JOINT EFFECT OF STALLING AND PRESENTATION QUALITY ON THE QUALITY-OF-EXPERIENCE OF STREAMING VIDEOS

Hojatollah Yeganeh[1], *Farzad Qassemi*[1,2] *and Hamid R. Rabiee*[1]

[1]AICT Innovation Center, Sharif University of Technology, Tehran, Iran
[2]Shahid Beheshti University, Tehran, Iran
Email: hyeganeh@ieee.org, f_qassemi@sbu.ac.ir, rabiee@sharif.edu

ABSTRACT

Over-the-top (OTT) video streaming services have been growing rapidly in the last decade, and thus increasing Quality-of-Experience (QoE) of end users is of great interest in emerging services. Although, numerous subjective studies have been conducted to investigate the impact of video presentation quality and playback interruptions, understanding the interactions between impairment types is still an open problem. In this work, we develop a streaming video dataset that contains compressed videos with different distortion levels as well as playback stalling events. Then, a subjective user study is performed to measure the QoE of these videos. The results of our experiment reveal strong dependency between video presentation quality and playback interruptions that provides useful insight for designing QoE models in video streaming.

Index Terms— Video streaming, Quality-of-Experience, Playback stalling, Video presentation quality.

1. INTRODUCTION

Video content accounted for 64% of all the world's internet traffic in 2014, and according to a new report from Cisco, by 2019, online video will be responsible for 80% of global Internet traffic [1]. Google's official statistics show since March 2014, the number of YouTube viewers per day has yearly increased by 40% [2]. The rise of demand for watching internet videos introduces many technical challenges that affects user's perceived Quality-of-Experience (QoE). The importance of monitoring user's QoE has motivated well-known streaming companies to focus on the performance of video delivery networks. Moreover, Internet Service Providers (ISPs) and other stakeholders have started to publish reports about expected video QoE [3, 4]. Therefore, assuring QoE of end users is highly desirable in streaming applications. However, the existing services lack reliable approaches to quantify QoE of online videos. Instead, parameters such as video encoding bitrate, video resolution or stall duration have been used to estimate the perceived QoE.

Playback interruption is known to be one of the major components that influences user experience and may even reduce viewing or engagement time [5]. Statistics reported in [6] states that 1% increase of stall duration reduces the engagement time by 14 minutes. Extensive effort has been made to understand the impact of stalling on perceived user experience and to propose techniques to alleviate re-buffering impairment. One of the first attempts to study stalling is presented in [7] that shows interruption is quite annoying for viewers, and suggest service providers to increase start-up delay to eliminate re-buffering during playback. The lower impact of start-up delay compared to playback stalling is also confirmed in [8] and [9]. A large-scale subjective experiment is conducted in [8] to investigate how frequency, position and duration of stalling events affect QoE of mobile videos. Few objective models are also developed to predict the perceptual impact of playback interruptions by taking initial buffering and re-buffering events into account [10–12].

Video compression introduces different types of visual degradation and alters video presentation quality. Considering no transmission or delivery issues such as stalling or quality switching, presentation quality of video is a major component of determining viewer's QoE. Numerous works have been done to study the effect of compression on video quality by either conducting subjective user study [13] or proposing objective quality assessment measures [14, 15]. The emerging internet video streaming applications have attracted researchers to carry out subjective user study on the presentation quality of online videos. Chen et al [16] performed subjective user study and developed a QoE database for HTTP adaptive streaming that contains compressed video sequences with temporally variable encoding bitrates. In [17], a subjective video quality assessment database was constructed where the impact of compression followed by scaling was observed.

