q17, 60), (stall, 12), (Q17, 80)
79	17	180	(Q38, 180)	97	17	180	(Q9, 180)
80	18	180	(Q1, 180)	98	18	180	(Q13, 180)
Experiment E							
99	01	420	(Q45, 50), (Q47, 30), (Q44, 30), (Q42, 60), (Q47, 70), (Q45, 50), (Q45, 110), (Q42, 20)	117	01	420	(Q17, 60), (Q23, 30), (Q8, 30), (Q28, 50), (Q17, 40), (Q24, 50), (Q28, 50), (Q17, 80), (Q23, 30)
100	13	180	(Q43, 180)	118	15	180	(Q17, 40), (stall, 12), (Q17, 60), (stall, 24), (Q17, 80)
101	14	180	(Q41, 180)	119	12	180	(Q17, 80), (stall, 12), (Q19, 100)
102	17	180	(Q47, 180)	120	04	180	(Q17, 40), (stall, 24), (Q19, 140)
103	08	300	(Q42, 20), (stall, 8), (Q42, 60), (Q46, 120), (Q47, 50), (Q43, 50)	121	06	300	(Q1, 40), (stall, 4), (Q8, 30), (stall, 4), (Q17, 20), (Q23, 30), (stall, 6), (Q29, 30), (stall, 4), (Q29, 30), (Q28, 50), (Q17, 70)
104	06	300	(Q44, 40), (stall, 4), (Q47, 30), (stall, 4), (Q47, 50), (Q42, 30), (stall, 24), (Q42, 30), (Q45, 50), (Q47, 70)	122	09	300	(Q36, 50), (Q28, 50), (Q24, 50), (stall, 36), (Q36, 50), (Q28, 50), (Q24, 50)
105	07	300	(Q41, 50), (Q47, 50), (stall, 36), (Q47, 50), (Q43, 100), (stall, 6), (Q45, 50)	123	08	300	(Q1, 30), (Q30, 120), (Q1, 150)
106	09	300	(Q44, 60), (Q42, 120), (Q47, 120)	124	05	300	(stall, 9), (Q1, 150), (Q19, 110), (Q17, 40)
107	05	300	(stall, 6), (Q44, 150), (Q44, 110), (Q47, 40)	125	07	300	(Q17, 80), (stall, 8), (Q23, 120), (Q28, 50), (Q24, 50)
108	10	180	(Q45, 180)	126	18	180	(Q17, 40), (stall, 12), (Q32, 140)
109	11	180	(Q43, 180)	127	02	180	(stall, 6), (Q17, 80), (stall, 18), (Q32, 100)
110	18	180	(Q44, 180)	128	16	180	(Q19, 180)
111	19	180	(Q48, 180)	129	03	180	(Q23, 180)
112	03	180	(Q42, 180)	130	13	180	(Q29, 180)
113	15	180	(Q44, 80), (Q41, 50), (stall, 12), (Q43, 50)	131	17	180	(Q32, 180)
114	04	180	(Q45, 40), (Q42, 40), (stall, 24), (Q41, 100)	132	14	180	(Q30, 180)
115	16	180	(Q45, 40), (stall, 12), (Q48, 140)	133	10	180	(Q9, 180)
116	02	180	(stall, 18), (Q45, 80), (stall, 8), (Q42, 100)	134	11	180	(Q14, 180)

quality are obtained independently using closed-form models described in Equations 1-4 with r being the frame rate and s being the number of pixels in a video frame. Once audio and video quality are estimated, audiovisual quality estimates are obtained by merging quality scores using 5. In these Equations, v_{1-7}, a_{1-3}, and m_{1-4} are model coefficients. Note that in previous work [51], model coefficients were obtained for users watching videos on TV. However, videos on TVs or mobile devices have different visibility in terms of coding artifacts, which has thus required two additional separate quality evaluation tests to be run for retraining and validating the model for a scenario where users watch videos on mobile devices. As this is not the main focus of this paper, these two experiments are not described and only retrained coefficients are listed in Table IV.

$$MOS_v = X + \frac{1 - X}{1 + (\frac{b_v}{Y})^{v_1}} \tag{1}$$

$$X = \frac{4 \times (1 - e^{-v_3 \times r}) \times s}{v_2 + s} + 1 \tag{2}$$

$$Y = \frac{v_4 \times s + v_6 \times log_{10}(v_7 \times r + 1)}{1 - e^{-v_5 \times s}} \tag{3}$$

TABLE III
SPATIAL AND TEMPORAL INFORMATION OF THE SRCS

SRC	Max SI	Mean SI	Max TI	Mean TI
1	51.41	28.22	81.49	4.96
2	89.56	54.61	79.94	6.91
3	136.36	69.54	99.05	11.72
4	89.27	44.64	104.15	12.12
5	103.31	54.72	95.24	20.24
6	94.41	45.62	106.66	6.94
7	97.26	46.93	86.39	7.34
8	96.05	37.63	104.19	6.57
9	99.86	44.82	95.11	9.27
10	54.21	37.44	85.55	4.22
11	67.80	52.97	24.05	8.81
12	88.88	43.44	108.87	5.22
13	133.74	61.44	91.25	13.53
14	84.28	39.02	104.50	16.77
15	81.78	49.45	75.20	10.15
16	61.83	41.22	97.31	3.78
17	83.75	36.97	87.85	11.12
18	102.85	57.08	93.78	13.75
19	140.07	85.59	84.02	12.39

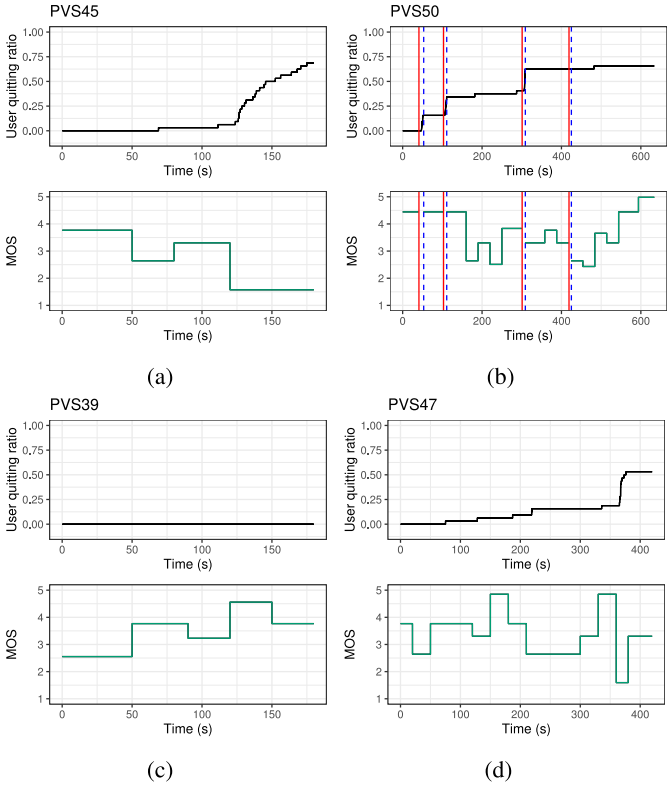


Fig. 1. Temporal evolution of users quitting the videos. MOS computed using computational algorithm [7], [51]. Vertical red lines and dashed blue lines are beginnings and ends of stalling events.

