Study on user quitting rate for adaptive bitrate video streaming

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

***Abstract*—In this study, the effect of coding degradation and stalling events on the percentage of users who quit watching videos midway is studied (referred to as the *user quitting rate*). The results are based on three laboratory-based subjective experiments involving up to 104 participants. The results show that for the coding condition only, a Mean Opinion Score (MOS) lower than 3 on a 5-point ACR quality scale will result in users quitting the video. The *quitting rate* was found to increase further when the MOS falls below 2.5. When a stalling event occurs, the *quitting rate* was found to be dependent on the stalling position, the stalling duration, the percentage of user who already quit, and the MOS. Statistical analysis allowed the identification of interaction terms between the stalling position and percentage of users who had already quit. Finally, the results were used to establish a model of the user *quitting rate* due to stalling.**

***Index Terms*—Viewing time, quitting rate, adaptive bitrate streaming, user engagement, quality, stalling**

I. Introduction

Video streaming is an important application for users. It is the most dominant application on the Internet and represented 73% of Internet traffic in 2016, which is expected to increase to 82% by 2021 [1]. To encourage users using their services, video streaming platforms need to ensure user satisfaction. Consequently, much research has then been conducted to evaluate and predict the Quality of Experience (QoE) using computational approaches within academia [2], [3] and standardization activities [4] allowing operators to monitor their services. However, a recent trend in research is to move from QoE to User Engagement (UE). UE is important as it provides a close view on whether users will keep using the services. Therefore, this work will focus on the investigation of the user's engagement in adaptive bitrate video streaming and how it relates to the perceived quality. Various methods are used to perform the quantification of UE. The most frequent ones are based on viewing time (the time a user watches a video including the loading delay for stalling events), or by studying action of users when the service is disrupted [5], [6]. In this work, it is the users quit events which will be studied. Having a good understanding of when users quit and its relation with technical parameters is of importance for operators as it allows them to understand when users may quit a video (or the whole service), enable them to react upfront to it, and ultimately allow them to build a healthy ecosystem with paying customers or getting revenue from advertisements. To achieve this, it is necessary to establish the link between users quitting and service quality.

978-1-7281-1817-8/19/$31.00 ©2019 IEEE

A. Viewing time, type of service and quality

Modeling the relationship between viewing time and quality is challenging. First, there is a complex nonlinear relationship between quality and viewing time [7]. Second, "confounding factors" [8] such as the type of content, the temporal effect (prime-time video vs. off-peak), the type of device used, the content, the user's interest, and the connectivity affect the results.

Nevertheless, a large amount of work has been performed to study the relationship between content properties, quality-related features, and viewing time. Results show that the type of video is of importance and that VOD and live content will have different quality requirements for users [7]. In the case of VOD, the *buffering ratio*, defined as the ratio between the duration of stalling events and the duration of uninterrupted playback, is a key feature for predicting viewing time [8]–[10]. However, in the case of live video, the average bit rate is the key feature [8], [10]. The type of network connectivity is also important as mobile users over cellular networks may prefer lower video quality to reduce data consumption [11].

Content duration is of importance, and short videos are more easily quit than long ones [10]. The percentage of users who watch video fully was also found to linearly decrease with content length [12]. In terms of quality, it was found that a video with a coding quality lower than 2.5 on the MOS scale is not acceptable [13]. Users also prefer constant quality over video with large quality variations [14]. Regarding stalling, it was found that an increase of 1% in the buffering ratio leads to a 3-min decrease in viewing time in the case of VOD [10]. Startup delay is also important, and a delay larger than 2 sec results in users dropping videos [15]. This phenomenon also depends on overall content duration and the type of Internet access (mobile or TV) [15]. Finally, recent studies in a laboratory setting have found that the number of users quitting videos because of a stalling event is a linear combination between the stalling position and the stalling duration [16], and going from a 12 sec to 24 sec stalling events adds a constant offset of 30% to the *user quitting rate*.

