Study on viewing completion ratio of video streaming

Pierre Lebreton and Kazuhisa Yamagishi
NTT Network Technology Laboratories
Musashino, Tokyo, Japan
{lebreton.pierre.mz,kazuhisa.yamagishi.vf}@hco.ntt.co.jp

***Abstract*—In this paper, a model is investigated for optimizing the encoding of adaptive bitrate video streaming. To this end, the relationship between quality, content duration, and acceptability measured by using the completion ratio is studied. This work is based on intensive subjective testing performed in a laboratory environment and shows the importance of stimulus duration in acceptance studies. A model to predict the completion ratio of videos is provided and shows good accuracy. By using this model, quality requirements can be derived on the basis of the target abandonment rate and content duration. This work will help video streaming providers to define suitable coding conditions when preparing content to be broadcast on their platform that will maintain user engagement.**

***Index Terms*—Acceptability, Completion ratio, Quality of Experience, Mobile video, Video streaming**

I. Introduction

To encourage users to use their services, video streaming service providers (e.g., Netflix, YouTube, Hulu, Twitch) need to ensure that the experience of the users is adequate. Therefore, to understand and improve the user's experience, Quality of Experience (QoE) has been researched [1], resulting in models able to predict QoE [2]–[6]. However, researchers has recently shifted their focus from QoE to User Engagement (UE). UE is of high interest for service providers as it gives information on whether users want to keep using the services. This gives a close-up view of income generated by a service and enables service providers to develop engagement control mechanisms to optimize services by considering users' willingness to use those services.

With this goal in mind, this paper addresses users' willingness to keep watching videos considering quality degradations due to coding and how it depends on content duration. Thus, the goal of this work is to study how videos should be encoded (resolution, bitrate, frame rate, etc.) so the video quality is high enough to keep the users engaged until the end. This differs from previous work, as the focus is not on QoE but on the completion ratio as a proxy for measuring acceptability of quality. The completion ratio provides a good understanding of whether users want to use the service considering its quality. A novel aspect of this work is to show how content duration should be considered when defining coding conditions. To this end, this study is based on laboratory experiments that enable the relationship between QoE, content duration, and completion ratio to be revealed. Finally, it is shown how existing QoE models can be used to predict the completion ratio while taking into account content duration.

978-1-7281-9320-5/20/$31.00 ©2020 IEEE

This paper is organized as follows: Section II describes the related work. Sections III and IV provide information on the evaluation procedure and experimental results, respectively. Section V describes a prediction model of the completion ratio. Section VI uses the model to identify quality requirements with respect to content duration. Section VII provides a discussion of the model, and Section VIII concludes this paper.

II. Related work

Multiple studies have been done to combine the work on QoE and acceptability. Acceptability can be defined as *a binary measure to locate the threshold of minimum acceptable quality that fulfills user quality expectations and needs for a certain application or system* [7]. ITU-T Recommendation P.10/G.100 specifies that this scale should be evaluated on the basis of a Yes/No answer that puts a low cognitive burden on the participants [8]. However, evaluating acceptability is not simple. Indeed, previous works have considered that what may be considered as an acceptable quality may not be enjoyable. Therefore, studies were conducted with different scales to address different concepts of acceptability. This was done by asking participants about not only "acceptability" but also "pleasing acceptability" [9] or whether the quality level was "annoying" [10]. Doing so enabled the steps leading from annoyance to unacceptability to be better understood. The transition from being unenjoyable to unacceptable will depend on stimulus duration. Therefore, this work considers how temporal aspects are involved in the construction of acceptability ratings. This is a new aspect that has not been addressed in previous studies.