TABLE IV
COEFFICIENTS FOR PREDICTING AUDIOVISUAL QUALITY OF VIDEO IN A MOBILE CONTEXT USING THE MODEL DESCRIBED IN [51]

HEVC				
v_1	v_2	v_3	v_4	v_5
0.986848842	115397.7115	0.128419476	5.79E-05	0.99697
v_6	v_7	a_1	a_2	a_3
229.8988474	1.490889043	4.964967	16.461	2.081840
m_1	m_2	m_3	m_4	
0.000	0.116041	0.524354	0.092393	
AVC				
v_1	v_2	v_3	v_4	v_5
1.635491012	108471.168	0.098819619	0.000186712	0.996968
v_6	v_7	a_1	a_2	a_3
10822.08877	0.003642812	4.964967	16.461	2.081840
m_1	m_2	m_3	m_4	
0.000	0.116041	0.524354	0.092393	

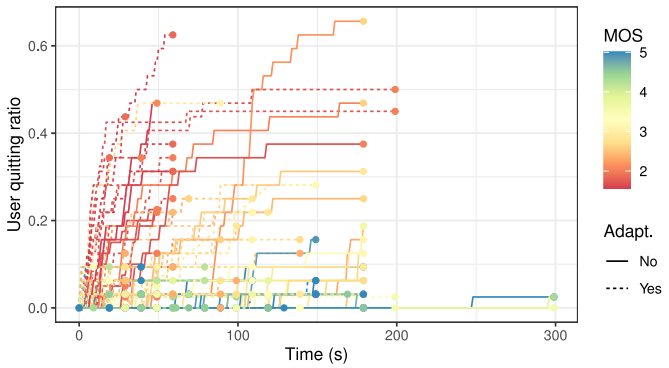


Fig. 2. Temporal evolution of users quitting the videos for coding conditions only with regards to MOS (MOS predicted using [7], [51]). Results reported per video segment of constant quality. Dashed and solid lines respectively indicate if a quality adaptation has occurred before the considered segment.

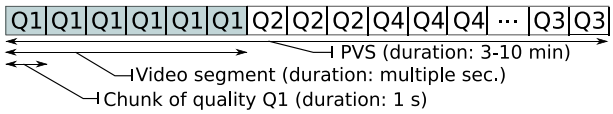


Fig. 3. Notations: Differences between chunk, video segment, and PVS.

$$MOS_a = a_1 + \frac{1 - a_1}{1 + (\frac{b_a}{a_2})^{a_3}} \tag{4}$$

$$MOS = m_1 + m_2 \times MOS_a + m_3 \times MOS_v + m_4 \times MOS_a \times MOS_v \tag{5}$$

On the basis of these quality estimates, Figure 2 gives an overview of the user quitting ratio for all coding conditions. Each segment of lines in the figure corresponds to a part of a PVS that had a constant quality (hereinafter referred to as a video segment of constant quality). Figure 3 illustrates this notation: a PVS can be decomposed into small pieces of video that are the chunks. Each chunk corresponds to a small portion of the video encoded at a given quality. A video segment of constant quality is then defined as one or multiple consecutive chunks at the same quality level. Video segments hence have variable duration (in the experimental data, from 10 to 300 s with an average duration of 67.78 s). To better visualize the effect of coding on user quitting ratio, Figure 2 shows the relationship between MOS and user quitting ratio as a function of time for each segment of constant quality. Each line segment is a segment of constant quality, and the graph shows the increase in the quitting ratio since the segment of constant quality started. Hence, each line segment in this graph starts from (0,0). Note that stalling is omitted, and the end of a stalling event would initiate a new segment of constant quality. In addition, in this figure, line segments are either represented using dashed and solid lines to respectively highlight cases where the segment of constant quality occurred after a quality adaptation and where no quality adaptation occurred before. According to this data, decreasing MOS increases the *user quitting ratio*, and low quality after a quality adaptation results in the *user quitting ratio* increasing more sharply.

To further investigate quality changes, Figure 4(a) illustrates a quality change where video coding quality is constant and equal to a MOS value M_1 until the time t_c where MOS decreases to

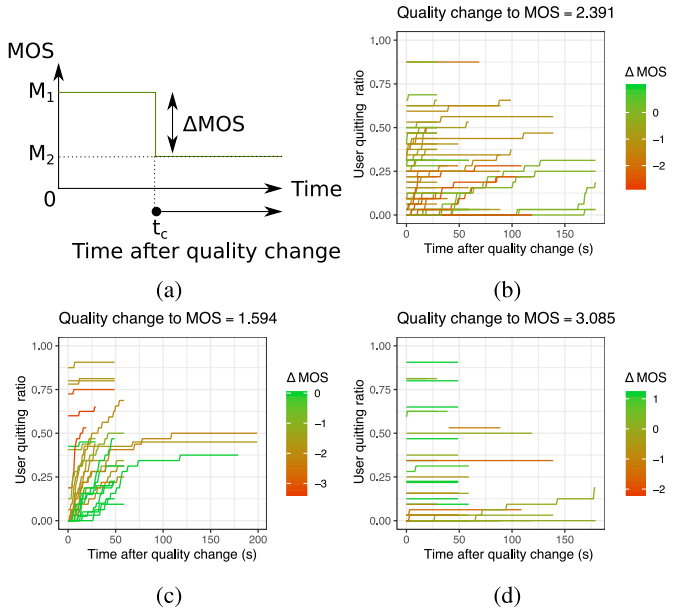


Fig. 4. Illustration of quality adaptation (a), and its impact on quitting ratio (b), (c), and (d).