These previous studies are of high interest as they allow the understanding of which factors are relevant for predicting viewing time. However, the quantification of the contribution of quality impairment on the viewing time is still unknown.

This is due to the use of real-life datasets which contains various combinations of coding and stalling condition across multiple context (mobile, TV, live, VOD) in a single session making difficult to quantify the impact of each impairment. Therefore, this work differs from previous studies as it addresses the effect of video coding quality and stalling (position and duration) in three systematic laboratory experiments allowing quantify the relative importance of each artifacts on the viewing time.

B. Viewing time and user's interest

Beyond quality issues, content dependency should also be addressed [17]. Only a few studies have considered the user's interest in the content in viewing time due to its complexity. It was found that the contents chosen by users mostly belong to top ranked channels [18], and the relationship between viewing time and channel popularity is logarithmic [19]. Moreover, the distribution of viewing time can be modeled using two distributions: an exponential distribution to address an early departure, and a truncated power law for cases where users continue watching after an initial screening phase [5], [10]. Finally, Tan et al. [20] developed an Extraction-Inference algorithm based on matrix factorization to address user interest towards content. Using this approach, the authors were able to find that there exists a multiplicative effect between user interest and QoS.

These previous studies are of high interest as they allow to tackle the difficult question of user interest and content dependency. However, with the exception of the work of Tan et al. [20], they do not quantify the effect of quality on the identified distribution, and in the case of Tan et al. [20], QoS is used although it does not reflect directly user experience as shown in the extensive past QoE modeling work [2]–[4]. Therefore, the relation between QoE, user interest and viewing time still needs to be further investigated. Although of interest, content dependency will not be addressed in this work and will be left to future research. This is due to the fact that interest towards content is difficult to address for operators, and operators can expect to address it by offering a large diversity of content allowing users to find the content they like. The quality of the service is on the other hand under their control.

C. Contributions

This study is motivated by the fact that previous studies still lack a comprehensive model summarizing the effect of coding and stalling on user quitting video services. This is because of the use of large-scale data which involves large variation of conditions across participants and many hidden factors are involved making it difficult to develop non-black box models. Laboratory experiments are proposed to tackle this challenge as they allow crafting test conditions and repeat the exact same stimuli across participants. By doing so, it is ensured that in the tests no factors others than the tested ones changes from one stimulus to another. Such precise design allows distribution analysis across participants, quantitative analysis of the effect of artifacts on user quit events. The second main difference from previous work is that this work addresses the percentage of users who quit at any instant in time (the *user quitting rate*). Indeed, while previous works have either predicted a single numeric value for viewing time of a given session, or performed distribution analysis mixing various viewing sessions with different quality conditions. The proposed approach performs distribution analysis per-quality condition (with a fixed quality adaptation and stalling pattern). This is possible thanks to the laboratory experiment and the repeated conditions across participants. Therefore, instead of providing a single numeric value for viewing time as in [8]–[10], this work addresses the probability that a user watches until a given point. Finally, it should be noted that working with the *user quitting rate* eases the problem of content dependency and user interest, as analysis across several participants are performed masking individual ratings variations.

This paper is organized as follows: Section II describes the three subjective experiments conducted, Section III provides details on the link between quality factors and user quitting rate, Section IV discusses the results and Section V concludes this paper.