In previous work, a strong correlation was found between "pleasing acceptability" and "acceptable quality." Pleasing acceptability can be achieved with a higher quality than an acceptable quality and can be derived from "acceptability" using a third order polynomial function [9]. However, authors did not consider content duration or study whether different coefficients are needed to map pleasing acceptability onto acceptable quality for different stimulus durations. Therefore, authors only focused on acceptability and developed multiple models on the basis of a five-parameter logistic function for various types of input (quantization parameter (QP), bitrate, peak signal-to-noise ratio (PSNR), or structural similarity index measure (SSIM)) [9]. Similarly, an acceptance model by mapping QoE scores estimated using the National Telecommunications and Information Administration (NTIA) Video Quality Metric (VQM) General Model [6] with a third order polynomial function was proposed by de Koning et al. [11]. Pessemier et al. [12] proposed an acceptability model based on a decision tree with parameters based on network-related features, watching behavior, and video quality. Other researchers [13]–[15] showed the relationship between acceptability and technical parameters such as bitrate, frame rate, resolution (but did not propose a model). Li et al. [10] categorized acceptability into three categories: acceptable, acceptable but annoying, and not acceptable. Then, a classification algorithm was proposed using quality predictions from a video quality prediction model as input (prediction based on Netflix video multimethod assessment fusion (VMAF) [5], NTIA VQM [6], SSIM, PSNR, or PSNR-human visual system (HVS)).

These studies attempted to evaluate acceptability and determine its relationship with QoE. However, one important aspect that has not been addressed so far is the effect of stimulus duration on acceptance ratings. Indeed, previous works addressing acceptability are based on stimuli having a constant duration (mostly of 10 seconds). However, as stimulus duration increases, participants' patience towards low quality decreases and so does the respective acceptance rating. This relates to the work of annoyance [9], [10], and how annoyance leads to unacceptability after a long period of time. Therefore, acceptance with regards to the temporal aspect needs to be addressed, which is a key contribution of this paper.

It should also be mentioned that although the temporal evolution of acceptability has not been studied before, it can be related to the viewing time studies for which the percentage of users who watch a video fully (the completion ratio) is studied. The completion ratio gives information on whether users watch a video fully or abandon it midway through. The completion ratio was shown to depend on content duration [16]–[20], but the effect of content duration with respect to QoE was not characterized. This is because most of these studies are based on real-life service data, which involve many factors other than quality. Therefore, in such scenarios, important factors include interest in the content, context of the viewing session (at home, on the train, etc.), and time available for the user to watch videos. This leads to video coding quality-related parameters such as average bitrate or even subjective quality ratings weakly correlating with viewing time [21], [22]. Bitrate variations were identified to be relevant, but their effect was not precisely quantified [23]. Therefore, although results on the completion ratio and its relationship with content duration can be found, further work is still needed to integrate QoE in a time-dependent acceptance model. This will be the main contribution of this work.

The contributions of this work are as follows. First, compared with past methodologies based on the evaluation of acceptability by asking users about acceptability in a post-viewing question, this paper proposes a non-intrusive evaluation of acceptability by measuring quitting behavior and monitoring the completion ratio. Then, a second contribution is to quantify the relationship between QoE, time, and completion ratio. Finally, it is shown how a previous model designed for QoE can be applied to predict the completion ratio and how such predictions can be time-dependent.

III. Subjective experiment

To investigate the completion ratio of videos while taking into account both QoE and stimulus duration, multiple subjective tests were conducted in a laboratory environment. In these experiments, participants watched videos of different coding quality and were told that they could quit watching whenever they desired. Quitting actions had to be motivated by coding quality and not because of a lack of interest in the content. By using this approach, the completion ratio with respect to QoE and time can be studied.

A. Test conditions

To investigate the effect of coding quality on abandonment, videos were encoded at different coding quality involving various combinations of resolution, bitrate, and frame rate. Two codecs (H.265/HEVC (High Efficiency Video Coding) and H.264/AVC (Advanced Video Coding)) were considered as H.265 and H.264 are frequently used for 4K and HD videos, respectively. 4K and HD videos were encoded using FFmpeg (version 4.0.2-4.2.0) with x265 and x264 codecs, respectively. A two-pass encoding with the preset "slower" was used. As for audio, the codec was always AAC-LC (low complexity advanced audio coding), and the encoding was performed using libfdk_aac from FFmpeg.

All test conditions were evaluated by the means of six subjective experiments conducted between 2017 and 2019. Experiments included constant quality conditions, conditions with quality adaptations, stalling events, and initial buffering enabling an adaptive bitrate video streaming service to be simulated. Quality adaptations included small to large quality changes, and stalling ranged from 2 to 36 sec. These were included in various positions to produce a large variety of conditions. Experiments had between 21 and 36 processed video sequences (PVS), and each PVS lasted between 3 and 5 min. This resulted in a total of 155 PVSs leading to 609 min of videos. PVSs were evaluated in experiments of approximately 1 hour 30 min with multiple breaks (after every 2 PVSs of 3 min, or after each video longer than 3 min). Each PVS was seen by 32-40 participants depending on the experiment. Special care was taken that in a single experiment, participants did not see the same content multiple times (source reference circuit (SRC)), and different groups of participants were hired to distribute PVSs across participants. In total, these experiments involved 304 participants across all tests (participants were aged from 18 to 30 years old, with a median age of 21. Exact gender balance was ensured in each experiment).

