IEEE International Conference on Multimedia and Expo, Seattle, WA, Jul. 2016

QUALITY-OF-EXPERIENCE PREDICTION FOR STREAMING VIDEO

Zhengfang Duanmu, Abdul Rehman, Kai Zeng and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: {zduanmu, abdul.rehman, kzeng, zhou.wang}@uwaterloo.ca

ABSTRACT

With the rapid growth of streaming media applications, there has been a strong demand of objective models that can predict end users' quality-of-experience (QoE) when watching the video being streamed to their display devices. Existing methods typically use bitrate and global statistics of stalling events as the QoE indicators. This is problematic for two reasons. First, using the same bitrate to encode different video content could result in drastically different presentation QoE. Second, the interactions between presentation visual quality and playback stalling are not accounted for. Here we propose a novel QoE prediction approach that takes into consideration the instantaneous quality degradation due to perceptual video presentation impairment, the playback stalling events caused by imperfect network delivery, and the instantaneous interactions between presentation quality and playback stalling. The proposed algorithm demonstrates strong promise when tested using a subject-rated video streaming QoE database.

Index Terms— quality-of-experience, video streaming, instantaneous quality, subjective experiment, adaptive streaming

1. INTRODUCTION

In the past decade, there has been a tremendous growth in streaming media applications, thanks to the fast development of network services and the remarkable growth of smart mobile devices. The total number of unique viewers of VoD in the US is more than 190 million in June 2015 and keeps increasing [1]. Adaptive HTTP streaming protocols such as HTTP Live Streaming (HLS) [2], Silverlight Smooth Streaming (MSS) [3], HTTP Dynamic Streaming (HDS) [4], and Dynamic Adaptive Streaming over HTTP (DASH) [5] achieve decoder-driven rate adaptation by providing multiple video streams of each content in a variety of bitrates and breaking these video streams into small HTTP file segments. The media information of each segment is stored in a manifest file, which is created at the server and transmitted to the client for the players to find the specification and location of each segment. Throughout the streaming process, the player adaptively switches among the available streams by selecting the corresponding segments based on the playback rate, the buffer condition and the instantaneous TCP throughput.

How to deliver videos over the network for optimal QoE of end consumers has been the central goal of modern video delivery services. A survey [6] is carried out to investigate the user preference on the type of video delivery services. Comparison categories consisting of content, timing, quality, ease-of-use, portability, interactivity, and sharing is presented to a group of respondents. Although such subjective user studies provide reliable evaluations, they are inconvenient, time-consuming and expensive. Highly accurate, low complexity *objective* models are desirable to enable efficient design of quality-control protocols for the media delivery systems.

1.1. Related work

Objective video quality assessment (VQA) of static video playback (i.e., with perfect playback smoothness) has been an active research topic in recent years [7][8]. In practice, for the sake of operational convenience, bitrate is often used as the presentation quality indicator. However, using the same bitrate to encode different video content could lead to drastically different visual quality. In addition, different encoders operated at the same bitrate but different operational or complexity modes could also cause large quality variations in the compressed video streams. In order to have a better estimation of video quality, it is necessary to look deep into the pixels of the decoded video frames. For this purpose, the simplest and most widely used VQA measures are the mean squared error (MSE) and peak signal-to-noise ratio (PSNR), which are simple to calculate and mathematically convenient in the context of optimization, but unfortunately are not well matched to perceived visual quality [9]. Perceptually more meaningful VQA models have been drawing significant attention in recent years, exemplified by the success of SSIM [10], MS-SSIM [11], MOVIE [12], VQM [13] and SSIMplus [14]. All these models are only applicable when the playback procedure can be accurately controlled. However, video streaming services, due to network impairments, may suffer from playback stallings that could significantly degrade user experience. Currently, research on QoE modelling for online video streaming is still at an early stage.

Hoßfeld *et al.* [15] made one of the first attempts to quantify QoE based on playback stallings. An exponential re-

lationship is observed between QoE and two global stalling factors: the number and length of stalling events. Oyman *et al.* [16] defined QoE as the probability of playback stalling but did not account for the significant difference between the impact of initial buffering and playback stalling [17][18]. Yeganeh *et al.* [19] modelled the dissatisfaction of playback stalling with a raised cosine function and the recovery of satisfaction level during the playback state with a linear model. Deepti *et al.* [20] employed a Hammerstein-Wiener model by using the stalling length, the total number of stalls, the time since the previous stall, and the inverse stall density as the key features to predict the instantaneous experience at each moment. One common problem of all these approaches is the lack of an effective way to characterize the interactions between video presentation quality and playback smoothness.

