

Study on viewing time with regards to quality factors in adaptive bitrate video streaming

Pierre Lebreton, Kimiko Kawashima, Kazuhisa Yamagishi, Jun Okamoto
NTT Network Technology Laboratories
Musashino-shi, Tokyo 180-8585

Abstract—In this work, the evaluation of user engagement’s characteristics in adaptive bitrate video streaming is addressed. To this aim, the viewing time and its relation with video quality is studied in two carefully designed subjective tests. Video quality and viewing time were addressed in distinct experiments. In case of viewing time, users were allowed to stop watching the videos when they desired. It was found that for low-quality videos, the number of users dropping the video increases logarithmically as a function of time. In addition, when a stalling event occurs, users start dropping video playback after a waiting period of 5 seconds. Then, when the stalling ends, the dropping rate returns to its baseline rate (which depends on video quality). The number of users stopping watching video after a stalling event was found to be a function of stalling position, stalling duration, and the quality affected by coding. A baseline model considering only stalling features is defined. Finally, a model for predicting the video completion rate is proposed that achieves a Pearson correlation of 0.96 and a root-mean-square error (RMSE) of 0.064.

I. INTRODUCTION

Adaptive bitrate video streaming has been used by many users in the past decade. Many users daily watch videos on multiple devices (e.g., TVs, smartphones, and tablets) and multiple services (e.g. YouTube, Netflix, and Twitch). However, due to variation of network conditions, perceived quality may be degraded. This results users having a worse viewing experience, and in the worst cases dropping the service. Therefore, viewers’ experience needs to be measured. User’s Quality of Experience (QoE) [1] has then been extensively investigated in terms of relating technical parameters to QoE in adaptive streaming scenarios. ITU-T Recommendation P.1203 [2] or video quality models such as those of Yamagishi and Hayashi [3] or Robitza et al [4] can be seen as examples of the large amount of research done on the topic. However, there has recently been a shift from the work on QoE to a new evaluation concept: user engagement. The goal of researching user engagement is to understand the extent of the engagement of users into the contents and how likely they are to quit the services. Previous work on user engagement can be categorized into two categories: studies that use data mining to understand the relationship between quality factors and user engagement, and studies that specifically attempt to establish prediction models. These will be the topics of the following subsections.

A. Understanding factors affecting user’s engagement

To study which factors affect user engagement, two types of data sources can be considered: data from laboratory

experiments and datasets collected from real life services. Each data source has its own benefits and weaknesses: laboratory experiments allow us to carefully test specific conditions but may not match real life conditions and cannot easily address complex user-related factors such as user’ preferences towards specific contents. On the other hand, real-life datasets allow us to obtain ecologically valid data but contain many sources of noise that can make the data difficult to interpret and result in fuzzy modeling.

In the specific context of laboratory experiments, Robiza and Raake [5] worked on subjective methodologies and defined a new method for measuring user engagement: participants were given a task, such as summarizing videos, to drive their attention away from the quality impairments added to a simulated video-on-demand (VOD) platform. Then, behavior of users was studied to evaluate their engagement in the given task. It was investigated to see if participants perform actions such as reloading pages, pause the video, or perform seeks in the video to restore the video playback when stalling events occur, allowing them to accomplish their task. A “good subject” phenomenon [6] showing that participants are afraid to break the experimental setup and will avoid performing any actions. Therefore, the evaluation of user engagement is a challenging topic in itself.

Such results make the use of real-life data analysis an appealing solution to study user engagement. Dobrian et al [7] used a real-life dataset that contained video player logs obtained from video streaming websites. The studied dataset contained three main categories of content: long VOD, short VOD (between 2 and 5 min), and live videos. User engagement was measured through the play time duration per video and for the overall session. These were then related with quality parameters based on the buffering ratio: the time spent watching with regards to the time spent waiting for the video to load. The buffering rate: the number of stalling events divided by the session duration, and the average bitrate. From a methodology point of view, the key result was that different analysis techniques should be used to avoid making incorrect conclusions due to specific cases in the data. Thus, three methodologies are proposed: correlation, information gain, and regression analysis. In terms of user engagement analysis results, the buffering ratio was found to be a key feature. More precisely, in the case of VOD, a 1% increase in the buffering ratio was found to lead to a 3-minute decrease in play time. However, depending on the studied case (VOD