II. Subjective experiments

A. Test conditions

This work aims at predicting the temporal evolution of the number of users watching videos with different coding and stalling conditions. Figure 1 provides an overview of the test conditions for each of the three tests conducted. For each processed video signal (PVS), the temporal evolution of video quality between different quality levels (QL) and their respective durations (in brackets) are provided. Each quality level is labeled by $Q[0-5]$ or $Q[11-15]$ and corresponds to a coding and resolution condition listed in Table I. Coding conditions were selected to balance quality from low to high quality using estimates from [2]. The audio codec was always AAC-LC. Stalling events were also inserted and marked in Figure 1 by white boxes along with their respective durations. The motivation behind the selection of these conditions was to simulate an adaptive bit rate streaming session which contains quality variations and stalling events. The specific design of test conditions allows testing the effect of low to high-coding quality and stalling events on viewing time. A wide range of stalling event is considered: from 6 to 36 seconds. Real-life stallings are mostly of short duration, and falls within this range [14]. Considering long stalling duration is motivated by the fact that it enables identifying non-linear relationship between viewing time and stalling which would not be possible otherwise. Similarly, coding conditions and quality adaptations offer a wide span of conditions, with real-life conditions falling into the range of considered quality levels [14]. Doing so, allows the relation between coding quality and viewing time to be investigated and allows model development which can support the various conditions which occur in real-life services. The tests involved videos with different durations:

two tests used 3-min long video sequences and one used 5-min long sequences. The source contents were either native 3840x2160 4K videos with a frame rate of 60 frames per second (FPS) or 1920x1080 HD content with a frame rate of 30 FPS. The content showed a large variety of scenes shot in Japan. Contents were similar to common TV shows in Japan and depicted scenes such as scenery, festivals, different types of sports events, documentaries, interviews, etc. The first test, “3 min-A”, only used the HD source content (as the 4K ones were not available at that time), while the two other tests, “3 min-B” and “5 min-A”, used the 4K videos. Information about the combinations between sources and processing is provided in Figure 1. While the test “3 min-A” contains multiple repetitions of the same source content, special care was taken in the tests “3 min-B” and “5 min-A” to avoid repetition. Therefore, test “3 min-B” had only one repetition and test “5 min-A” had no repetition.

TABLE I
LIST OF QUALITY LEVELS. RES. IS THE VIDEO RESOLUTION, B.R. ARE THE BIT RATE VALUES IN KBPS, AND FR. IS THE FRAME RATE.

QL	Codec	Res.	Video B.R.	FR.	Audio B.R.
Q0	HEVC	144	100	30	64
Q1	HEVC	360	450	30	96
Q2	HEVC	480	640	30	96
Q3	HEVC	720	1000	60	128
Q4	HEVC	2160	4000	60	128
Q5	HEVC	1080	4000	60	128
Q11	AVC	1080	10000	30	128
Q12	AVC	240	200	30	128
Q13	AVC	480	1000	30	128
Q14	AVC	720	1600	30	128
Q15	AVC	360	500	30	128

B. Subjective evaluation

Subjective tests were conducted on smartphones with a video player designed to track user behavior and viewing time. In the case of the test “3 min-A”, the playback was performed on 5.2-inch smartphones with full HD resolution, while the tests “3 min-B” and “5 min-A”, used 5.5-inch 4K smartphones. In every case, participants used headphones to listen to the audio. The viewing distance was set to 5-7 H (H being the picture height) and listened to each sequence PVS at -21 dB. The room was a standardized laboratory environment designed for video tests: gray room, controlled lighting, etc. The illumination was set at 20 lux, which corresponds to a dark room. After passing visual acuity tests, participants were provided with written instructions and went through a training phase that consisted of watching six 3 min long videos (in 3 sessions). After this, the main task of the experiment started. The main task was divided into several sessions (11 in the 3 min tests, 13 in the 5 min test). In each session, the user tap on a start button which starts the playback of a video. During playback, the participant can make playback controls appear by touching the screen. When he desires, he can hit the stop button. Once playback is stopped, a button to switch to the next video is enabled allowing moving to a second video (in the 3 min test), or end the session (in the 5 min test, or after the second video in the 3 min test). Between sessions, participants were asked to rest. To avoid participants