In video streaming applications, the throughput of delivery channels may vary significantly, and thus stalling occurs when the available bandwidth falls below receiving video bitrate. A straightforward solution to avoid stalling events is to lower video presentation quality by reducing encoding bitrate or resolution, but this may affect user's experience. Therefore, a reliable QoE measure that considers both factors is highly demanding. Little has been done to study the joint effect of playback stalling and video presentation quality. Garcia et al. [18] observed an additive impact of playback interruptions and compression on the users QoE and suggest that the effect of stalling is independent of video contents at higher bitrates. It is shown in [19] that playback interruption and video presentation quality are independent of each other, while this hypothesis is challenged in [9]. As one of first attempts , a subjective study of user's experience on encoded videos with playback stalling is performed in [9]. Despite presenting valuable results, the scope of their study is quite limited as they investigated the impact of only initial buffering and a single stall on perceived experience of videos with different encoding bitrate. Moreover, their conclusion suffer from the lack of enough statistical analysis. In this work, we construct a new database that contains compressed videos with varied bitrates and resolutions, and insert stalling events at various positions and with different frequency of occurrence. Then, a subjective test was conducted to collect users QoE of the test videos. The analysis of subjective scores reveals new findings about dependency between video presentation quality and playback stalling.

2. DETAILS OF THE SUBJECTIVE STUDY

To study the interactions between video presentation quality and playback stallings, we need a set of rated videos by human subjects. To the best of our knowledge, most of existing datasets for video quality assessment are meant to study either the effects of different types of distortions caused by video codecs or the impact of re-buffering only on perceived quality of experience [13, 19], making it difficult to observe the dependencies and interactions between these artifacts that are usually concurrent in internet video streaming. Therefore, we aim to conduct a subjective study and develop a new dataset that can be used to observe the joint effect of stalling and video presentation quality.

Based on recently published ITU standard for subjective assessment of internet videos [20], we chose 7 high-quality and common licensed video sequences of 1920x1080 resolution and 10-second long [21]. The source videos are selected to cover various contents, including animation, sports, natural scenes, plants, humans, and movie. The detailed specifications of those videos are listed in Table 1.

Table 1: The details of reference source videos (SRCs).

Index	Name	Frame Rate	Bitrate(Mbps)
1	Big-buck Bunny	24	189
2	Costa-Rica	30	163
3	Football	30	143
4	Elephant Dream	24	105
5	Sky Dubai	24	46
6	Sport	24	68
7	Bungy Jumping	24	180

The H.264 codec was adopted to encode each source video at three bit-rates and resolutions shown in Table 2 to obtain different video presentation quality levels. The bitrate-resolution pairs are selected based on values used in real online video streaming applications to cover a wide range of presentation quality.

Table 2: The bit-rates and resolution pairs used to encode videos.

Video presentation quality level (Q)	Bitrate(Mbps) - Resolution
High	10 Mbps - 1080p
Moderate	500 Kbps - 480p
Low	200 Kbps - 240p

Re-buffering impairment is quite common in internet videos. It is shown that the frequency of stalls, their location, and the duration of stalling events are the most important factors that contribute to the experience of users [8] compared with motion information and total video length. Here, we focus on frequency and position of stalls and thus varying stall duration is avoided. A 5-second stalling event with different frequency is inserted at either the beginning, the middle or the end point of encoded sequences. To observe the impact of stall frequency, the 5-second re-buffering is broken into two stalls. Moreover, to simulate practical scenarios, a 2-second start-up delay is added to all sequences that include playback re-buffering. Fig. 1 depicts a summary of stalling patterns used in our experiment.

Considering three video presentation quality levels, uninterrupted videos and 6 stalling patterns, there are 21 so-called hypothetical reference circuits (HRC) or test conditions. Therefore, combination of 7 source sequences (SRC) and 21 HRC results in 147 processed video sequences (PVS) that contain different video presentation quality levels as well as playback interruptions.