M_2. The change in quality is noted as ΔMOS and defined by $\Delta MOS = M_2 - M_1$.

Figure 4(b) and (c) show the effect of ΔMOS on the increase of the user quitting ratio as a function of time since the segment of constant quality stated. In these graphs, color encodes the value of ΔMOS. In each plot, the quality after the change, noted as M_2 in Figure 4(a), is kept constant (at ± 0.05 unit of MOS). This enables only the effect of ΔMOS on the *user quitting ratio* to be tested. Figure 4(c) shows that for a constant low quality $M_2 \approx 1.6$, the quitting ratio depends on ΔMOS and increases further as quality largely changes. In addition, when many users have already left, the quitting ratio only slightly increases even though MOS can be low and ΔMOS is large. This can be observed in Figure 4(b) where the quitting ratio is initially above 75% and remains constant even though MOS is low and ΔMOS is large. Finally, Figure 4(d) shows cases with $\Delta MOS < -1.5$, along with a low initial quitting ratio, but the quitting ratio remains constant. This shows that even when quality greatly changes, if M_2 is high enough ($M_2 \approx 3.1$ in this case), quitting does not increase.

To summarize this analysis, the quitting ratio increases when coding quality is low and further increases when quality changes largely. The temporal evolution of the quitting ratio depends on the initial quitting ratio value as identified in the segment of constant quality analysis. Finally, quality changes do not necessarily increase the quitting ratio, and quitting ratio also depends on the quality level after the change (M_2).

C. Quitting Ratio and Stalling Events

The second main characteristic that needs to be addressed in this work is the effect of stalling and initial buffering on the user quitting ratio. Figure 5(a) shows an example of the evolution of the user quitting ratio as a function of time (notations are

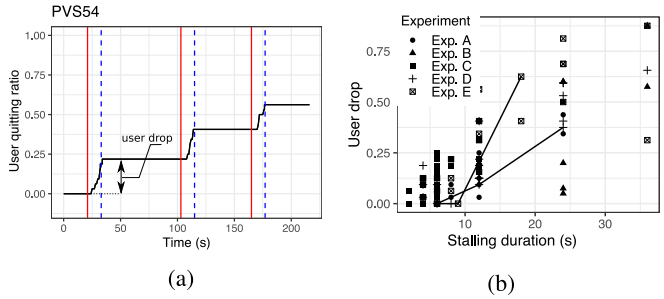


Fig. 5. Effect of stalling on user quitting ratio. Lines on the graph (b) highlight the relationship between initial buffering condition duration and quitting ratio in experiments D and E.

identical to those in Figure 1). The magnitude of quitting because of stalling will be hereinafter referred to as a "user drop."

Figure 5(b) shows the relationship between user drop and stalling duration. When stalling duration increases, user drop increases but with various amplitudes. In our previous work, we showed that other relevant factors are the stalling position, the audiovisual quality before the stalling occurred, and the quitting ratio at the time of the stalling [29], [54]. Section V-B on modeling will further elicit the relationship between them.

One novelty in this paper can be easily identified in the special case of initial buffering, which was tested in experiments D and E and is marked in Figure 5 by lines segment. It can be seen that user drop increases non-linearly with the initial buffering duration (as expected from past studies [26], [28]), but large differences between experiments D and E are found. Indeed, in experiment E, users reacted more sharply to initial buffering than in experiment D due to the differences in total stalling duration between experiments D and E (272 and 341 s, respectively). This shows that even in the simplest case of initial buffering (fixed stalling position, fixed quality, and initial quitting ratio of 0), there is more to consider than the properties of stalling events themselves, and overall session-based features should be considered. This makes the work of testing initial buffering as proposed previously [24], [25] more difficult as multiplying stalling conditions biases the results. This phenomenon is quantified in Section VII-D.

V. Modeling of User Quitting Ratio

In the following, a model of the *user quitting ratio* is described.

A. Coding Module

First, the effect of coding on the user quitting ratio is given in 6. It is defined as depending on time, $t \in [0, +\infty]$, and provides the user quitting ratio bounded in $[0, 1]$. It should be 0 at the beginning of the PVS, and 1 when time tends to the infinite. In 6, t is the time since the beginning of the PVS, and t_c is the time when the playback of the video segment of constant quality starts. Therefore, $t - t_c$ provides the time elapsed since the beginning of the video segment. In this equation, λ accounts for the effect of features identified in Section IV-B on the user quitting ratio. As audiovisual contents are used, λ considers both

video (MOS_v) and audio (MOS_a) quality along an interaction term as usually done in QoE estimation models [55]. Although the QoE estimation model used in this work ([7], [51]) can provide audiovisual QoE estimation, it was chosen to re-optimize the weight for audio and video components as their importance may vary depending on the objective: QoE or quitting ratio. In addition to the coding quality, λ accounts for the quitting ratio when the segment of constant quality starts, which is referred to as $D(t_c)$. It also accounts for the video quality variation between the current segment of constant quality and the previous one referred to as ΔMOS_v. ΔMOS_v is null at the beginning of the video. To avoid division by zero, λ is minimized by $\epsilon = 0.0001$ (value defined empirically), and C_{1-6} are the parameters.

$$D(t) = 1 - (1 - D(t_c)) \times e^{-\frac{t-t_c}{\lambda}} \tag{6}$$

$$\begin{aligned}\lambda = max(\epsilon, C_1 + C_2 \times MOS_v + C_3 \times MOS_a \\ + C_4 \times MOS_v \times MOS_a + C_5 \times \Delta MOS_v \\ + C_6 \times D(t_c))\end{aligned} \tag{7}$$

Finally, it should be stressed that the module handles interaction between past quality impairment events and the current condition in two ways: the first is with the gradient of quality (ΔMOS_v), which addresses the consequence of the immediate quality change, and the second is by using $D(t_c)$, which provides a feature about the long-term aspect of quality on the user quitting ratio.