Apparently both video presentation quality and playback smoothness play important roles in QoE, but very few works have investigated the connections between them. Garcia *et al.* [21] focused on the progressive download video services and investigated the quality impact due to initial loading, stalling, and compression for high definition sequences. They observed an additive impact of stalling and compression on perceived QoE and reported that the stalling effect is independent of the video content at high bitrate. Ricardo *et al.* [22] approximated the effect of frame drop and image sharpness separately, and took the product of the two terms to predict the overall QoE. Xue *et al.* [23] estimated the packet-level video quality from QP [24] and introduced the concept of intensity of the storyline to weight the impact of stalling. However, neither work provides insights on the interaction between the visual image quality and the stalling events.

1.2. Proposed scheme and major contributions

In this work, we consider the QoE of streaming video as cumulative presentation quality altered by interruptive streaming events; which include initial buffering and playback stallings. The instantaneous quality of the video is captured by advanced VQA models which have been proven to be effective in static video quality prediction. The quality loss due to stalling is modelled with a exponential decaying function, adapted to the user expectation of the video presentation quality.

Our major contributions are twofold. First, we investigate the interactions between video presentation quality and playback smoothness. Our experiment shows that the video presentation quality of the freezing frame correlates with the dissatisfaction level of the stalling event. This is perhaps the first time to explicitly identify the dependence of the two QoE influencing factors. Second, we formulate a joint video streaming QoE model that incorporates both the presentation VQA and the cognitive influence of playback smoothness. The proposed model is not only superior to state-of-the-art models in overall QoE prediction, but also computational efficient. This may help us better understand the perceptual experience of video streaming services in realistic scenarios. The instantaneous QoE prediction has the potentials to be employed in the optimization of adaptive media streaming systems.

2. INTERACTION BETWEEN PLAYBACK SMOOTHNESS AND PRESENTATION QUALITY

2.1. Subjective experiment

Despite the significant amount of effort on the presentation VQA and stalling-centric QoE models, not much has been dedicated to the fundamental relationship between the two QoE factors. Here we briefly present a subjective study we carried out to understand this relationship. Due to space limit and the major focus of the current paper, more details of the subjective study will be reported in other publications.

A video streaming database of 20 pristine high-quality videos of size 1920×1080 are selected to cover diverse content types, including humans, plants, natural scenes, architectures and computer-synthesized sceneries. All videos have the length of 10 seconds. We compressed each of the videos using an x264 encoder at three bitrates (500 kbps, 1500 kbps and 3000 kbps), resulting in 80 videos. We then simulated the initial buffering and stalling events at the beginning and in the middle of each video. All buffering and stalling events last 5 seconds. This results in a total of 240 test videos. A total of 25 naïve subjects, including 13 males and 12 females, participated in the subjective test, after which the mean opinion score (MOS) of each test video (including the videos with and without buffering/stalling) is computed as the average of all subjective scores on the same video. Two outliers are removed based on the outlier removal scheme suggested in [25]

State-of-the-art VQA models were tested on the videos without buffering/stalling and SSIMplus [14] turns out to have the highest correlation with MOS, suggesting that SSIMplus provides a reasonable prediction on the video presentation quality. Another significant advantage of SSIMplus is that it provides device and viewing condition dependent quality evaluations, and thus the same video viewed on a different device (e.g., a TV versus a smart phone) would be given different SSIMplus scores, a desirable feature that is lacking in other VQA approaches.

2.2. Analysis

The major purpose of this subjective experiment is to gain insights about whether the stalling events are independent of the video presentation quality. If the answer is yes, then regardless of the presentation quality, stallings will have the same impact on the overall QoE scores. Assuming an additive relationship between stalling and presentation quality as in [21], we are expecting a near constant quality drop across different compression levels when a stalling event occurs in the middle of the sequences.

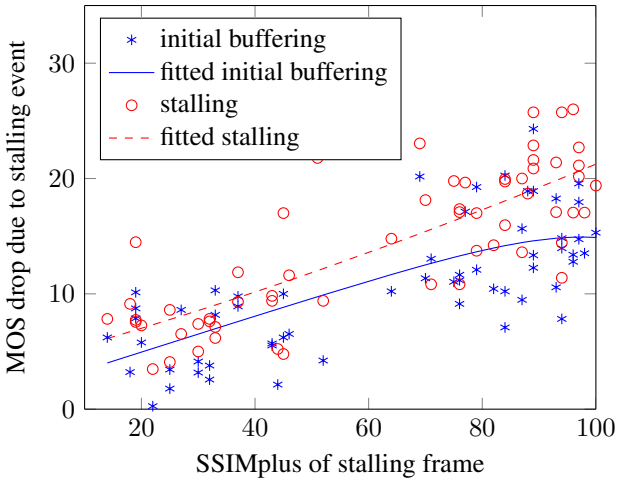


Fig. 1: Presentation video quality vs. penalty of stalling.