Table I lists the test conditions addressed in this paper. A parameter-based audiovisual quality model [3], [24] was used to give information on audio, video, and audiovisual quality

TABLE I

LIST OF CODING CONDITIONS. VBR AND ABR ARE RESPECTIVELY VIDEO AND AUDIO BITRATE IN KBPS. RES IS THE RESOLUTION OF THE VIDEO GIVEN BY ITS HEIGHT (ASPECT RATIO OF 16:9). FR IS THE FRAME RATE. A, V, AND AV ARE RESPECTIVELY AUDIO, VIDEO, AND AUDIOVISUAL QUALITY ESTIMATES USING AN AUDIOVISUAL QUALITY MODEL [3], [24]

Codec	VBR	ABR	Res.	FR.	A	V	AV	(PVS, Dur., Exp)
HEVC	15000	128	2160	60	4.91	4.70	5.00	(P38,3,D), (P48,2,E), (P49,2,E), (P50,2,E), (P51,3,E)
HEVC	8000	196	2160	60	4.94	4.52	5.00	(P09,3,B), (P10,1,B), (P29,5,C), (P30,3,C)
HEVC	4000	128	2160	60	4.91	4.18	4.66	(P11,3,B)
HEVC	800	32	2160	30	4.17	2.85	3.08	(P39,1,D)
HEVC	200	32	2160	30	4.17	1.73	2.06	(P52,3,E)
HEVC	7000	192	1080	60	4.94	4.49	4.98	(P40,3,D), (P53,2,E), (P54,2,E)
HEVC	4000	128	1080	60	4.91	4.30	4.77	(P31,3,C)
HEVC	800	48	1080	30	4.57	3.27	3.63	(P55,3,E)
HEVC	800	384	1080	30	4.95	3.27	3.79	(P81,3,F)
HEVC	600	32	1080	30	4.17	3.01	3.22	(P41,1,D)
HEVC	300	32	1080	30	4.17	2.39	2.66	(P56,3,E)
HEVC	150	32	1080	30	4.17	1.86	2.17	(P57,3,E)
HEVC	10000	128	720	60	4.91	4.37	4.85	(P42,3,D), (P58,3,E)
HEVC	4000	48	720	30	4.57	4.12	4.44	(P43,1,D), (P44,1,D), (P59,3,E)
HEVC	4000	384	720	30	4.95	4.12	4.63	(P82,3,F)
HEVC	1000	128	720	60	4.91	3.35	3.85	(P12,1,B), (P13,2,B), (P14,3,B), (P15,1,B), (P16,1,B), (P32,5,C), (P61,1,E)
HEVC	1000	48	720	30	4.57	3.41	3.76	(P60,1,E), (P62,1,E)
HEVC	1000	384	720	30	4.95	3.41	3.93	(P83,1,F), (P84,1,F), (P85,1,F)
HEVC	400	32	720	30	4.17	2.66	2.91	(P45,2,D)
HEVC	250	32	720	30	4.17	2.27	2.55	(P63,3,E)
HEVC	250	384	720	30	4.95	2.27	2.81	(P86,3,F)
HEVC	3000	32	480	30	4.17	3.68	3.83	(P46,1,D)
HEVC	900	32	480	30	4.17	3.09	3.30	(P64,3,E)
HEVC	900	384	480	30	4.95	3.09	3.62	(P87,3,F)
HEVC	640	96	480	30	4.86	2.86	3.35	(P17,1,B), (P18,1,B), (P19,3,B), (P20,1,B), (P21,1,B), (P22,1,B), (P23,2,B), (P33,3,C), (P34,5,C), (P35,1,C), (P36,1,C), (P37,1,C)
HEVC	450	96	360	30	4.86	2.38	2.89	(P24,1,B), (P25,2,B), (P26,3,B), (P27,1,B)
HEVC	350	32	360	30	4.17	2.22	2.50	(P47,1,D), (P65,1,E), (P66,1,E), (P67,3,E)
HEVC	350	384	360	30	4.95	2.22	2.76	(P88,3,F)
HEVC	200	384	360	30	4.95	1.88	2.42	(P89,3,F)
HEVC	800	32	240	15	4.17	2.14	2.43	(P68,3,E)
HEVC	800	384	240	15	4.95	2.14	2.68	(P90,3,F)
HEVC	100	64	144	30	4.74	1.19	1.70	(P28,1,B), (P70,3,E)
HEVC	100	32	144	30	4.17	1.19	1.57	(P69,1,E)
AVC	10000	256	1080	30	4.95	4.53	5.00	(P01,3,A), (P02,1,A), (P03,1,A)
AVC	8000	384	1080	30	4.95	4.48	4.97	(P71,3,F)
AVC	7500	384	720	60	4.95	4.32	4.82	(P72,1,F), (P73,1,F)
AVC	5000	384	720	30	4.95	4.19	4.69	(P74,3,F)
AVC	2500	384	480	30	4.95	3.61	4.13	(P75,3,F)
AVC	1000	256	480	30	4.95	3.15	3.67	(P04,2,A), (P05,1,A)
AVC	500	384	480	30	4.95	2.68	3.21	(P76,3,F)
AVC	300	384	480	30	4.95	2.30	2.83	(P77,3,F)
AVC	1000	384	360	30	4.95	2.86	3.38	(P78,3,F)
AVC	400	384	360	30	4.95	2.31	2.84	(P79,3,F)
AVC	300	384	240	30	4.95	1.80	2.34	(P80,3,F)
AVC	200	256	240	30	4.95	1.33	1.88	(P06,3,A), (P07,1,A), (P08,1,A)