B. Stalling Module

Regarding stalling, previous works have studied the impact of stalling on the user quitting behavior and have identified that *user drop* could be estimated using a linear combination between the stalling duration and stalling position [29]. However, this model was identified to only be able to predict quitting from stalling events where no users had already quit the video when the stalling occurred [29]. Further research then identified an interaction term between the user quitting ratio before the stalling and stalling characteristics. Indeed, if no users left the video before the stalling, user drop significantly increased with the stalling position. However, if users had already left, user drop significantly decreased with the stalling position. In addition to these results, past studies have shown an interaction between MOS and the response of users in terms of quitting due to stalling. To account for these phenomena, the method of Lebreton *et al.* [29] was further extended to account for the effect of MOS and the quitting ratio when the stalling occurred (noted, t_s, and corresponding to $D(t_s)$) on the estimation of user drop [54]. 8 gives the method to estimate user drop resulting from these works. The equation accounts for the contribution of the stalling duration (S_d) and stalling position (S_p), both provided in seconds, and their respective non-linear interactions with $D(t_s)$ and the audiovisual quality (MOS). S_{1-8} are model parameters.

$$\begin{aligned}U = S_1 + S_d \times e^{S_2 + S_8 \times D(t_s)} \\ + S_3 \times S_p + S_4 \times M \times [D(t_s)]^{S_6} \\ + S_5 \times MOS \times S_p \times [D(t_s)]^{S_7}\end{aligned} \tag{8}$$

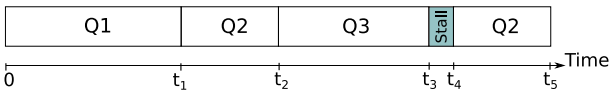


Fig. 6. Example of a viewing session with quality adaptation and stalling events.

C. Combined Model for Quitting Ratio Prediction

Once the coding and stalling module are defined, a model is introduced to predict the user quitting ratio with mixed quality adaptation and stalling conditions. It is made by combining the two previous modules from Sections IV-B and IV-C in an iterative process. Figure 6 illustrates a PVS with two consecutive quality changes: one occurring at t_1 with a quality change from Q1 to Q2 and the other at t_2 with a quality change from Q2 to Q3. In this scenario, the computation of the user quitting ratio is as follows. The user quitting ratios for each segment of constant quality are computed one after the other using 6. First, when $t \in [0, t_1]$, $D(0) = 0$. Therefore, 6 can be simplified into 9. Then, to compute the quitting ratio values with $t \in [t_1, t_2]$, 6 is used with $t_c = t_1$. In this step, λ (see 7) accounts for both the quitting ratio at the beginning of the segment ($D(t_1)$), and the difference in perceptual quality between consecutive segments (ΔMOS_v). Therefore, the computation of the quitting ratio $t \in]t_1, t_2]$ also depends on past computations.

$$\forall t \in [t_0, t_1], D(t) = 1 - e^{-\frac{t}{\lambda}} \tag{9}$$

In the example presented in Figure 6, a stalling event occurs at t_3 and lasts until t_4. To predict the user quitting ratio in the interval $]t_3, t_4]$, 8 is used where $D(t_3)$ MOS, S_p and S_d are replaced with their respective values on the basis of the previous computations as described in 10. In this scenario, MOS_{Q3} corresponds to the estimate of the audiovisual quality of the quality level "Q3," which corresponds to the quality of the video before the stalling occurs. In the special case of an initial buffering event, the value of MOS is set to 5, as high quality would be expected.

$$\begin{aligned}\forall t \in]t_3, t_4], D(t) = D(t_3) + \frac{t - t_3}{t_4 - t_3} \times \\ \{+S_1 + (t_4 - t_3) \times e^{S_2 + S_8 \times D(t_3)} \\ + S_3 \times t_3 + S_4 \times MOS_{Q3} \times [D(t_3)]^{S_6} \\ + S_5 \times MOS_{Q3} \times t_3 \times [D(t_3)]^{S_7}\}\end{aligned} \tag{10}$$

By using this approach, the user quitting ratio can then be estimated at any time for any session containing quality adaptation and stalling event while also accounting for their respective interactions.

VI. Performance Evaluation

This section addresses performance and training procedures.

A. Coding Module Performance

To train the coding module, special care is needed as there are two cases of user quit events: when users reach the end of

TABLE V
PERFORMANCE OF INDIVIDUAL MODULES ACROSS DIFFERENT TRAINING AND VALIDATION (REPARTITION MARKED BY CROSSES)

	Experiment					Coding module		Stalling module	
Phase	A	B	C	D	E	RMSE	PCC	RMSE	PCC
Training	x	x	x	x	x	0.07790	0.92564	0.08749	0.9184
Training		x	x	x		0.08298	0.91619	0.06440	0.9500
Validation	x				x	0.07029	0.94336	0.11589	0.8785
Training	x	x	x			0.07729	0.94407	0.06460	0.9344
Validation				x	x	0.08194	0.92105	0.11627	0.9094
Training			x	x		0.08184	0.91055	0.06490	0.9220
Validation	x	x			x	0.07106	0.94556	0.11151	0.9030
Training	5-fold cross validation					0.083520	0.90699	0.09241	0.9025
Validation	5-fold cross validation					0.070414	0.94542	0.09386	0.8957

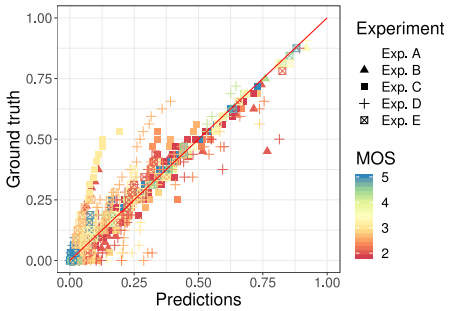


Fig. 7. Prediction accuracy of the quitting ratio module for coding conditions. Plot based on training performed on Exp. C, D, and E.

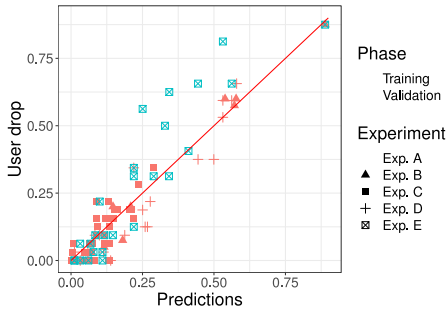


Fig. 8. Stalling module prediction accuracy. Plot based on training performed on Exp. B, C, and D and validated on A and E.