or live streaming), the contribution of features is not equal: the buffering ratio is the most important feature for VOD, whereas the average bitrate is the most important feature for live contents. Pursuing this work, Vilas et al. [8] studied a real-life dataset from a VOD service in Spain. Two categories are used: short videos (0-5 min) and long videos (5-50 min). It was found that short and long videos do not have the same distribution of completion rates: short videos are more easily dropped than long videos. It was also observed that the histogram of viewing completion is a mixture of two distributions: one located within $[0\%, 10\%]$ and the other in $[90\%, 100\%]$. This corresponds to two different cases: users who are not interested in the content and stops the video immediately, and those who are interested and watch the video entirely. Expanding on the work of Dobrian et al. [7], Krishnan and Sitaraman [9] enhanced the results by considering causality in user engagement prediction. They used the Quasi-Experimental Design (QED) paradigm that aims to identify causality by matching a “treated” user with a “untreated” one. Using this method, they showed that an initial delay over two seconds causes viewers to abandon the video. Then, after these two seconds, a one-second increase in delay increases the abandonment rate by 5.5%. They also found that users are less tolerant of startup delays for short videos than long videos. This shows the importance of video duration when studying engagement. These results can be put into relation with those of Ahsan et al. [10] who studied the impact of the duration of the analysis window on the performance at evaluating the overall video streaming experience. In their results, it was revealed that using Koolmogorov-Smirnov test acceptance rate and autocorrelation analysis and assuming the spatio-temporal complexity of the video does not largely vary, that an analysis window of 3 min can provides sufficient information on the network performance to evaluate the overall streaming experience. This is due to the small temporal variation of typical network condition in non-mobile scenarios. Nam et al. [11] addressed the abandonment rate of YouTube videos using a plugin that allowed them to relate technical parameters with viewing time. The data, revealed that frequent short re-buffering events resulted in a lower abandonment rate than a unique buffering event with the same length. Rebuffering was found to induce an abandonment rate six times higher than start up delay, and a single buffering event had three times the impact of bitrate change. The effect of the bitrate on user engagement was addressed specifically in the work of Wang et al [12]. This was based on large data collection and showed that viewing time not only increases for a higher bitrate but also strongly highly depend on the content (cartoon, sport, entertainment, etc.). Finally, Ahmed found that quality adaptation also strongly affects viewing time [13].

B. User engagement prediction

Previous work have leveraged the large scale data analysis to establish models enabling viewing time to be predicted. Balachandran et al. [14] developed a decision tree model based on technical features such as buffering ratio, buffering rate,

average bitrate, content duration, and type of content (VOD or Live). Consistently with previous work [7], [8], Balachandran et al. found that the importance of features differs depending on the type and length of the videos. Therefore, several decision trees had to be developed to address each use case. Decision trees were chosen instead of a black box model such as support vector regression or a neural network as they allow an interpretable model to be obtained in terms of feature contribution and also perform well in this study case. It could then be observed that for VOD, the top node is the buffering ratio, while for live content, the top node is the average bitrate, then buffering ratio and then buffering rate. These results are consistent with those of Vilas [8]. Pursuing this work, Shafiq et al. [15] took the point of view of a network operator rather than a VOD service provider. The key difference in their work lies in the feature selection: a network operator has difficulty obtaining information on the video player status needed to predict the user engagement. This would rely on deep packet inspection, which requires using much processing power, collecting too much video playback data or HTTP-related data, and collecting too much private data about the users. Therefore, the model was designed to address mobile networks and use features specific to wireless networks in addition to network parameters. Similarly to Balachandran et al. [14], Shafiq et al. [15] used decision trees and found that they predicted completion of video with high accuracy.

C. Contributions

A large amount of work has been done on user engagement, but the key limit of the previous studies is the current lack of a unified and comprehensive framework allowing viewing time to be modeled. Indeed, many studies have provided key results on factors affecting viewing time duration, but since this topic was mostly addressed using real-life datasets many complex interactions and hidden confounding variables are involved. This makes a comprehensive viewing time model difficult to be establish and results in the use of machine learning, which provides models with limited interpretability. Therefore, in this work, it was decided to base the modeling on laboratory experiments, enabling the final goal of a unified and comprehensive model to be obtained.

In the remainder of this paper, Section II describes the design of two distinct subjective experiments that were conducted. Section III analyzes the subjective results relating stalling and coding with perceived quality and viewing time. Then, Section IV presents a model that uses quality parameters to predict completion rate of video watching. Section V addresses the limits and future work. Finally, Section VI concludes this paper.