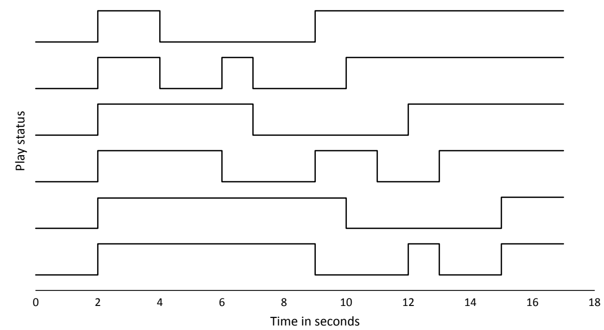

Fig. 1: Re-buffering patterns.

A computer with Intel(R) Core(TM) i7-4790 dual 3.60GHz CPU was used in the subjective user study. The test environment was setup as a normal indoor office workspace with ordinary illumination levels. All videos are displayed at 1920x1080 resolution on an HD LCD monitor at full screen resolution with Truecolor (32bit) at 60Hz. The Viewing distance and the monitor were adjusted and calibrated in accordance with the recommendations of ITU-T P.913 [20]. A customized graphical user interface (GUI) was used to render the videos on the screen with random order during the test.

We adopted the Absolute Category Rating (ACR) method [22, 23] to obtain quality ratings on video sequences where the subjects scored their viewing experience on a continuous scale ranging between worst (0) to best (100). To mitigate the effect of variation in the quality of test videos, an ACR with hidden reference method was used in our study. A total of 29 naïve observers, 12 females and 18 males aged between 22 and 35, were invited to participate in the subjective experiment. A brief introduction was given to each subject before running the test and a training session was prepared to make subjects more familiar with impairment types and the test environment. During the test, subjects were asked to watch a single video at one time, and give their opinions about the video QoE after playback. For each subject, the whole study took about one hour, and to minimize the influence of fatigue on users ratings, subjects scored videos in two sessions with a 10-minute break in-between.

3. ANALYSIS

After running the subjective test, a statistical analysis was performed based on the outlier detection scheme in [22], and four subjects were identified to be outliers and the corresponding scores were removed. The remaining 25 subjective scores are then converted to Z-scores based on the sample mean and standard deviation [13]. Finally, the subjective scores for each video sample were averaged to a Mean Opinion Score (MOS).

In the subsequent analysis, we considered the MOS computed for each video as the "ground truth", which can be used to evaluate the performance of an individual subject by comparing his/her scores with MOS for all the test sequences. We utilized the Pearson Linear Correlation Coefficient (PLCC) and Spearman's Rank-order Correlation Coefficient (SRCC) as the comparison criteria. Both PLCC and SRCC ranged within 0 and 1, corresponding to the worst and the best performance, respectively. The results of computing PLCC and SRCC criteria for all subjects are depicted in Fig. 2. It can be observed that most of subjects are in high conformance with MOS scores as they have high PLCC and SRCC values with MOS. The average performance across all individual subjects and the standard deviation between them are provided in Table 3, and are also plotted in Fig. 2.

Table 3: Mean/Std of subject performance

	PLCC	SRCC
μ	0.84	0.83
σ	0.064	0.058

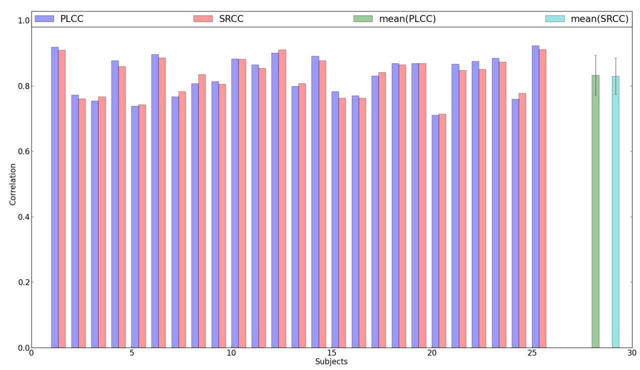

Fig. 2: Performance evaluation of individual subjects using MOS as the ground truth.