TABLE VI
PERFORMANCE OF THE MODEL WITH BOTH QUALITY ADAPTATION AND STALLING EVENTS. "OVERALL RESULT" AND "W.O. PVSS 36, 43, 74" DIFFERS ON WHETHER PVSS 36, 43 AND 74 ARE CONSIDERED OR NOT

	Experiment					Overall. result		w.o. PVSs 36, 43, 74	
Phase	A	B	C	D	E	RMSE	PCC	RMSE	PCC
Training	x	x	x	x	x	0.09716	0.8706	0.09010	0.8933
Training			x	x		0.09411	0.8677	0.08387	0.9000
Validation	x	x			x	0.10080	0.8824	0.09712	0.8925
Training	x	x		x	x	0.09834	0.8725	0.09134	0.8939
Validation			x			0.09888	0.8540	0.09288	0.8814
Training			x	x	x	0.09753	0.8577	0.08778	0.8910
Validation	x	x				0.09841	0.8830	0.09480	0.8946
Training	5-fold cross validation					0.09863	0.8643	0.09069	0.8900
Validation	5-fold cross validation					0.09421	0.8776	0.08989	0.8945

the video or quit midway through. To train the quitting ratio module, only cases where the users quit midway through are considered as the end of the video is independent of the decision of the user. The collected data contain 542 video segments, and 890 events of users quitting midway through that can be used to train the module. The module was trained with non-linear least square regression using the optimizer *nls* from the statistical software *R* across different combinations of training and validation sets. At this point in the analysis, the quitting ratio at the beginning of the segment is based on ground truth data. Table V lists performance results in terms of Root Mean Square Error (RMSE) and Pearson Correlation Coefficient (PCC) for different trainings and validations across experiments. A 5-fold cross-validation test is provided and based on splitting the 134 PVSs into training and validation sets.

Figure 7 depicts the performance accuracy of the proposed model. Across the 134 PVSs, the quitting ratio of three PVSs (PVS36, PVS43, and PVS74) was largely underestimated as content dependency issues were identified in the audio and video perceived quality and perceived quality changes were identified (further details in Section VII). To conclude, although simple, the temporal evolution of the user quitting ratio for coding conditions was accurately predicted for most cases using only a parameter-based model as required for encrypted video bitstream.

B. Stalling Module Performance

Across the 5 experiments, 104 stalling events were encountered by the participants. These ranged from 2 to 36 s with an average duration of 12.86 s. To train the stalling module, data was split into training and validation datasets, and the module was trained using non-linear least square regression using the optimizer *nls* from the statistical software *R*. Table V reports the performance across various combinations of training and validation either across experiments or by using cross-validation. This table shows that the module achieved consistently good accuracy, and Figure 8 illustrates the performance when the model was trained on experiments B, C, and D and validated on A and E. These results show that the model provides an overall good accuracy but under-predicts the effect of long stalling events from experiment E. Indeed, as previously described, user sensitivity to stalling was higher in experiment E and shows that stalling events should not be considered individually but as part of an overall viewing session involving all PVSs seen in the experiment. Further details are given in Section VII.

C. Overall Model Performance

Similarly to previous analysis, the overall model was evaluated by splitting training and validation across datasets and cross-validation analysis (by splitting data on a per-PVS basis). Due to the iterative, time-dependent, and non-linear properties of the model, the coding and stalling modules should be trained jointly, and this was achieved using the generalized reduced gradient (GRG) non-linear optimizer from Microsoft Excel. Table VI provides a quantitative performance analysis of the overall model, and Figure 9 illustrates the prediction accuracy. From these results, across different trainings and validations, the model performs consistently and with good overall accuracy. Table VI reports RMSE and PCC across the different training and validation combinations with and without the PVSs 36, 43, and 74. These were identified as more challenging than

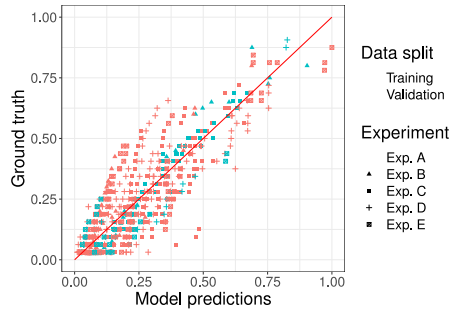


Fig. 9. Overall model performance. Training based on random selection of PVSs during 5-fold cross-validation.

TABLE VII
MODEL COEFFICIENTS

Coefficient	S_1	S_2	S_3	S_4	S_5
Estimate	-0.12788	-3.67803	0.00085314	0.056463	-0.00030539
Coefficient	S_6	S_7	S_8	C_1	C_2
Estimate	0.2600	0.2000	-2.23901	4271.17309	-3911.3628
Coefficient	C_3	C_4	C_5	C_6	ϵ
Estimate	-1118.0445	1034.16072	25.3443875	116.951587	0.0001

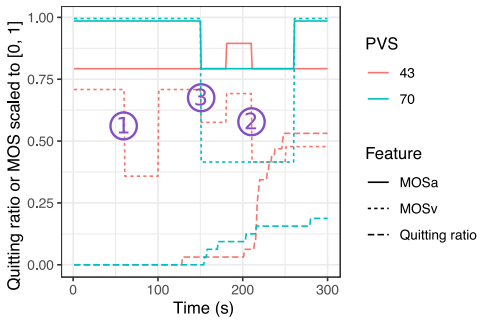


Fig. 10. Impact of content dependency aspects on the user quitting ratio.

in Section IV-B as differences in coding complexity made estimations of audiovisual quality less accurate. This shows one of the challenges of the proposed iterative approach, as the previous computation serves as input for subsequent steps and there is a risk of propagating errors. Nevertheless, the proposed model was able to overcome these challenges in most cases. Final coefficients are listed in Table VII.

VII. DISCUSSION

The proposed model is the first of its kind that aims at predicting the user quitting ratio and its temporal evolution on a per-PVS basis as a function of quality-related factors. This is a challenging task, and across the model development, several points of interest were identified and addressed bellow.

A. Challenging Scenarios

One challenge with predicting the user quitting ratio was that perceived quality and quality change estimations were inaccurately predicted. Indeed, as shown in Figure 7, this led to several PVSs being inaccurately estimated. To demonstrate this, Figure 10 show two PVSs: one accurately estimated (PVS70) and one with large error (PVS43). Although these two PVSs share similar characteristics, the ground truth quitting ratio data differs largely (see long dashed lines in the figure). For PVS43, the event marked as (1) shows a large change in video quality that did not result in an increase in the user quitting ratio, while the quality change (2) although similar to (1), resulted in a large increase in user quitting ratio. One difference between events (1) and (2) is that event (2) had a change in audio quality as depicted by the continuous line in the figure. However, if PVS70 is considered at the time of event (3), the quitting ratio only slightly increases although both video quality and audio quality change in the same way as in event (2). This example shows the issue of content dependency and visibility of audio and video coding degradation on the user quitting ratio. In this example, audio quality change was more perceivable in PVS43 than in PVS70 as it occurred during speech instead of music and may justify the differences. Similarly, title screens in videos have a very low coding complexity and resulted in underestimated video quality and overestimated user quitting ratio. The model does not consider content dependency, so it failed to address these issues and resulted in incorrect predictions shown in Figure 7. However, these are constraints from working with encrypted bitstreams.