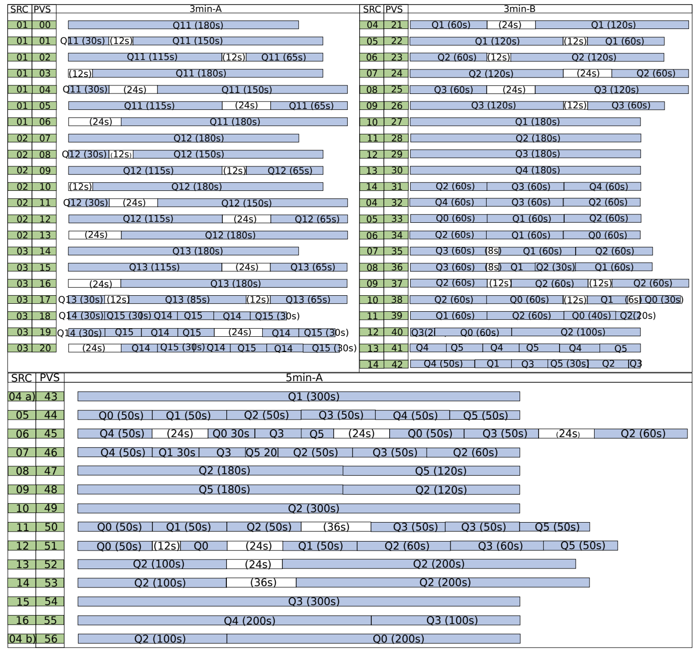

Fig. 1. Test conditions. PVS stands for process video signal, and SRC for source reference circuit.

rushing into the test, the experimenter told participants when to start a new video. The total experiment duration including breaks took about 2 hours 20 minutes. Presentation order of PVS was randomized. Participants were not given any task and were told that they could stop watching whenever they wanted to, but they should not quit because of lack of interest into the contents. Forty participants participated in the “3 min-A” experiment, and 32 participants participated in each of the “3 min-B” and “5 min-A” experiments. Tests were conducted on different days and involved different participants. Special care was taken to ensure gender balance in each test. Participants’ age was in 18-29 years old (avg. 21).

III. ANALYSIS OF USER QUITTING RATE

A. Results overview

Figure 2 depicts examples of the data obtained from the experiments. It shows the “*user quitting rate*, ” which correspond to the percentage of users who quit the video as a function of time since the playback started. The scatter plot illustrates cases where the users experienced stalling events and quit the video because of them. The stalling events begin and end in the figure with a red continuous line and a blue dashed line. The scatter plot includes information about video quality using a green line. Considering that the subjective evaluation did not include a continuous quality evaluation task, it was chosen to estimate video quality using a computational algorithm [2], [21]. Using this model, a per-second video quality estimation could be obtained, and the results show that users could also quit due to low video quality. In the following text, the paper will refer to “*user drop event* amplitude” to address the effect of a single degradation (stalling or low coding quality) on the "*user quitting rate*". The "*user quitting rate*" is an absolute quantity resulting from the sum of all "*user drop events*" in the considered playback session.

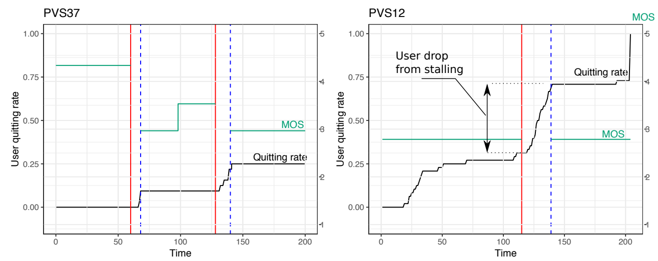

Fig. 2. Temporal evolution of users stopping video playback. MOS is estimated per second using the model [2], [21]

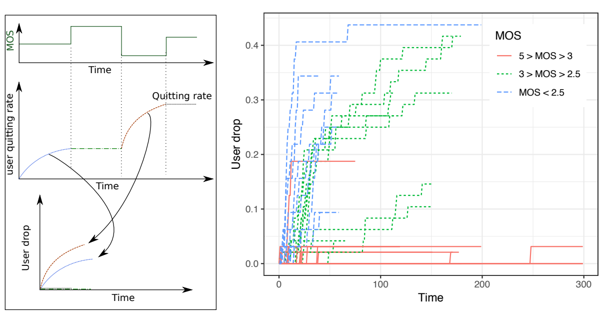

Fig. 3. Temporal evolution of users stopping video playback. Analysis is per segment of constant quality.