II. SUBJECTIVE EXPERIMENT

Two distinct subjective experiments were conducted. In the first, participants were asked to watch 3 min long videos and rate their overall quality. In the second participants were able to watch the same videos as in the first experiment and were allowed to stop watching the video at any point. The goal of

TABLE I
LIST OF PROCESSED VIDEO SEQUENCES (PVS)

SRC	PVS	Coding condition	Initial loading delay	Stalling duration @ 30 sec	Stalling duration @ 115 sec
01	01	1920 x 1080 @ 10Mbps	0	0	0
01	02	1920 x 1080 @ 10Mbps	0	12	0
01	03	1920 x 1080 @ 10Mbps	0	0	12
01	04	1920 x 1080 @ 10Mbps	0	24	0
01	05	1920 x 1080 @ 10Mbps	0	0	24
01	06	1920 x 1080 @ 10Mbps	12	0	0
01	07	1920 x 1080 @ 10Mbps	24	0	0
02	08	426 x 240 @ 0.2Mbps	0	0	0
02	09	426 x 240 @ 0.2Mbps	0	12	0
02	10	426 x 240 @ 0.2Mbps	0	0	12
02	11	426 x 240 @ 0.2Mbps	0	24	0
02	12	426 x 240 @ 0.2Mbps	0	0	24
02	13	426 x 240 @ 0.2Mbps	12	0	0
02	14	426 x 240 @ 0.2Mbps	24	0	0
03	15	854 x 480 @ 1Mbps	0	0	0
03	16	854 x 480 @ 1Mbps	0	0	24
03	17	854 x 480 @ 1Mbps	24	0	0
03	18	854 x 480 @ 1Mbps	0	12	12
03	19	Bitrate adaptation	0	0	0
03	20	Bitrate adaptation	0	0	24
03	21	Bitrate adaptation	24	0	0

the experiments was to relate quality and playback issues such as stalling events with the viewing time. In the following, the test conditions common to the two experiments will be detailed followed by information specific to either experiment.

A. Test conditions

In the experiments, three different source contents (SRC) were used. These were full high definition (HD) content with a frame rate of 30 frames per second (FPS). Each SRC was 180 seconds long. The videos were all shot in Japan and depict scenery in Enoshima (SRC 1), a samba festival (SRC 2), and rafting (SRC 3). SRC 1 had a relatively low-coding complexity, while SRC 2 and SRC 3 had large and complex motion making them challenging to encode. Considering the large amount of previous work in the literature addressing video quality and viewing time, key factors affecting the viewing time were identified: the video quality, the video quality variation, the stalling duration, and the stalling position. Table I depicts the full matrix of test conditions. In all cases, the frame rate was maintained at 30 FPS, and the codec was H.264 in High Profile. Different coding conditions and resolutions were considered. This allows the effect of the coding condition on quality and viewing time to be studied. The “bitrate adaptation” case corresponds to an oscillation between 720p @ 1.6 Mbps and 360p @ 0.5 Mbps with a periodicity of 30 seconds. Then, from the table, it can be seen that stalling conditions were designed to occur at three different position in time: initial loading, towards the beginning of the video (30 seconds), and towards the end of the video (115 seconds). In addition, stalling events were repeated with durations of 12 and 24 seconds to allow the effect of the stalling duration to be studied for different positions compared with no stalling. All stalling scenarios were repeated for the different coding conditions to allow considering the interaction

with coding distortions. Noted that one 12 second stalling event was replaced by multiple stalling events at 30 second and 115 seconds to allow single and multiple stalling events of equal total stalling duration to be compared. Finally, audio was always encoded at 256kbps using low-complexity advanced audio coding (AAC-LC) and had no perceivable distortions. This resulted in 21 processed video sequences (PVS).

B. Common conditions and procedure

The 21 PVS previously described were then used in the subjective experiments. The playback was performed on a 5.2 inch smartphone with a full HD resolution. Participants used headphones to listen to the audio. The viewing distance was set to 5-7H (H being the picture height) and listened to each PVS at -21 dB. The room was a standardized laboratory environment designed for video tests: gray room, noise cancellation, controlled lighting, etc. The illumination was set 20 lux, which correspond to a dark room. After passing the visual acuity tests, participants were provided written instructions about their task and then, went through a training phase in which six 90-second PVSs were shown allowing them to practice their task. After this, the main task of the experiment started. The main task was divided in three sessions of seven PVSs. After each session, participants were given a rest. The presentation order of PVS was randomized across subjects.