B. Content Duration

Content duration is widely identified as an important factor affecting abandonment [1], [30], [31], [33], [35], [39], [42]. In this work, the user quitting ratio was found to increase with time, and longer PVS may result in a higher quitting ratio and lower completion ratio [31], [39], [41], but these were only because longer time spans are used. The temporal evolution of the user quitting ratio was found to be similar across content with various durations, and no evidence of fundamental differences between 3- to 10-min contents could be found.

C. Interest Towards Contents

As discussed in Section II-A3, in real-life settings, user abandonment depends on the context and users' interest in content. This work proposed addressing the content dependency by explicitly requesting users to quit the videos only because of quality-related issues and not because of a lack of interest in the content. However, it is worth mentioning that content may have unconsciously affected users' decisions to quit as having interest or not in the content may have made them more quickly react to quality impairments. A possible approach to quantify this issue could be to ask users whether they were interested into the content after choosing to quit (because of low quality) or fully watching the video. This approach would then enable us to test the hypothesis that users' behavior is unconsciously affected by users' level of interest. In this work, no such question was asked, so this issue is difficult to address. Researchers further studying this topic may be advised to ask such questions. In this work, this issue may have been mitigated by the repeated design where multiple participants each having individual interests were involved. Therefore, some may have been more sensitive and others less sensitive to content-related issues, resulting in an average quitting ratio across participants. This will be further tested in future research.

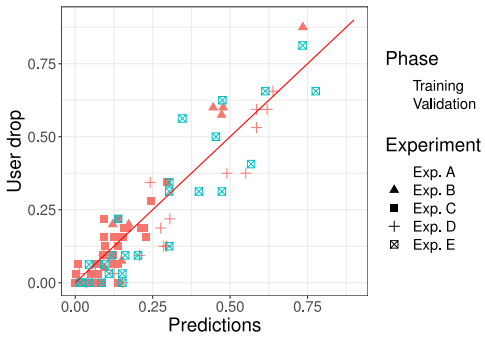


Fig. 11. Prediction of user drop due to stalling while accounting for total stalling duration (11).

D. Dataset Bias in Stalling Conditions

In our experiments, users in experiment E reacted more strongly to stalling than users in other experiments. To explain this phenomenon, differences between experiments were analyzed, and it was found that because experiment E had a longer total stalling duration than other experiments (178 s in Exp. A, 204 in Exp. B, 256 in Exp. C, 272 in Exp. D, and 341 in Exp. E), users became more sensitive. To further investigate this, the total stalling duration per experiment was considered as an additional feature to 8. 11 is then defined and scales the model prediction by the total stalling duration per experiment. In this regression analysis, coefficients S_{1-8} from 8 are frozen and kept to their value identified during the previous training (see Table VII). Therefore, only a scaling factor as a function of the overall stalling duration is trained. Figure 11 shows the results of the adjusted model after only training coefficient $S_9 = 0.0\,040\,628$. This plot is directly comparable with Figure 8 as it uses the same coefficients and shows that considering the total stalling duration of the test enables a large performance improvement. RMSE at validation was then improved going from 0.11 589 to 0.09 488 (validation done on Exp. A and E), which shows improvement across all experiments. Therefore, session-related factors exist. These results can be directly applied to the proposed model from Section V-C by swapping the estimation of *user drop* from stalling using 8 by 11 and will enable the dataset bias to be handled. However, note that applying these results to real-life monitoring is not trivial as total stalling duration experienced by the users is unknown. Viewing session based analysis could be considered, but it will require further research.

$$U_b = S_9 \times S_{d,total} \times U \tag{11}$$

VIII. Comparison With Alternative Approaches

Although the user quitting ratio has never been addressed before, this work is related to the estimation of the Cumulative Distribution Function (CDF) of the viewing time on a per-PVS basis. However, although viewing time CDF has been studied [16], [26], [41], [47], [48], the analysis was always performed on a per-dataset basis and not on a per-PVS basis. This is problematic as per-dataset and per-PVS CDF have different properties, and traditional approaches such as Pareto distribution or Log-Normal distributions [16], [47] cannot be applied to per-PVS analysis. Indeed, as shown in Figure 1, there are discontinuities in the per-PVS CDF due to changes between the quality level and stalling events. Therefore, the quitting ratio does not follow these distributions. Doing a per-PVS analysis requires designing a mixture of multiple functions that needs to be parametrized on the basis of features about stalling and coding properties. This requires new models and corresponds to the work done in this paper that has not been done before.

Alternatively, survival analysis can be performed [50]. In survival analysis, the estimation of the user quitting ratio corresponds to estimating the cumulative hazard function. The hazard function can be based on quality-related features as in the Cox proportional-hazards model. However, as working on a per-PVS basis has introduced discontinuities in the temporal evolution of the quitting ratio, this type of model cannot be applied at the PVS-scale as it would violate the time-independence criteria of covariates. Therefore, the discontinuities in the quitting ratio need to be addressed by splitting the PVSs per segment and designing a piecewise model. This relates to the work conducted in this paper and has inspired the module for low quality coding conditions proposed in this paper. Regarding stalling events, due to them lasting only a couple of seconds, the temporal aspects of the prediction of quitting ratio due to stalling can be easily approximated with a linear function without involving complex survival analysis modeling.

Finally, another candidate alternative is to use the Long Short-Term Memory (LSTM) recurrent neural network architecture to estimate probability density functions [49]. However, this type of modeling is challenging to apply to this work for multiple reasons. First, the quality-related features are estimated using a no-reference parameter-based audiovisual quality estimation algorithm, which results in features having constant values when coding parameters are constant. Therefore, input features were found to be constant over long periods, making the fitting the LTSM-based model difficult. This requires the use of a large temporal window and leads to a second issue: a more complex model requires more data. Moreover, although many experiments were conducted in this work, LTSM-based models are challenging to train and lead to poor prediction accuracy.

IX. Conclusion

This paper described a study aiming to understand the characteristics of the user quitting ratio and use this knowledge to develop a computational model to predict the user quitting ratio.