(listed respectively as A, V, and AV in Table I). This model has a closed form and takes as input audio and video bitrate values as well as the frame rate and resolution. Then, it outputs audio, video, and audiovisual quality estimates. The model was trained to predict quality of videos seen on mobile devices and can then predict the audiovisual quality of videos for various resolutions when videos are watched on such devices.

As this paper focuses on the effect of quality on the completion ratio with the ultimate goal of defining coding conditions on the server side, PVSs that contained stalling events or quality adaptations are not considered. Although not addressed here, the motivation for including quality adaptations and stalling conditions in the subjective experiments was twofold. First, it enables future work to address the completion ratio with broader conditions. Second, it is beneficial for this study focusing on defining how to encode the videos, as the range of quality conditions in subjective tests affects the user behavior. Therefore, although stalling and quality adaptation are not directly relevant to the definition of how to encode videos on the server, these allow users to put into perspective low coding quality with other common impairments such as stalling and quality adaptations. Test conditions were selected to cover a large span of video quality (V from 1.2 to 4.7). Audio quality also varied, but very low quality was avoided (A from 4.17 to 4.95). Conditions were designed to disambiguate audio and video quality enabling testing individual components.

Table I list tuples (PVS, Duration, Experiment) to report on the conditions that were used. The first tuple is the PVS number, the second tuple is the duration of the constant coding condition, and the last tuple is the experiment in which the PVS was evaluated. Durations of 1 and 2 min are listed in this table, as these PVSs contained quality adaptations and only the initial 1- and 2-min segments of constant quality segment before the quality changed could be used in this work.

Regarding SRCs, thirteen 3840x2160 4K-UHD videos with a frame rate of 60 frames per second (FPS) were used. The videos showed a large variety of content corresponding to common TV shows in Japan. Scenes included natural scenery, traditional festivals, documentaries, sports, interviews, etc. These videos were available in a raw format and were recorded by video professionals using professional grade cameras. Finally, audio had two channels and a sampling rate of 48 kHz and was available in an uncompressed format.

B. Experimental setup and 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 3840x2160 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 (gray room, controlled ambient light, acoustic treatments, etc.). The illumination was set to 20 lux, which corresponds to a dark room.

Regarding their task, participants were instructed to watch the videos and were 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, participants were asked to base their decision to quit only on quality-related reasons and not because of the content, giving this work a focus on acceptability of coding quality.

To take the test, participants first needed to pass vision tests: visual acuity (with correction glasses if needed) and color vision. The experiment included a training phase with six 3-min videos over three sessions (SRCs were different than the main experiment).