By analyzing the collected subjective data, we may have a number of empirical observations. Although these observations are only qualitative, they provide useful insights in understanding the joint impact of playback re-buffering and video quality on users experience and in developing quantitative models that approximate human judgment. These observations are presented in the following sections.

3.1. Stalling Position & Video Quality

To vary perceived video presentation quality, each SRC is encoded using three different bitrate-resolution pairs given in Table 2. The MOS for the test videos with no stalling impairment is plotted in Fig. 3, where all MOS values can be classified into three distinct quality groups with almost no overlap. It is also seen that subjective scores for uninterrupted videos span from “Poor” to “Excellent” by using the ACR method. For the sake of clarity and throughout this paper, we use green, orange and red color codes to represent high, moderate and low video presentation quality, respectively.

In order to investigate the impact of stalling position in different presentation quality levels, we computed the user’s drop off experience caused by stalling events, using the Difference Mean Opinion Scores (DMOS) of the videos with and without stallings as:

$$\mathrm{DMOS}_i(Q, L, n) = \mathrm{MOS}_i(Q) - \mathrm{MOS}_i(Q, L, n), \qquad (1)$$

where $i \in \{1, \ldots, 7\}$ labels videos according to Table 2. $\mathrm{MOS}_i(Q)$ and $\mathrm{MOS}_i(Q, L, n)$ denote MOS of i^{th} video with presentation quality $Q \in \{\text{High}, \text{Moderate}, \text{Low}\}$ without any stalling and stalls at location $L \in \{\text{Beginning}, \text{Middle}, \text{End}\}$ with stalling frequency $n \in \{1, 2\}$, respectively. Table 4 shows the average $\mathrm{DMOS}(Q, L)$

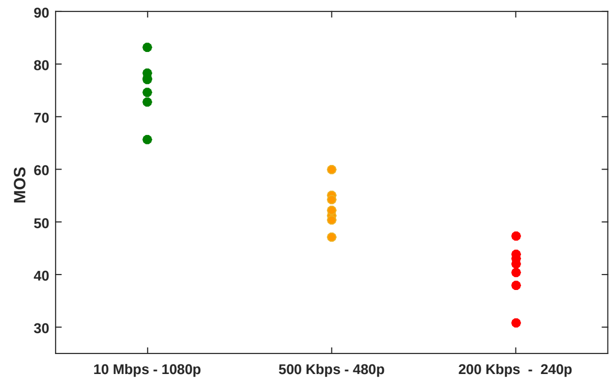

Fig. 3: MOS of uninterrupted videos.

Table 4: The average $\mathrm{DMOS}(Q, L)$ of videos with stalling events occurring at different positions.

Encoding parameters	Beginning	Middle	End
10 Mbps - 1080p	10.83	10.31	10.34
500 Kbps - 480p	6.49	6.49	6.5
200 Kbps - 240p	7.14	6.84	6.98

of videos with stalling events happening at different positions including single, $n = 1$, and two, $n = 2$, re-buffering events.

Several interesting observations can be made from Table 4. First, stalling always results in drops in MOS, and the drops could be significant. Second, the impact of playback interruption is considerably stronger when video presentation quality is fairly high. This is noticeable by looking at the first row in Table 4. Third, comparing videos of relatively poor quality and of moderate quality, drop in user experience as a result of re-buffering events is more evident in former, which implies playback interruption in videos with moderate presentation quality is more tolerable.

Although for videos with moderate quality MOS drops in all three positions are almost identical, the second and the last columns in Table 4 imply that interruptions occurring at the beginning and the end of videos with high and low presentation quality have greater influence on the perceived quality of experience compared to stalling events at the middle. This difference, might be interpreted as an evidence for existing the serial position phenomena [24] or equivalently “primacy and recency” effects. The so-called “primacy and recency” effects means people highly remember the first and the last items in a series while they most likely forget items in the middle of a given sequence. Moreover, the average DMOS values reported in the table suggest that for our videos, primacy effect is more noticeable than the recency effect. The above observations might be seen in contradiction with well-known recency effect that indicates human judgment is biased to the most recent unpleasant viewing experience [25–27]. This inconsistency is perhaps due to the fact that the current duration of test videos is 10 seconds, which is less than the human short-term working memory [26].