It was shown that when coding quality becomes low, users quit the video. Quitting was found to be amplified by quality changes, but the magnitude of the quality change is not the only factor, and quality after the change occurred is also a significant factor. The quitting ratio as a function of time was also found to depend on previous quitting ratio values and highlight interactions between quality events. Another key result from this work was to show that the impact of stalling events on user quitting ratio depends on more parameters than the stalling properties themselves, and the entire session should also be considered. This is a novel point that was quantified thanks to the large number of subjective experiments done in this work. In this study, the quitting ratio

for a stalling event scaled linearly with the total stalling duration of the experiments.

Then, these results were used to design a computational model of user quitting ratio that can handle low coding quality, stalling, and quality changes. The model can be used in monitoring scenarios where low complexity is needed and bitstream is encrypted. It is suitable for active real-time control of quality of videos watched by users. Although content dependency was still found to be challenging, the model provided consistently good results across multiple trainings and validations.

Future research will focus on optimizing services on the basis of engagement measurements and integrating the proposed user quitting ratio estimation into state-of-the-art service optimization algorithms (chunk selection algorithms for video players, bitrate selection algorithms for encoding videos, etc.). This will improve previous approaches, by considering quitting-based features instead of only quality-related features. Finally, testing and model performance will be improved. In subjective evaluations, the interest of the users towards content will be addressed. Then, further investigation will be pursued to better address the effect of quality changes, as well as dataset bias and session-based analysis. These studies will enable the overall prediction accuracy of the quitting ratio model to be improved.

REFERENCES

[1] Z. Shafiq *et al.*, "Understanding the impact of network dynamics on mobile video user engagement," in *SIGMETRICS*, vol. 42, pp. 367–379, Jun. 2014.

[2] Akamai, "Maximizing audience engagement: How online video performance impacts viewer behavior," White Paper, 2012.

[3] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior : Inferring causality using quasi-experimental designs," *ACM Tran. Netw.*, vol. 21, no. 6, pp. 2001–2014, 2013.

[4] K. Brunnström *et al.*, "Qualinet white paper on definitions of quality of experience," *Eur. Netw. Qual. Experience Multimedia Syst. Serv.*, 2012.

[5] ITU-T Recommendation P.1203, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport," ITU-T, 2017, pp. 1–22.

[6] W. Robitza, M. Garcia, and A. Raake, "A modular HTTP adaptive streaming QoE model - candidate for ITU-T P.1203 ("P.NATS")," in *Qual. Multimedia Experience, 9th Int. Conf.*, 2017, pp. 1–6.

[7] 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.

[8] C. G. Bampis and A. C. Bovik, "Learning to predict streaming video QoE: Distortions, rebuffering and memory," in 2017, *arXiv:1703.00633*.

[9] W. Robitza *et al.*, "Measuring YouTube QoE with ITU-T P.1203 under constrained bandwidth conditions," in *10th Int. Conf. Qual. Multimedia Experience*, 2018, pp. 1–6.

[10] L. Plissonneau, E. Biersack, and P. Julur, "Analyzing the impact of YouTube delivery policies on user experience," in *ITC Proc. 24th Int. Teletraffic Congr.*, 2012, pp. 1–8.

[11] D. Raca *et al.*, "Incorporating prediction into adaptive streaming algorithms: A QoE perspective," in *Proc. NOSSDAV ACM SIGMM Workshop Netw. Operating Syst. Support Digit. Audio Video*, 2018, pp. 49–54.

[12] B. Pan, X. Wang, C.-P. Hong, and S.-D. Kim, "AMVP-cloud: A framework of adaptive mobile video streaming and user behavior oriented video prefetching in the clouds," in *Proc. 12th Int. Conf. Comput. Inf. Technol.*, 2012, pp. 398–405.

[13] A. Balachandran *et al.*, "Developing a predictive model of quality of experience for internet video," in *SIGCOMM'13*, 2013, pp. 339–350.

[14] I. Ayad, Y. Im, E. Keller, and S. Ha, "A practical evaluation of rate adaptation algorithms in HTTP-based adaptive streaming," *Comput. Netw.*, vol. 133, pp. 90–103, 2018.

[15] S. Sahoo, M. Nidhi, K. S. Sahoo, B. Sahoo, and A. K. Turuk, "Video delivery services in media cloud with abandonment: An analytical approach," in *Proc. IEEE Int. Conf. Adv. Netw. Telecommun. Syst.*, 2017, pp. 1–6.

[16] V. Gopalakrishnan, R. Jana, and K. K. Ramakrishnan, "Understanding couch potatoes: Measurement and modeling of interactive usage of IPTV at large scale," in *ACM SIGCOMM Conf. Internet Meas.*, 2011, pp. 225–242.

[17] K.-W. Hwang *et al.*, "Leveraging video viewing patterns for optimal content placement," in *Proc. Int. Conf. Res. Netw.*, 2012, pp. 45–58.

[18] W. Robitza and A. Raake, "(Re-)Actions speak louder than words? a novel test method for tracking user behavior in web video services," in *Qual. Multimedia Experience*, 2016, pp. 1–6.

[19] A. Sackl, P. Zwickl, S. Egger, and P. Reichl, "The role of cognitive dissonance for QoE evaluation of multimedia services," in *Proc. GC'12 Workshop: Qual. Experience Multimedia Commun.*, 2012.

[20] W. Robitza, S. Schönfellner, and A. Raake, "A theoretical approach to the formation of quality of experience and user behavior in multimedia services," in *Proc. Workshop Perceptual Qual. Syst.*, 2016, pp. 39–43.

[21] M. Seufert *et al.*, "Unsupervised QoE field study for mobile YouTube video streaming with YoMoApp," in *Qual. Multimedia Experience*, 2017, pp. 1–6.

[22] M.-N. Garcia *et al.*, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. Qual. Multimedia Experience, 3rd Int. Conf.*, 2014, pp. 141–146.

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

[24] M. Zhibin, A. Raake, W. Robitza, and N. Zhangyan, "Training test results for G.QUIT and model structure," *Int. Telecommun. Union, Study Group 12, ITUI Contribution SG12-C370R2*, 2019.

[25] J. Song *et al.*, "Initial perceived quality analysis for dash video streaming," in *Proc. Vis. Commun. Image Process. Conf.*, 2018, pp. 1–4.