B. Coding quality and user drop events

To study the relationship between video coding quality and *user drop events*, this section only considers cases without stalling. Figure 3 (right) depicts the temporal evolution of *user drop event* amplitude per segment as having a constant quality level (See Table I, and Figure 1). Considering that the experiment simulated adaptive bitrate streaming, changes occurred in the quality. Therefore, to study the relationship between *user drop event* amplitude and video quality, it is proposed to perform the analysis per segment with constant quality. Figure 3 (left) depicts the process that was employed: the temporal evolution of the *user quitting rate* was measured for each segment with constant quality. Then, the axes were normalized such that the origin corresponds to the beginning of each segment. This was done both in terms of the number of users who quit and time, giving the *user drop event* amplitude per segment. Finally, the temporal evolution of the *user drop event* amplitude was related to the quality of each segment using three categories: $MOS > 3$, $3 > MOS > 2.5$, and $MOS < 2.5$, MOS values were, as previously indicated, estimated using a computational video quality model [2], [21]. This classification is motivated by the work of Schatz et al. [13], who found a threshold of acceptance for a MOS value of 2.5 and the experimental results of the current experiment. Using this classification, it can be seen that when the MOS is lower than 3, users begin to quit the video; otherwise, most users will continue watching. Going to a MOS below 2.5 result in a further increase of user quitting video. A linear model between time and *user drop* during the first 30 seconds of the segment shows that going to category $3 > MOS > 2.5$ to $MOS < 2.5$ result in an increase of *user drop* by a factor of 3.96. One exception of $MOS > 3$ and a large *user drop event* is visible in Figure 3, but it may be a residual effect of the stalling event that occurred before the beginning of the segment (as *user drop* amplitude is constant after the initial increase). Finally, it should be noted that in this analysis segments are considered independently. However, interactions are expected, and large quality changes may results in even larger *user drop* amplitude. Future work will consider these cases.

C. Stalling and user drop events

The second quality-related reason for users quitting the videos are the stalling events. The *user drop* amplitude due to a stalling event (illustrated by a double sided arrow on the right side of Figure 2) is related with key characteristics about the stalling events below. Figure 4 depicts the relationship between the percentage of users who quit the video because of the stalling event, and the stalling event position and duration. It can be seen that *user drop* amplitude increases linearly with the stalling position. With the exception of the 36 sec stalling event occurring at 150 sec (this exception being discussed further below), increasing the stalling duration adds a constant offset in terms of *user drop* amplitude. The results are consistent with the findings of Lebreton et al. [16], who found the amplitude of users quitting due to stalling linearly dependent on stalling position and stalling duration. The quality of the video before the stalling event was also computed (using the model [2], [21]) and added to the figure. No clear effect of MOS on *user drop* amplitude can be found. In Equation 1, the model from [16] is presented. In this equation, U is the *user drop* amplitude, S_d and S_p are respectively the stalling duration and position. a_{1-3} are model parameters. The performance of this model on the dataset established in this work can be found in Figure 5.

$$U = a_1 \times S_d + a_2 \times S_p + a_3 \tag{1}$$

Although simple, this model shows interesting performance (a Pearson correlation coefficient (PCC) of 0.728, and root mean square error (RMSE) of 0.153). In their work, the authors theorized that prediction accuracy could be improved by accounting for the video quality in the overall model. This was motivated by the outliers identified in the low video quality cases. However, the experimental results from our study indicate that this is the *user quitting rate* before the stalling event occurred, e.g., the percentage of users who already quit (noted Q_{alr} in Figure 5), which is of primary importance (as when cases with $Q_{alr} > 0.2$ are omitted, RMSE drop to 0.112). Nevertheless, video quality still has an effect on *user drop*, as Q_{alr} depends on it, and it increases when the video quality is low (see Section III-B).