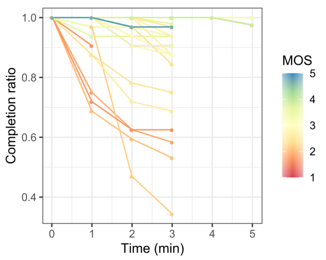

Fig. 1. Effect of quality on completion ratio with respect to time.

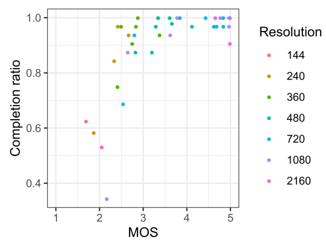

Fig. 2. Completion ratio at 3 min as a function of MOS.

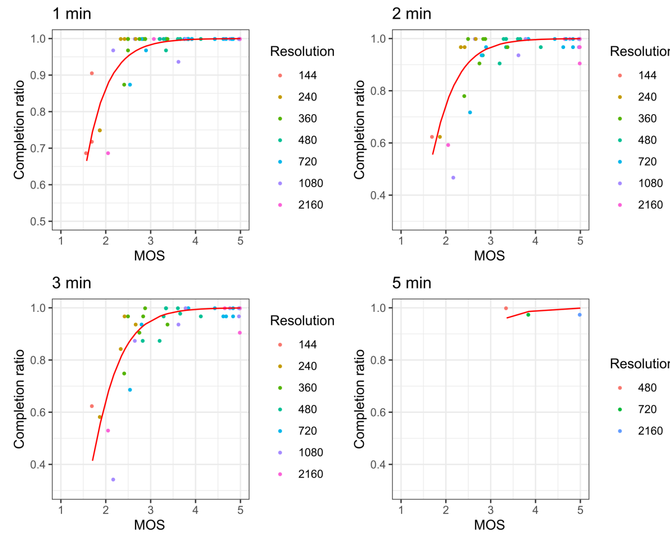

Fig. 3. Prediction of completion ratio as a function of quality and time.

IV. Overview on Experimental Results

In the experiments, the times of user quitting actions were recorded. This enabled the completion ratio of each PVS to be estimated on a per-minute basis on the basis of the percentage of users who were still watching after every minute. Thus, the relationship between quality, time, and completion ratio could be revealed. Figure 1 shows the temporal evolution of the completion ratio on a per-minute basis for all PVSs. Audiovisual quality (hereafter, mean opinion score (MOS)) is indicated using a prediction model [3], [24]. This figure shows that there is a temporal aspect in the decision of users to quit because of bad quality and that a logarithmic decay of the completion ratio as a function of time can be found. Then, consistent with previous work on acceptability, as quality decreases, the completion ratio also decreases logarithmically. To further show the effect of quality on user quitting actions, Figure 2 depicts the completion ratio at 3 min as a function of MOS and shows a clear relationship between the quality and completion ratio.

V. Prediction of Completion Ratio

In this section, the completion ratio is predicted as a function of quality and time. Considering that the subjective experiments did not include a quality evaluation task, an audiovisual quality prediction model is used to obtain quality estimates [3]. To predict the completion ratio as a function of time, the following approach is proposed. First, the relationship between quality and completion ratio after 1 min is estimated using Equation 1. In this equation, $C(Q)$ is the completion ratio at 1 min as a function of the video quality (Q), and c_{1-2} are model parameters.

$$C(Q) = 1 - e^{-c_1 \times Q + c_2} \tag{1}$$

By using this equation, the completion ratio at 1 min can be predicted. Then, the completion ratio after multiple minutes, t, of videos with a quality, Q, noted $C(Q,t)$, is estimated by the means of a power function as described in Equation 2.