Fig. 1 shows a scatter plot of the instantaneous quality of the freezing frame predicted by SSIMplus and the MOS degradation for both initial delay and playback stalling. It can be observed that for the stalling at the same temporal instance and of the same duration, human subjects tend to give a higher penalty to the video with a higher instantaneous presentation quality at the freezing frame. One explanation may be that there is a higher viewer expectation when the video presentation quality is high, and thus the interruption caused by stalling make them feel more frustrated.

3. QOE MODEL

Motivated by the observation and analysis above, we develop a QoE prediction model by incorporating the video presentation quality and the impact of buffering/stalling events.

3.1. Video Presentation Quality

For each frame in the streaming video, its instantaneous video presentation quality P_n can be estimated at the server side by a frame-level VQA model before transmission

$$P_n = V(X_n, R_n), \tag{1}$$

where X_n and R_n are the n-th frame of the streaming video and pristine quality video, and $V(\cdot)$ is a full reference VQA operator. The computed quality score $V(X_n, R_n)$ can either be embedded into the manifest file that describes the specifications of the video, or carried in the metadata of the video container. The manifest is transmitted to the client side such that its information is available to the client. In commonly used streaming protocols such as MPEG-DASH, the partially decoded frame will not be sent for rendering, and thus viewers will see the last successfully decoded frame during the stalling interval. Thus, for a stalling moment n in the interruption period $[i, j]$, the video presentation quality at the instance, P_n, is the same as the quality of the last decoded frame

$$P_n = P_{i-1}, \quad \text{for } n = i, i+1, ..., j. \tag{2}$$

3.2. Stalling Experience Quantification

To simplify the formulation, we assume the influence of each stalling event is independent and additive. As such, we can analyze each stalling event separately and compute the overall effect by aggregating them. Note that each stalling event divides the streaming session time line into three non-overlapping intervals, i.e., the time intervals before the stalling, during the stalling, and after the stalling. We will discuss the three intervals separately because the impact of the stalling event on each of the intervals are different.

First, we assign zero penalty to the frames before the stalling occurs when people have not experienced any interruption. Second, as a playback stalling starts, the level of dissatisfaction increases as the stalling goes on till playback resumes. The exponential decay function has been successfully used in previous studies [26][15]. The use of exponential decay assumes an existence of QoE loss saturation to the number and length of stalling, and low tolerance to jitters comparing to the other commonly used utility function such as logarithm and sigmoid. Here we approximate the QoE loss due to a stalling event with an exponential decay function similar to [15]. Third, QoE also depends on a behavioural hysteresis "after effect" [27]. In particular, a previous unpleasant viewing experience caused by a stalling event tends to penalize the QoE in the future and thus affects the overall QoE. The extent of dissatisfaction starts to fade out at the moment of playback recovery because observers start to forget the annoyance. To model the decline of memory retention of the buffering event, we employ the Hermann Ebbinghaus forgetting curve [28]

$$M = \exp\left\{-\frac{t}{T}\right\}, \tag{3}$$

where M is the memory retention, T is the relative strength of memory, and t is the time instance.

Assume that the k-th stalling event locates at $[i_k, i_k + l_k]$, where l_k is the length of stall, a piecewise model is constructed to estimate the impact of each stalling event on the QoE

$$S^k(t) = \begin{cases} P_{i_k-1}\left(-1 + \exp\left\{-\left(\frac{tf - i_k}{T_0}\right)\right\}\right) & \frac{i_k}{f} \le t \le \frac{i_k+l_k}{f} \\ P_{i_k-1}\left(-1 + \exp\left\{-\left(\frac{l_k}{T_0}\right)\right\}\right) & \\ \cdot\left(\exp\left\{-\left(\frac{tf - i_k - l_k}{T_1}\right)\right\}\right) & t > \frac{i_k+l_k}{f} \\ 0 & otherwise \end{cases} \tag{4}$$

where f is the frame rate in frames/second, and T_0, T_1 and $S^k(t)$ represent the rate of dissatisfaction, the relative strength of memory and the experience of the k-th stalling event at

time t, respectively. P_{i_k-1}, the scaling coefficient of the decay function, has two functions: 1) it reflects the viewer expectation to the future video presentation quality, and 2) it normalizes the stalling effect to the same scale of VQA kernel. This formulation is qualitatively consistent with the relationship between the two QoE factors discussed in the previous section. In addition, since the impact of initial buffering and stalling are different, we have two sets of parameters: $\{T_0^{init}, T_1^{init}\}$ for initial delay and $\{T_0, T_1\}$ for other playback stallings, respectively. We also assume that the initial expectation P_0 is a constant. In this way, the initial buffering time is proportional to the cumulated experience loss.