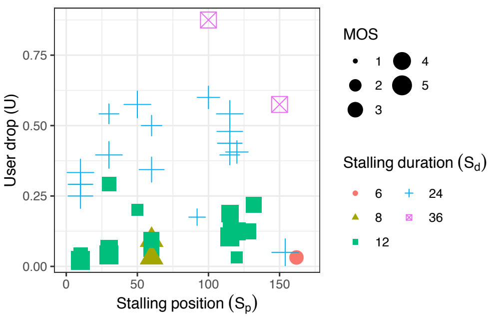

Fig. 4. Analysis of the user drop event amplitude as a function of MOS, stalling position and duration.

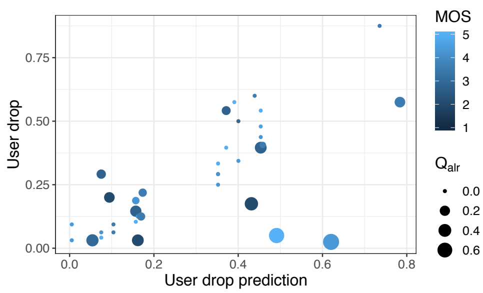

Fig. 5. Prediction using linear combination between stalling position and stalling duration (Equation 1).

In this work, an improved model to predict *user drop* amplitude due to stalling will be presented. First, several issues with Equation 1 needed to be addressed. In this formula, the stalling position is a constant positive term. Although at first look Equation 1 is in agreement with Figure 4, this result is counterintuitive because a stalling event occurring after a long uninterrupted playback would have a large effect on *user drop* amplitude. This is problematic as previous work indicates otherwise: having long uninterrupted playback results in a small buffering ratio, which is preferable [7], [10]. On the other hand, this term cannot always be negative, as users are expected to be more tolerant towards initial buffering compared to stalling during playback [15]. Therefore, the model needs to handle an inversion of the effect of stalling position on *user drop* amplitude.

To address this point, further investigation on the relationship between stalling position and *user drop* amplitude was performed, and an interaction with Q_{alr} was found. Indeed, in Equation 1, it was found that when the *user quitting rate* is null before the stalling event, a_2 has a significant contribution to the equation ($t\ value = 5.089, p < 0.001$) and is positive: $a_2 = 0.00141$, while when users have already quit, a_2 still significantly contributes to the equation ($t\ value = -3.153, p = 0.00832$) but is negative: $a_2 = -0.00160$. A multiplicative term between the stalling position and Q_{alr} is then necessary. Beyond the effect of Q_{alr}, video quality is also expected to contribute to the *user drop event* amplitude due to a stalling event. To address this, the work of Tan et al. [20], who found an interaction term between quality and user interest, can be considered. Although user interest was not evaluated in this work, a possible approach to address it is to use results on viewing time distribution: users mostly quit after an initial scan of the video, or watch fully [10]. Therefore, the amount of time the user already spent watching videos and the *user quitting rate* before the stalling event (e.g., Q_{alr}), can be thought of as ways to measure user interest towards content.

TABLE II
REGRESSION TABLE OF EQUATION 2.