$$C(Q,t) = [C(Q)]^t \tag{2}$$

To show the performance of the proposed approach, the quality of the video (Q) is estimated using [3], [24], and predictions are compared with ground truth data. Figure 3 shows the prediction accuracy of the model at 1, 2, 3, and 5 min (ground truth values are obtained from the proportion of users still watching the videos after t-min). The four plots show the relationship between MOS and ground truth completion ratio values. In each scatter plot, the red line corresponds to the relationship between MOS and the completion ratio estimated by the model defined in Equations 1 and 2. In this figure, the model was trained on data corresponding to the completion ratio at 1 min, and predictions at 2, 3, and 5 min are obtained using Equation 2 without involving further retraining (coefficients are $c_1 = 2.1010$ and $c_2 = 2.2134$). Table II provides a quantitative evaluation of the model at different points in time. The table also compares a full-retraining of the model at each point in time based on Equation 1 and the proposed approach based only on a training using completion ratio at 1 min along with the use of a power function as described in Equation 2. This table shows that with a longer time frame, the root mean square error (RMSE) increased significantly. This is due to differences in data distributions, as after 1 min, most users were still watching videos, so the completion ratio is mostly equal to 1. However, at 2 and 3 min, most PVSs had a non-null completion ratio, making the prediction more difficult and leading to increased RMSE. Interestingly, it can be seen that for 2 and 3 min, having a direct training of Equation 1 on 2- and 3-min completion ratio data provides comparable performance results with the training of Equation 1 on 1-min completion ratio data put respectively to the square or the cube. This shows that the power-based approach can be used to predict the completion ratio at different times. In terms of the Pearson correlation coefficient (PCC), the proposed approach shows a similar correlation to a full retraining while only providing a loss of ≈ 0.02 in terms of RMSE. Finally, note that the table provides data only up to 3 min as not enough data points were available to perform analysis at 4 and 5 min (see Figure 3 bottom-right).

TABLE II
QUANTITATIVE EVALUATION OF PREDICTION ACCURACY. PCC IS THE PEARSON CORRELATION COEFFICIENT, AND RMSE IS THE ROOT MEAN SQUARE ERROR. TIME IS GIVEN IN MINUTES.

Time	Metric	Model based on Eq. 1 - 2	Full retraining of Eq. 1	4-fold cross val. Eq. 1 - 2)
1	PCC	0.8567	0.8567	0.9385
	RMSE	0.03690	0.03690	0.02243
2	PCC	0.8002	0.7989	0.8013
	RMSE	0.07268	0.07080	0.07244
3	PCC	0.8016	0.8069	0.8042
	RMSE	0.08765	0.08542	0.08689

TABLE III
COMPARISON BETWEEN DIFFERENT MAPPING FUNCTIONS.

Time	Metric	Logistic	Polynomial
1	PCC	0.86389	0.82349
	RMSE	0.03596	0.08198
2	PCC	0.80710	0.80032
	RMSE	0.07155	0.13429
3	PCC	0.80678	0.8217
	RMSE	0.08727	0.1832

To analyze the performance, a four-fold cross-validation was performed. Considering that the distribution of ground truth completion ratio data at 1 min is unbalanced (most PVSs have a completion ratio of 1), special care was taken to create four random but balanced sets. Therefore, each set contained an equal number of non-equal-to-one completion ratio data points. Then, training is performed only based on the 1-min data (using 3/4 of the data), and completion ratios at 2 and 3 min are predicted using Equation 2. Table II provides average validation results and shows consistent performance with previous training.

Then, as described in Section II, several previous works have attempted to predict acceptability of videos based on QoE ratings. These models were either based on a logistic-based function [9], [13] or third order polynomial function [11]. It is then proposed to train these types of model to predict the completion ratio at 1 min and show performance results. As these previous works did not address temporal aspects, Equation 2 is used to extend predictions to 2 and 3 min. Table III compares these mapping functions (using quality values based on [3], [24]). These results are comparable with those in Table II and show that the logistic function performs equivalently to Equation 1. As for the third order polynomial mapping, performance was found to be significantly worse. These performances indicate that previous work on acceptability for a fixed duration can be extended to a longer time span using the proposed approach from Equation 2.

VI. QUALITY REQUIREMENTS WITH RESPECT TO TIME

The proposed model relates quality, time, and completion ratio. An interesting result that can be derived from it is the identification of quality requirements to preserve a low abandonment rate as a function of time. This is given in Equation 3. In this equation, c_1 and c_2 are the coefficients previously obtained when training Equation 1. From Figure 2, there are different completion ratio values for a given MOS