To further investigate the joint effect of re-buffering position and video perceived quality, we carry out Wilcoxon rank-sum test [28] to examine whether above mentioned statements are statistically confirmed. Wilcoxon is a non-parametric test and does not require any assumption for underlying distribution.The null hypothesis in our test is that the MOS drop at different stalling positions are independent random samples with equal medians against the alternative hypothesis. More precisely, we are interested to observe whether differences reported in the columns of Table 4 are statistically significant.

The definition of null hypothesis is given in Eq. (2) where $\mathrm{DMOS}(Q, L)$ denotes the subjective scores given to the video with presentation quality Q, with stalls occurring at different locations.

$$H_0 : \mathrm{DMOS}(Q, L) = \mathrm{DMOS}(Q', L'). \quad (2)$$

We consider two scenarios: first, we study the effect of location for every presentation quality, i. e., $Q = Q', L \neq L'$. Second, we consider the effect of changing presentation quality for every stalling location, i. e., $Q \neq Q', L = L'$.

Results of the Wilcoxon rank-sum test performed on the subjective data collected for videos of the same presentation quality ($Q = Q'$) is presented in Table 5. Each entry in the the table is a codeword consisting of 3 symbols corresponding to video presentation quality. As such symbols denote the "High", "Moderate" and "Low" in that order (for clarity, it has been also color coded as Green Yellow Red, respectively). A symbol value of "−" indicates that the p-value resulted from the Wilcoxon rank-sum test between the DMOS given to the videos with interruption at the location indicated in the row and that of in the column is greater than 0.05 suggesting weak evidence against the null hypothesis. To avoid confusion, we did not present the result of Wilcoxon test for second scenario, $Q \neq Q'$, in Table 5. Our analysis shows that for all possible variations, the difference between different presentation qualities is statistically significant for every location. That is,

$$p(\mathrm{DMOS}(Q, L), \mathrm{DMOS}(Q' \neq Q, L)) \equiv +. \quad (3)$$

The $+$ symbol signifies the fact that the p-value resulted from the Wilcoxon rank-sum test between the DMOS for different presentation quality is less than 0.05, i. e., the null hypothesis is rejected in the favor of alternatives.

Table 5: Results of the Wilcoxon test performed on DMOS of videos with various stalling locations.

	$Q = Q'$		
	Begin	Middle	End
Begin	− − −	− − −	− − −
Middle	− − −	− − −	− − −
End	− − −	− − −	− − −

Our statistical analysis demonstrate that for short videos of the same quality, we may not observe a remarkable difference in human satisfaction as stalling location varies. This result may also be interpreted as a strong support for the premise of extending human working memory beyond 10 to 15 seconds that are typical sequence duration people use while conducting subjective studies of perceived QoE of videos [21, 22, 29].

3.2. Stall frequency & Presentation quality

The previous studies have shown that QoE of streaming video is influenced by stall frequency [8, 10]. In particular, subjects prefer a single stalling compared to multiple short freezes. The DMOS of test videos with different stall frequencies at three positions are shown in Table 6. Each entry in the the table is a codeword consisting of 3 symbols corresponding to video presentation quality from high to low, as denoted in 3.1. The DMOS values demonstrate that regardless of video presentation quality and position of stalling events, the impact of multiple playback interruption on viewer's QoE is more profound than a single re-buffering.

Table 6: DMOS of the test videos with different stall frequency and presentation quality.

	Single Buff.	Two Buffs.
Begin	9.61 6.06 5.86	12.03 6.92 8.41
Middle	8.62 5.54 6.69	12.07 7.54 7.28
End	8.86 5.56 5.75	11.77 7.42 7.93

Comparing data in the second and the third column of Table 6, we observe drop in the MOS for two buffering is always large than single buffering on average. However, in the following we analyze how statistically significant is this observation.