Factor	Estimate	Std. Err	t value	$Pr(>\|t\|)$
Intercept (a_3 in equations 1 & 2)	-0.2967	0.03265	-9.086	$p < 0.001$
S_d	0.02577	0.001431	18.001	$p < 0.001$
S_p	0.001352	0.0002635	5.131	$p < 0.001$
$MOS \times Q_{alr}$	0.9764	0.27046	3.610	$p = 0.00128$
$S_p \times MOS \times Q_{alr}$	-0.007783	0.001094	-7.116	$p < 0.001$

Based on these observations, the final Equation 2 was established. Table II provides the regression analysis table for single stalling events. It shows that the factors considered have statistically significant contributions at 95% to the model's performance. t values report on the accuracy of coefficients estimation. A high value can be found for stalling duration, indicating a clear contribution. The other coefficients having rather balanced t values is a good sign for the model.

$$U = a_1 \times S_d + a_2 \times S_p + a_3 + MOS \times Q_{alr} \times (a_4 - a_5 \times S_p) \quad (2)$$

Figure 6 (left) shows the performance of this model in predicting *user drop event* amplitude due to a stalling event. Several outliers can be identified, which correspond to cases with multiple stalling with an already high user quitting rate before the stalling event. In these cases, the model overpredict the effect of the stalling event. Few cases are involved; therefore, the model extension to account for these is described in the discussion section as further testing will be needed.

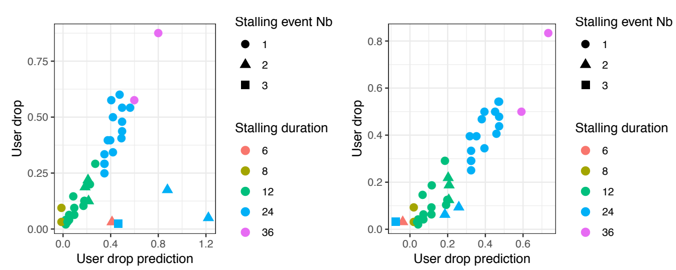

Fig. 6. Prediction of the *user drop* due to stalling. The result from Equation 2 are on the left. On the right, a nonlinear contribution was added to Q_{alr} (see Section IV-B).

IV. Discussion

A. Subjective results

In terms of subjective methodology, several points of interest were found. First, model and coefficients were constant across dataset when modeling *user drop*, therefore it seem there are no major differences between experiments "3 min-A" and "3 min-B". This is interesting as these experiments had differences in terms of experimental design: one had more repetition in the source content than the other. Therefore, repetition did not seem to alter the result of the first experiment. Future research will keep the approach without repetition, but there is some flexibility in this regard. In terms of the number of participants, using 32 participants in experiment "3 min-B" instead of 40 participants in "3 min-A" did not show any limitation. Therefore, future work will use 32 participants.

B. Regression analysis

Regarding the regression analysis, the results were found to be consistent with [16], but extra factors also had to be considered: the *user quitting rate* before the stalling event (e.g., the percentage of users who already quit), the MOS and the respective interactions with the stalling position. The model successfully addresses most cases; however, when the *user quitting rate* is already high, it was found that the model overpredicts the effect of further stalling events. It is theorized that this issue is due to a nonlinear contribution of Q_{alr} in Equation 2. Therefore, it is suggested to use a power function: $Q_{alr}^{a_6}$ (with $a_6 = 0.26$, obtained via non-linear regression) instead of Q_{alr} only. This term was also added as a separate term without interactions with MOS. Performance is depicted in Figure 6 (right), and a 5-fold cross test/validation was performed, which shows a RMSE of 0.0686, a PCC of 0.9346, and a Spearman correlation of 0.9065. Due to the interaction terms, further analysis is needed to report on the effect of MOS on the *user drop event* amplitude. By investigating the sign of $\frac{\partial U}{\partial MOS}$, it was found that an increase of MOS results in a lower *user drop* amplitude when the stalling event occurs after 125 seconds. However, before 2 min, an increase in MOS increases the *user drop* amplitude. This result means that towards the beginning of the video, it may be better to start with lower quality and then increase it rather than starting with a very high quality. This result can be used for the design of video players. Finally, it should be noted that the model described in Equation 2 uses the *user quitting rate* before the stalling event (e.g., Q_{alr}) to predict the impact of a stalling event. In this study, this value was obtained using ground truth data, but it will be calculated once the effect of low coding quality on *user drop* amplitude can be estimated, as Q_{alr} is the sum of all past *user drop events*.