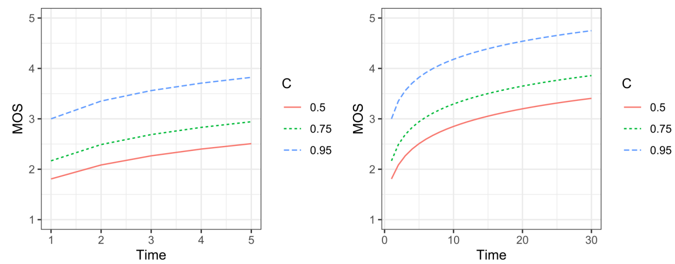

Fig. 4. Quality requirements for reaching different completion ratio (C) after different durations. Left plot depicts short-term predictions (less than 5 min), while the right plot shows requirements for videos up to 30 min.

value. Therefore, a quality-preservation model was trained by taking into account only the PVSs with lower completion ratio. To this end, intervals of 1 unit are defined: $[1, 2]$, $[2, 3]$, $[3, 4]$, $[4, 5]$, and in each interval PVSs with a completion ratio within the 10% lowest values are identified. These PVSs are then used to train the coefficients c_1 and c_2 of Equation 1 enabling the model to be trained for most critical values. By using this approach on 1-min completion ratio data, $c_1 = 1.9307$ and $c_2 = 2.7973$. These are used in Equation 3, and Figure 4 shows the resulting quality $Q(C, t)$ that needs to be reached to have a completion ratio C as a function of content duration t. To ensure that less than 5% of users quit the video after 1 min, a MOS larger than 3 is needed. Thus, 50% of the users will quit after 1 min if MOS is below 2. To ensure that less than 5% of users quit after 5 min of video, MOS needs to be 4. Whereas 25% of users quit after 5 min even if the MOS is 3, 50% of users will quit if MOS is 2.5. Figure 4 right) shows predictions extrapolated up to 30 min and shows that retaining users for long duration requires high video quality as maintaining 75% of users for 30 min would require a MOS of 4. Such extrapolation requires further testing but shows the need for high quality when dealing with long videos. Finally, note that identified quality requirements may be high, but the key result of this work is not only the absolute quality requirements themselves but also the relative differences between short and long videos and that long video content requires higher quality.

$$Q(C, t) = \frac{c_2 - ln(1 - \sqrt[t]{C})}{c_1} \tag{3}$$

VII. DISCUSSION

First, one limitation of this work is the temporal resolution at which completion ratio estimates are made. Estimations are obtained at multiples of the base duration used for training (1 min in this work). Although working with smaller base duration is possible, having too short a duration such as 10 sec may not leave enough time for users to leave and would prevent the model to be trained. Then, a second point that should be mentioned is that this work does not handle quality adaptation and the model is only defined for constant quality conditions. Quality adaptations raise new challenges as users perceive changed quality differently from constant quality [25]. Such a restriction is reasonable for this study as the goal here is to find acceptable coding conditions with respect to content duration when encoding videos. Therefore, quality adaptation does not apply. A third point to address is the impact of context and content. This work is based on laboratory experiments, but previous work has shown that acceptability depends on context. Participants are more sensitive to low quality in laboratory tests than in real-life settings [13], [26], [27]. However, this may not be too critical as this only results in more conservative results. As for content, although users were told to only consider coding quality, their interest in the content may unconsciously affect their quitting behavior. This problem may be mitigated by using a repeated design with multiple participants as done in this work. Finally, it should be mentioned that the analyses of this paper were performed on the basis of audiovisual quality scores, and individual audio and video quality components are expected to affect the quitting behavior differently. However, it is difficult to test this in this work as the importance of audio and video quality depends on the content. Moreover, the quality ranges of audio and video conditions were not equal. Therefore, results are limited to impacts of audiovisual quality on the completion ratio.

VIII. Conclusion

This paper addresses how to encode videos of a video streaming platform. To this end, the relationship between quality, time, and completion ratio of videos as a proxy for acceptability was studied. Although not addressed in previous studies, stimulus duration was shown to be important for acceptability. A model of the completion ratio for video of constant quality was developed and showed good accuracy. By using this model, quality requirements can be defined to reach a minimum completion ratio considering the content duration. This result can be used by video service providers while encoding videos and enables coding parameters to be selected while focusing on acceptability with respect to content duration. Future work will extend this study by focusing on estimating the optimal bitrate ladder for adaptive bitrate video streaming to maintain user engagement.