We use null hypothesis testing to inspect whether multiple re-buffering events are almost always more annoying in compared to single stall. We consider the effect of stalling frequency for all locations. Here, our null hypothesis is as following,

$$H_0 : \mathrm{DMOS}(Q, L, n = 2) = \mathrm{DMOS}(Q, L, n = 1).$$

where Q and L are defined as before, and n is the frequency of buffering ($n = 1$ and $n = 2$ is denoted as single and two buffering, respectively). Results of the Wilcoxon rank-sum test performed on our data is presented in Table 7. The same result holds true for all possible stalling locations, i. e., $L \in \{\mathrm{Begin}, \mathrm{Middle}, \mathrm{End}\}$.

Table 7: Results of the Wilcoxon test performed on DMOS of videos with different stall frequencies.

	Single Buff.	Two Buffs.
Single Buff.	− − −	+ − −
Two Buffs.	+ − −	− − −

Table 7 shows that the Wilcoxon rank-sum test confirms the higher impact of two re-buffering events with respect to single stall only for videos with high presentation quality, and we may not clearly generalize the claim to videos with moderate and low presentation quality. This can be realized by looking at off-diagonal entries in the table.

4. CONCLUSION

To study the joint effect of presentation quality and playback stalling on QoE of streaming videos, as one of initial attempts, we conducted a subjective user study and developed a database of 10-second compressed videos with different stalling patterns. We made several useful observations from the experiment. First, it was found that the impact of stalling location as well as stall frequency on overall QoE is a function of video presentation quality. In particular, the negative impact of stalling is more profound as the level of presentation quality reaches its high and low ends, and thus people are more tolerable to see stalls when videos are of moderate presentation quality. Second, the results of statistical test suggest that to observe recency effect, we have to extend the duration of test sequences beyond 10 to 15 seconds. Third, it is commonly believed that users tolerate stalls with low frequency in general; however, our statistical analysis demonstrates that for videos with moderate and low presentation quality, this hypothesis is challenged.

5. REFERENCES

[1] Cisco Inc, "Visual networking index: Global mobile data traffic fore-cast update, 2012 to 2017," http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.htm, Feb. 2013.

[2] YouTube Statistics, ," http://www.youtube.com/yt/press/statistics.html.

[3] Google Video Quality Report., ," https://www.google.com/get/videoqualityreport/.

[4] The ISP speed Index from Netflix., ," http://ispspeedindex.netflix.com/.

[5] S. S. Krishnan and R. K. Sitaraman, "Video stream quality impacts viewer behavior: Inferring causality using quasi-experimental designs," in *Proceedings of the ACM SIGCOMM Internet Measurement Conference, IMC*, 2012, pp. 211–224, Cited By (since 1996):2.

[6] Conviva's consumer survey report 2015: how consumer judge their viewing experience., ," http://www.conviva.com/csr-2015/.

[7] R. Schatz M. Fiedler K. Masuch T. Hossfeld, S. Egger and C. Lorentzen, "Initial delay vs. interruptions: between the devil and the deep blue sea," in *Fourth Int. Workshop on Qualty of Multimedia Experience (QoMEX)*, 2012, pp. 1–6.

[8] Deepti Ghadiyaram, Alan C Bovik, Hojatollah Yeganeh, Roman Kordasiewicz, and Michael Gallant, "Study of the effects of stalling events on the quality of experience of mobile streaming videos," in *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*. IEEE, 2014, pp. 989–993.

[9] Kai Zeng, Hojatollah Yeganeh, and Zhou Wang, "Quality-of-experience of streaming video: Interactions between presentation quality and playback stalling," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2016, pp. 2405–2409.