V. Conclusion

In this work, the results of three different subjective experiments addressing the user quitting rate in video streaming were presented. It was found that the effect of stalling on user drops is linearly dependent on the stalling position and stalling duration. It was shown that an inversion of the effect of stalling position should be considered. An interaction between stalling position and quitting rate before the stalling event was found. The MOS was also found to be of importance in predicting user drops in stalling cases. These analysis allowed a model of *user drop* for stalling events to be proposed. The model was tested using a 5-fold cross test/validation and shows a RMSE of 0.0686, and PCC of 0.9346. Furthermore, it was found that when video quality is lower than 3, users will start to quit videos. Having a MOS lower than 2.5 was translated by an even higher number of user drops. Future work will pursue the modeling of the relationship between quality and *user drops*. Content with durations up to 10 min will also be considered to study source duration and the effect of multiple stalling events.

References

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2017–2022," available online: cisco.com.

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

[3] Z. Duanmu, K. Ma, and Z. Wang, "Quality-of-experience for adaptive streaming videos: An expectation confirmation theory motivated approach," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6135–6146, 2018.

[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] Y. Chen et al, "On distribution of user movies watching time in a large-scale video streaming system," in *IEEE ICC*, 2014.

[6] W. Robitza and A. Raake, "(Re-)Actions Speak Louder Than Words? A Novel Test Method for Tracking User Behavior in Web Video Services," in *QoMEX*, 2016.

[7] F. Dobrian, A. Awan, D. Joseph, A. Ganjam, J.Z. 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.

[8] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," in *SIGCOMM'13*, 2013.

[9] S. Shunmuga Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior : Inferring causality using quasi-experimental designs," *ACM Tran. on networking*, vol. 21, 2013.

[10] M. Vilas, X.G. Panneda, R. García, D. Melendi, and V.G. García, "User behaviour analysis of a video-on-demand service with a wide variety of subjects and lengths," in *EUROMICRO*, 2005.

[11] Y. Li, Y. Zhang, and R. Yuan, "Measurement and analysis of a large scale commercial mobile internet TV system," in *ACM SIGCOMM conference on Internet measurement conference*, 2011.

[12] 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 Transactions on Multimedia*, vol. 15, no. 8, pp. 2087, 2013.

[13] R. Schatz, T. Hoßfeld, L. Janowski, and et al., "From packets to people: Quality of experience as a new measurement challenge," in *Data traffic monitoring and analysis*, 2013, p. 219–263.

[14] H. Nam and H. Schulzrinne, "Youslow: What influences user abandonment behavior for internet video?," *Columbia University Rcp*, 2016.

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

[16] P. Lebreton, K. Kawashima, K. Yamagishi, and J. Okamoto, "Study on viewing time with regards to quality factors in adaptive bitrate video streaming," in *Workshop on Multimedia Signal Processing*, 2018.

[17] 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 *Conf. FSKD*, 2015.

[18] S. Wu, M.A. Rizoiu, and L. Xie, "Beyond views: Measuring and predicting engagement in online videos," in *Int. Conf. on Weblogs and Social Media (ICWSM)*, 2018.

[19] C. Zhou, Y. Guo, Y. Chen, X. Nie, and W. Zhu, "Characterizing user watching behavior and video quality in mobile devices," in *23rd International Conference on Computer Communication and Networks (ICCCN)*, 2014.

[20] X. Tan, Y. Guo, M.A. Orgun, L. Xue, and Y. Chen, "An Engagement Model Based on User Interest and QoS in Video Streaming Systems," *Wireless Communications and Mobile Computing*, pp. 1–11, 2018.

[21] K. Yamagishi, "Audio-visual quality extimation device, method for estimating audio-visual quality and program," in *US Patent 15/776425*, 2018.