References

[1] K. Brunnström, S. A. Beker, K. D. Moor, A. Dooms, S. Egger, M.-N. Garcia, T. Hoßfeld, S. Jumisko-Pyykkö, C. Keimel, and M.-C. L. et al., "Qualinet white paper on definitions of quality of experience," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, 2012.

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

[3] K. Yamagishi and T. Hayashi, "Parametric quality-estimation model for adaptive-bitrate-streaming services," *IEEE Trans. on Multimedia*, vol. 19, no. 7, pp. 1545–1557, 2017.

[4] ITU-T Recommandation P.1203, "Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport," *ITU-T*, 2017.

[5] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric," in *tech blog with VMAF's open sourcing on Github*, 2016.

[6] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," in *ITU-T*, 2004.

[7] S. Jumisko-Pyykkö, V. K. M. Vadakital, and M. M. Hannuksela, "Acceptance threshold: A bidimensional research method for user-oriented quality evaluation studies," in *International Journal of Digital Multimedia Broadcasting*, 2008.

[8] I.-T. R. P.10/G.100, "Vocabulary for performance, quality of service and quality of experience," in *ITU-T*, 2017.

[9] W. Song and D. W. Tjondronegoro, "Acceptability-based QoE models for mobile video," *IEEE Trans. on Multimedia*, vol. 16, no. 3, 2014.

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

[11] T. C. M. de Koning, P. Veldhoven, H. Knoche, and R. E. Kooij, "Of MOS and men: Bridging the gap between objective and subjective quality measurements in mobile TV," in *Multimedia on Mobile Devices 2007 IS&T/SPIE Symposium on Electronic Imaging*, 2007.

[12] T. D. Pessemier, K. D. Moor, A. J. Verdejo, D. V. Deursen, W. Joseph, L. D. Marez, L. Martens, and R. V. de Walle, "Exploring the acceptability of audiovisual quality for mobile video session based on objectively measured parameters," in *Quality of Multimedia Experience*, 2011.

[13] M. Sasse and H. Knoche, "Quality in context - an ecological approach to assessing QoS for mobile TV," in *Proc. 2nd ISCA/DEGA Workshop Perceptual Quality System*, 2006.

[14] P. Spachos, W. Li, M. Chignell, L. Zucherman, and J. Jiang, "Acceptability and quality of experience in over the top video," in *IEEE ICC 2015 - Workshop on Quality of Experience-based Management for Future Internet Applications and Services (QoE-FI)*, 2015.

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

[16] H. Nam and H. Schulzrinne, "Youslow: What influences user abandonment behavior for internet video?" in *Columbia University, Tech. Rcp*, vol. 44, no. 4, 2014.

[17] S. Wu, M. Rizoiu, and L. Xie, "Beyond views: Measuring and predicting engagement in online videos," in *ICWSM*, 2018.

[18] 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. on Multimedia*, vol. 15, no. 8, p. 2087, 2013.

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

[20] P. Lebreton and K. Yamagishi, "Study on user quitting rate for adaptive bitrate video streaming," in *IEEE 21th International Workshop on Multimedia Signal Processing*, 2019.

[21] F. Dobrian, A. Awan, D. Joseph, A. Ganjam, J. Conviva, V. Seka, I. Stoica, and H. Zhang, "Understanding the impact of video quality on user engagement," *Communications of the ACM*, vol. 56, no. 3, pp. 91–99, 2013.

[22] M. Seufert, N. Wehner, F. Wamser, P. Casas, A. D'Alconzo, and P. Tran-Gia, "Unsupervised QoE field study for mobile YouTube video streaming with YoMoApp," in *Quality of Multimedia Experience*, 2017.

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

[24] P. Lebreton and K. Yamagishi, "Transferring adaptive bit rate streaming quality models from H.264/HD to H.265/4K UHD," *IEICE Transactions on Communications*, vol. E102-B, no. 12, pp. 2226–2242, 2019.

[25] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Communication Surveys & Tutorials*, vol. 17, no. 1, pp. 469–492, 2015.

[26] R. Schatz, S. Egger, and A. Platzer, "Poor, good enough or even better? bridging the gap between acceptability and qoe of mobile broadband data services," in *IEEE ICC*, 2011.

[27] S. Jumisko-Pyykkö and M. Hannuksela, "Does context matter in quality evaluation of mobile television?" in *Proceedings of the Conference on Human-Computer Interaction with Mobile Devices and Services*, 2008.