C. Quality experiment

The video quality followed the absolute category rating (ACR) methodology using a five-grade scale. The participants were asked in Japanese, “How would you rate the video quality?”, the grades being: “5: Excellent”, “4: Good”, “3: Fair”, “2: Poor”, “1: Bad”. For this experiment, 24 participants (12 males and 12 females) passed the screening tests and participated. They were non-experts with no previous experience of assessing audiovisual quality as part of their work.

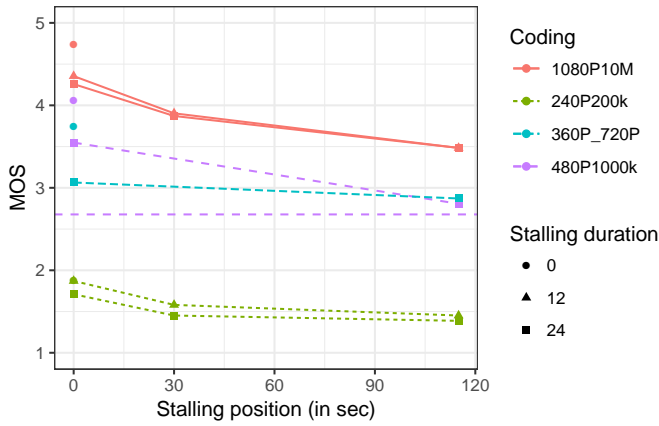


Fig. 1. Quality evaluation as a function of stalling position, duration, and coding quality. The horizontal dashed line corresponds to PVS18 which had two stalling events at 30 and 115 seconds.

D. Viewing time experiment

In the viewing time experiment, participants were not given any task and only had to watch the videos on the smartphone. They were instructed that they were allowed to stop watching whenever they wanted. In this test, 40 participants (20 males and 20 females) took part. The viewing time experiment was conducted on a different day from the quality experiment, hence the participants were different.

III. EXPERIMENTAL RESULTS

This section presents the relationship between the test condition (Hypothetical Reference Circuit, HRC) and each individual test. The relationship between experiments is addressed in Section IV.

A. Video quality experiment

Figure 1 depicts the results of the quality test. It can be observed that coding conditions were distributed such as the HD contents received a high mean opinion score (MOS) (4.7), while 240p contents received a low one (1.9). Quality adaptation between 360p and 720p received a lower MOS than the constant 480p case (3.7 vs. 4.1). An initial stalling event of 12 seconds caused a quality drop of 0.39 MOS in the HD case but did not affect the 240p case. An initial stalling event of 24 seconds caused a drop of 0.5 MOS in the higher quality scenarios (1080p and 480p) but a drop of 0.16 MOS in the 240p case. This shows that the perceived quality of initial loading depends on both duration and the video coding quality. Moreover, it can be observed that quality decreases when stalling events occur towards the end of the video. This relates to the well-known recency effect. Although not statistically significant, it can be observed that 24-second stalling events induced a lower quality rating than 12-second ones. Consistent with the initial buffering scenario, the effect of stalling events at different positions in time induced quality degradation, the amount of which depends on the coding quality. Finally, it can be observed that the PVS containing two

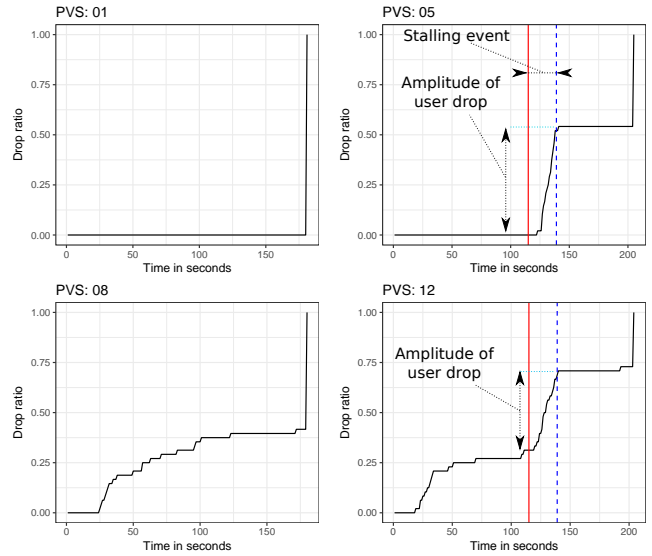


Fig. 2. Temporal evolution of users stopping video playback for different conditions.