[26] S. Takahashi, K. Yamagishi, P. Lebreton, and J. Okamoto, "Impact of quality factors on users' viewing behaviors in adaptive bitrate streaming services," in *Qual. Multimedia Experience*, 2019.

[27] M. T. Diallo, F. Fieau, and J.-B. Hennequin, "Impacts of video quality of experience on user engagement," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2014, pp. 1–6.

[28] X. Tan, Y. Guo, M. Orgun, L. Xue, and Y. Chen, "An engagement model based on user interest and QoS in video streaming systems," *Wirel. Commun. Mobile Comput.*, vol. 2018, pp. 1–11, 2018.

[29] P. Lebreton, K. Kawashima, K. Yamagishi, and J. Okamoto, "Study on viewing time with regards to quality factors in adaptive bitrate video streaming," in *Proc. Workshop Multimedia Signal Process.*, 2018, pp. 1–6.

[30] H. Nam, K. Kim, and H. Schulzrinne, "QoE matters more than QoS: Why people stop watching cat videos," in *INFOCOM*, 2016, pp. 1–9.

[31] H. Nam and H. Schulzrinne, " YouSlow: What influences user abandonment behavior for internet video?" vol. 44, no. 4, pp. 1–14, 2014.

[32] A. Ahmed, Z. Shafiq, H. Bedi, and A. Khakpour, "Suffering from buffering? detecting QoE impairments in live video streams," in *Proc. IEEE Int. Conf. Netw. Protoc.*, 2017, pp. 1–10.

[33] Y. He, A. Wei, W. Zhang, and H. Xiao, "Understanding user behavior in large scale internet video service," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov.*, 2015, pp. 261–267.

[34] F. Dobrian *et al.*, "Understanding the impact of video quality on user engagement," *Commun. ACM*, vol. 56, no. 3, pp. 91–99, 2013.

[35] Y. Chena *et al.*, "Understanding viewer engagement of video service in wi-fi network," *Comput. Netw.*, no. 91, pp. 101–116, 2015.

[36] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming," in *IEEE/IFIP IM*, 2011, pp. 485–492.

[37] X. Wang, A. Wei, Y. Yang, and J. Ning, "Characterizing the correlation between video types and user quality of experience in the large-scale internet video service," in *Proc. 12th Int. Conf. Fuzzy Syst. Knowl. Discov.*, 2015, pp. 2086–2092.

[38] E. Aguiar, S. Nagrecha, and N. V. Chawla, "Predicting online video engagement using clickstreams," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics*, 2015, pp. 1–10.

[39] S. Wu, M. Rizoiu, and L. Xie, "Beyond views: Measuring and predicting engagement in online videos," in *Proc. Int. Conf. Weblogs Social Media*, 2018, pp. 1–10.

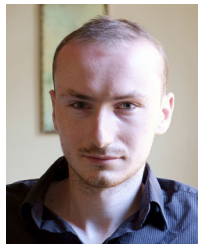
[40] H. Yan *et al.*, "On migratory behavior in video consumption," in *Proc. ACM Conf. Inf. Knowl. Manage.*, 2017, pp. 1109–1118.

[41] 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.

[42] Z. Li *et al.*, "Watching videos from everywhere: A study of the PPTV mobile vod system," in *Proc. 2012 Internet Meas. Conf.*, 2012, pp. 185–198.

[43] Y. Chen, B. Zhang, Y. Liu, and W. Zhu, "On distribution of user movies watching time in a large-scale video streaming system," in *IEEE ICC - Commun. Softw., Serv. Multimedia Appl.*, pp. 1825–1830, 2014.

[44] P. Reichl *et al.*, "Towards a comprehensive framework for QoE and user behavior modelling," in *Proc. Qual. Multimedia Experience, 4th Int. Conf.*, 2015, pp. 1–6.s
[45] C. Moldovan and F. Metzger, "Bridging the gap between QoE and user engagement in HTTP video streaming," in *Proc. 28th Int. Teletraffic Congr.*, vol. 1, 2016, pp. 103–111.
[46] L. Yang, M. Yuan, Y. Chen, W. Wang, Q. Zhang, and J. Zeng, "Personalized user engagement modeling for mobile videos," *Comput. Netw.*, vol. 126, pp. 256–267, 2017.
[47] Y. Li, Y. Zhang, and R. Yuan, "Characterizing user access behaviors in mobile TV system," in *Proc. IEEE ICC Commun. Softw. Serv. Multimedia Appl. Symp.*, 2012, pp. 2093–2097.
[48] Y. Li, Y. Zhang, and R. Yuan, "Measurement and analysis of a large scale commercial mobile internet TV system," in *IMC '11 Proc. ACM SIGCOMM Conf. Internet Meas.*, 2011.
[49] K. Yeo, I. Melnyk, N. Nguyen, and E. K. Lee, "DE-RNN: Forecasting the probability density function of nonlinear time series," in *Proc. IEEE Int. Conf. Data Mining*, 2018, pp. 697–706.
[50] R. G. Miller, *Survival Analysis*. Hoboken, NJ, USA: Wiley., 1997.
[51] 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. E102-B, no. 12, pp. 2226–2242, 2019.
[52] W. Robitza, "Bufferer: Inserts fake rebuffering events into video," 2017. [Online]. Available: https://github.com/slhck/bufferer
[53] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," ITU-T, 2008.
[54] P. Lebreton and K. Yamagishi, "Study on user quitting rate for adaptive bitrate video streaming," in *Proc. IEEE 21th Int. Workshop Multimedia Signal Process.*, 2019, pp. 1–6.
[55] M. H. Pinson, W. Ingram, and A. Webster, "Audiovisual quality components," *IEEE Signal Process. Mag.*, vol. 28, pp. 60–67, Nov. 2011.



Pierre Lebreton received the engineering Degree in computer science from Polytech Nantes, France, in 2009. In 2010, he was with the Group Assessment of IP-based Applications, TU-Berlin, where he is working toward the Ph.D. degree in 3D video QoE. After graduating in 2015, he joined the group of Audio Visual Technology with the TU-Ilmenau, Germany, in 2015 and the group of Networked Sensing and Control, Zhejiang University, China, in 2016. His work addressed various topics including aesthetic appeal, large scale video quality monitoring, and bike sharing systems. In 2017, he joined NTT Laboratories, where he now 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, in 2001 and the M.E. and Ph.D. degrees in electronics, information, and communication engineering from Waseda University, Japan, in 2003 and 2013, respectively. 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. He was the recipient of the Young Investigators Award (IEICE) 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.