[10] Tobias Hofeld, Raimund Schatz, Ernst Biersack, and Louis Plissonneau, "Internet video delivery in youtube: From traffic measurements to quality of experience," in *Data Traffic Monitoring and Analysis*, vol. 7754 of *Lecture Notes in Computer Science*, pp. 264–301. Springer Berlin Heidelberg, 2013.

[11] Hojatollah Yeganeh, Roman Kordasiewicz, Michael Gallant, Deepti Ghadiyaram, and Alan C Bovik, "Delivery quality score model for internet video," in *Image Processing (ICIP), 2014 21th IEEE International Conference on*. IEEE, 2014.

[12] Deepti Ghadiyaram, Janice Pan, and Alan C Bovik, "A time-varying subjective quality model for mobile streaming videos with stalling events," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2015, pp. 959911–959911.

[13] Kalpana Seshadrinathan, Rajiv Soundararajan, Alan Conrad Bovik, and Lawrence K Cormack, "Study of subjective and objective quality assessment of video," *IEEE transactions on image processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[14] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[15] The SSIMplus Index for Video Quality-of Expeirnce Assessment, ," http://ece.uwaterloo.ca/~z70wang/research/ssimplus.

[16] Chao Chen, Lark Kwon Choi, Gustavo de Veciana, Constantine Caramanis, Robert W Heath, and Alan C Bovik, "Modeling the timevarying subjective quality of http video streams with rate adaptations," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2206–2221, 2014.

[17] Joe Yuchieh Lin, Rui Song, Chi-Hao Wu, TsungJung Liu, Haiqiang Wang, and C-C Jay Kuo, "Mcl-v: A streaming video quality assessment database," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 1–9, 2015.

[18] Marie-Neige Garcia, Dominika Dytko, and Alexander Raake, "Quality impact due to initial loading, stalling, and video bitrate in progressive download video services," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. IEEE, 2014, pp. 129–134.

[19] Tobias Hoßfeld, Raimund Schatz, Ernst Biersack, and Louis Plissonneau, "Internet video delivery in youtube: From traffic measurements to quality of experience," in *Data Traffic Monitoring and Analysis*, pp. 264–301. Springer, 2013.

[20] ITU-T, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of internet video and distribution quality television in any environment," Recommendation P.913, International Telecommunication Union, 2016.

[21] R. Schatz-M. Muhlegger K. Masuch B. Gardlo P. Frohlich, S. Egger, "Qoe in 10 seconds: Are short video clip lengths sufficient for quality of experience assessment?," in *Fourth Int. Workshop on Quality of Multimedia Experience (QoMEX)*, 2012, pp. 242–247.

[22] ITU-T, "Methodology for the subjective assessment of the quality of television pictures," Recommendation BT.500, International Telecommunication Union, Geneva, January 2012.

[23] ITU-T, "Subjective video quality assessment methods for multimedia applications," Recommendation P.910, International Telecommunication Union, Geneva, September 1999.

[24] Andrew M Colman, *A dictionary of psychology*, Oxford University Press, USA, 2015.

[25] R Aldridge, J Davidoff, M Ghanbari, D Hands, and D Pearson, "Recency effect in the subjective assessment of digitally-coded television pictures," in *Image Processing and its Applications, 1995., Fifth International Conference on*. IET, 1995, pp. 336–339.

[26] Don E Pearson, "Viewer response to time-varying video quality," in *Photonics West'98 Electronic Imaging*. International Society for Optics and Photonics, 1998, pp. 16–25.

[27] David S Hands and SE Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Applied cognitive psychology*, vol. 15, no. 6, pp. 639–657, 2001.

[28] G. Snedecor and W. Cochran, *Statistical Methods*, Iowa State University Press, 1989.

[29] V Seferidis, M Ghanbari, and DE Pearson, "Forgiveness effect in subjective assessment of packet video," *Electronics Letters*, vol. 28, no. 21, pp. 2013–2014, 1992.