12-second stalling events (PVS 18, represented by a horizontal long dashed purple line in Figure 1), was rated lower than the PVS containing a single 24-second stalling event. However, the differences are not statistically significant.

B. Viewing time experiment

The main novelty of this work lies in the viewing time experiment. Due to limited space, not all the data can be depicted. Therefore, Figure 2 illustrates four characteristic distributions of the users' dropping rates. This figure shows the temporal evolution of the number of users who decided to stop the videos as a function of time. A "drop ratio" of 0.3 indicates that 30% of the 40 users decided to quit the video. Figure 2 illustrates four different scenarios: high/low coding quality, with/without stalling events. In the case of high-quality videos without stalling events (scatter plot in the top-left), all users watched the video until the end. However, when the same high-quality video contained a stalling event of 24 seconds (scatter plot in the top-right), several users stopped the video while the stalling event was occurring. First, when a stalling event occurs, a delay of around 5 seconds was found before users start dropping the video. This was consistent across PVSs with high coding quality (1080p, 480p). As for the other PVSs, due to the shape of the temporal evolution of drop, this delay is less easily identifiable (see PVS 12 in Figure 2). Then, across all PVSs, it was found that when a stalling event occurs, users will drop the video. Then, when the stalling event stops, users will behave the same as before the stalling event. Therefore, the data shows that if the users stay with the video until the end of the stalling event, they will likely stay with the video until the next quality degradation occurs. In the case of low coding quality, an initial delay of 25 seconds is visible before users start dropping the video. An example is depicted in PVS 08, Figure 2, bottom-left.

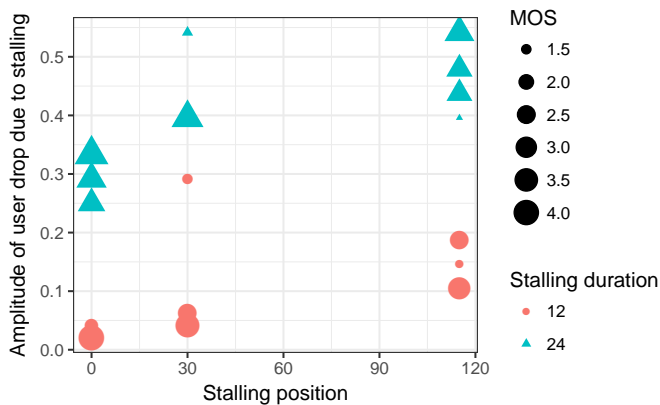


Fig. 3. Analysis of the number of users who quit the video due to stalling event as a function of MOS and stalling duration and position.

The dropping rate is particularly interesting for the low coding quality cases, as the number of users dropping the video was found to be logarithmically related to time. Then, when the stalling event occurs, consistent with the high quality case, the number of users who stop watching the video increases largely during the stalling event. In the special case in Figure 2, the number of users stopping the video because of the 24 second stalling event is constant and is approximately 50%. An in-depth analysis of the consequences of a stalling event can be found in Figure 3. The number of users stopping the video because of the stalling event is a linear function of the stalling position. If the stalling happens towards the end of the video, a larger number of users will quit the video. The proportion of users stopping the video also depends on the stalling duration: increasing the stalling duration from 12 to 24 seconds resulted in a constant increase of 30% of users stopping the video. It should be stressed that this value is constant for each stalling position. Therefore, in a model of amplitude of user drops, the stalling duration could be considered as an additive term. Finally, several outliers corresponding to low quality coding conditions can be identified. It can then be concluded that, in a model for predicting amplitude of user drops, stalling duration, stalling position, and MOS should be considered.

IV. RELATIONSHIP BETWEEN QUALITY AND VIEWING TIME

In this section, the results from the video quality experiment will be put into relation with the results from the viewing time experiment. In this paper, the analysis is restricted to the analysis of the number of users who watched the video fully. We omit the temporal evolution of the number of users watching videos. Figure 4, depicts for each video the percentage of users who completed the playback as a function of the video quality ratings obtained from the video quality experiment. In this plot, three main results can be seen: 1) a logarithmic relationship between MOS and completion rate can be observed, 2) stalling events of 24 seconds greatly affected the completion rate, and 3) quality degradation due to stalling events of 12 seconds affected similarly the completion

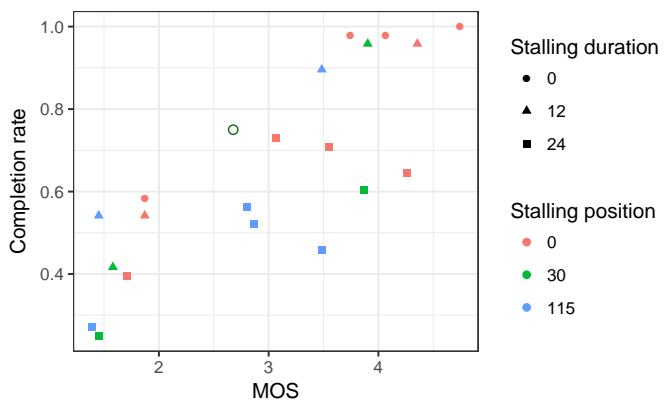


Fig. 4. Relationship between the percentage of users who watched the entire videos in the viewing time experiment and subjective MOS rating.

rate as the degradation from coding conditions only. Considering only coding conditions, a simple base model as defined in Equation 1 can be used to predict the completion rate. This model shows a root-mean-square error (RMSE) of 0.034 and a Spearman correlation of 0.948 (the Pearson correlation is artificially high due to the data distribution). Interestingly, according to this model, in case of coding condition only, the completion rate increases roughly by half of the MOS's logarithm.

$$C.R = 0.48 \times \ln(MOS) + 0.296 \quad (1)$$

To consider the cases with stalling conditions, the residual error of the base model defined by Equation 1 was studied and was found to be related to both MOS values and stalling duration. A candidate model is then defined as in Equation 2.

$$C.R = 0.521 \times \ln(MOS) + 0.244 - 1.78 \times MOS \times \exp\left(\frac{-73.98}{Stall.Dur + \epsilon}\right) \quad (2)$$

Figure 5 depicts the relationship between the predictions of the model defined in Equation 2 and the ground truth data. The accuracy is defined by a Pearson correlation of 0.96, a Spearman correlation of 0.93 and a RMSE of 0.064.

V. DISCUSSION

Several limitations of this work need to be discussed. First, only three source contents were used. Since this work does not relate bitrate and video quality, the limited number of sources may not have been a big issue. However, for the viewing time experiment, the usage of repeated source content may have induced fatigue in the participants and biased the results. This statement should be balanced by the fact that a large number of participants took part (40), three breaks were added, and PVS were randomized across participants. Therefore, even though the conditions were not ideal from the participants' point of view, the repetition of source content may not have affected the results too much. Nevertheless, in the future steps of this work, more sources will be used, and we may use

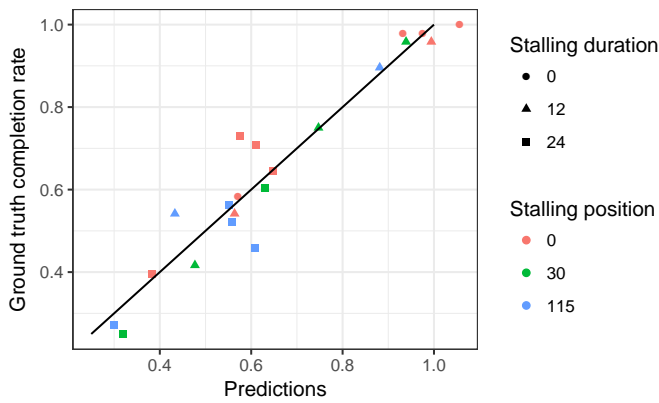


Fig. 5. Prediction of video completion based on video quality rating and stalling duration.

an immersive design [16]. The second limit of this work was found in the definition of the coding condition, as can be seen in Figure 1, coding conditions were not sufficiently balanced. This has been problematic for the modeling of the number of users dropping video due to a stalling event: the 240p conditions were greatly different from the other conditions. It was found that without considering these conditions, users dropping videos as a consequence of stalling can be modeled as a function of stalling position and duration (see Equation 3). This model has a high accuracy: a Pearson correlation of 0.97 and a root mean square error (RMSE) of 0.04. However, to integrate the 240p resolution cases, the effect of coding on video quality will need to be accounted for and this will require further testing.

$$User.Drop = A \times Stall.Pos + B \times Stall.Dur + C \quad (3)$$

Future work will pursue the modeling of the temporal evolution of the number of users as a function of time. A first basis for this model was put in this paper in the user drop analysis (Section III-B) and user drop model (Equation 3). However, video quality still needs to be integrated, which will require further subjective tests to be conducted that include more intermediate coding, stalling conditions, and SRCs.

VI. CONCLUSION

In this work, viewing time of adaptive bitrate video streaming was evaluated. Two well-defined subjective tests were conducted to relate quality factors such as stalling duration, position, and coding quality with both video quality and viewing time. It was found that for low-quality videos, the number of users dropping the video increase logarithmically as a function of time. When a stalling event occurs, users start dropping video playback after a waiting period of 5 seconds. Then, when the stalling ends, the dropping rate returns to its baseline rate (which depends on video quality). The number of users who stop watching a video due to a stalling event was found to be a function of stalling position, stalling duration, and mean opinion score (MOS). A baseline model using only

stalling features was defined. Finally, a model for predicting the video completion rate was proposed and achieve a high accuracy.

REFERENCES

- [1] Brunnström et al., “Qualinet white paper on definitions of quality of experience,” in *www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf*, 2013, Novi Sad Version 1.2.
- [2] “Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport,” in *ITU-T Recommendation P.1203*, Nov. 2016.
- [3] Kazuhisa Yamagishi and Takanori Hayashi, “Parametric quality-estimation model for adaptive-bitrate-streaming services,” *IEEE TRANSACTIONS ON MULTIMEDIA*, vol. 19, no. 7, pp. 1545–1557, 2017.
- [4] Werner Robitza, Marie-Neige Garcia, and Alexander Raake, “A modular HTTP adaptive streaming qoe model - candidate for itu-t p.1203 (“p.nats”),” in *Quality of Multimedia Experience (QoMEX), 2017 Ninth International Conference on*, 2017.
- [5] Werner Robitza and Alexander Raake, “(re-)actions speak louder than words? a novel test method for tracking user behavior in web video services,” in *Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016.
- [6] A. L. Nichols and J. K. Maner, “The good-subject effect: Investigating participant demand characteristics,” *The Journal of general psychology*, vol. 135, no. 2, pp. 151–165, 2008.
- [7] Florin Dobrian, Asad Awan, Dilip Joseph, Aditya Ganjam, Jibin Zhan Conviva, Vyas Seka, Ion Stoica, and Hui Zhang, “Understanding the impact of video quality on user engagement,” *Communications of the ACM*, vol. 56, no. 3, pp. 9199, 2013.
- [8] X.G. Panñeda M. Vilas, 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 *Proceedings of the 2005 31st EUROMICRO Conference on Software Engineering and Advanced Applications (EUROMICRO-SEAA05)*, 2005.
- [9] S. Shunmuga Krishnan and Ramesh K. Sitaraman, “Video stream quality impacts viewer behavior : Inferring causality using quasi-experimental designs,” *ACM Transaction on networking*, vol. 21, no. 6, pp. 2001–2014, 2013.
- [10] Saba Ahsan, Varun Singh, and Jörg Ott, “Impact of duration on active video testing,” in *NOSSDAV '16 Proceedings of the 26th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, 2016.
- [11] Hyunwoo Nam, Kyung-Hwa Kim, and Henning Schulzrinne, “Qoe matters more than qos: Why people stop watching cat videos,” in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, 2016.
- [12] 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 *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2015.
- [13] Adnan Ahmed, Zubair Shafiq, Harkeerat Bedi, and Amir Khakpour, “Suffering from buffering? detecting qoe impairments in live video streams,” in *IEEE International Conference on Network Protocols - ICNP*, 2017.
- [14] Athula Balachandran, Vyas Sekar, Aditya Akella, Srinivasan Seshan, Ion Stoica, and Hui Zhang, “Developing a predictive model of quality of experience for internet video,” in *SIGCOMM13*, 2013.
- [15] Zubair Shafiq, Jeffrey Erman, Lusheng Ji, Alex X. Liu, Jeffrey Pang, and Jia Wang, “Understanding the impact of network dynamics on mobile video user engagement,” in *SIGMETRICS14*, 2014.
- [16] Werner Robitza, Marie-Neige Garcia, and Alexander Raake, “At home in the lab: Assessing audiovisual quality of http-based adaptive streaming with an immersive test paradigm,” in *7th International Workshop on Quality of Multimedia Experience*, 2015.