The instant QoE drop due to stalling events is computed by aggregating the QoE drop caused by each stalling event and is given by

$$S(t) = \sum_{k=1}^{N} S^k(t), \tag{5}$$

where N is the total number of stalling events.

3.3. Overall QoE

The instantaneous QoE at each time unit n in the streaming session can be represented as the aggregation of the two channels

$$Q_n = P_n + S_n. \tag{6}$$

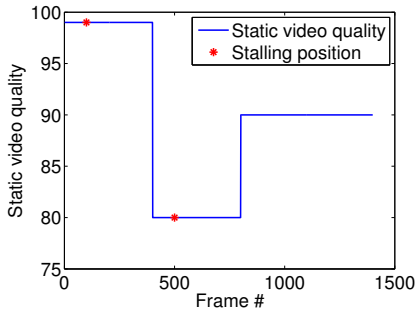
In practice, one usually requires a single end-of-process QoE measure. We use the mean value of the predicted QoE over the whole playback duration to evaluate the overall QoE. To reduce the memory usage, the end-of-process QoE can be computed in a moving average fashion

$$A_n = \frac{(n-1)A_{n-1} + Q_n}{n}, \tag{7}$$

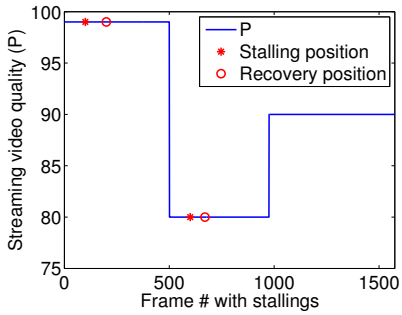
where A_n is the cumulative QoE up to the n-th time instance in the streaming session. An example of each channel and the final output of the model is illustrated in Fig. 2.

4. TEST

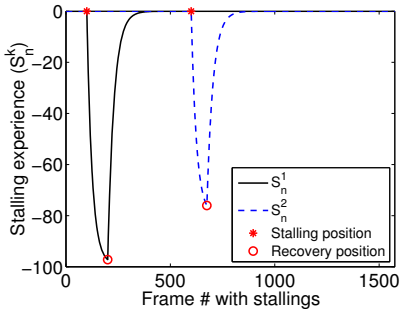
To demonstrate the effectiveness of our framework, four VQA algorithms, namely PSNR, SSIM [10], MSSSIM [11] and SSIMplus [14], are employed as the frame-level video presentation quality measure. Throughout the paper, the proposed model uses the following parameter settings: $T_0^{init} = 2$, $T_1^{init} = 0.5$, $T_0 = 1$, $T_1 = 1.2$ and $P_0 = 0.8 \cdot |(V(\cdot)|$, where $|V(\cdot)|$ is the range of adopted VQA kernel. Then Spearman rank order correlation coefficients (SRCC) and Pearson's linear correlation coefficients (PLCC) between the predicted and the ground truth MOS were computed to assess the prediction monotonicity and prediction accuracy. All presentation VQA measures mentioned above without incorporating the proposed method are also included in the comparison as the



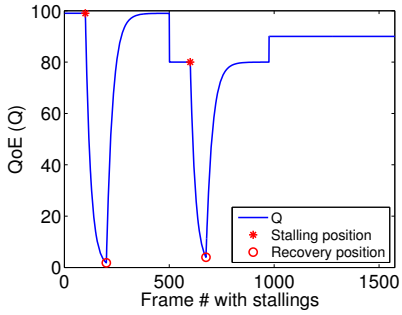
(a)



(b)



(c)



(d)

Fig. 2: An illustrative example of and channel responses at each frame. (a) video presentation quality of the static video at each frame. '*' indicates the position of stalling. (b) video presentation quality of the streaming video during playback at each frame. '*' indicates the position of stalling and 'o' indicates the position of recovery. (c) QoE drop due to each stalling events at each frame. The solid curve shows the QoE drop due to initial buffering and the dashed curve shows the QoE drop due to playback stalling. (d) Overall QoE at each time instance during playback.

Table 1: SRCC and PLCC performance comparison of QoE models.

	FTW [15]	PSNR	PSNR+proposed	SSIM [10]	SSIM+proposed	MS-SSIM [11]	MS-SSIM+proposed	SSIMplus [14]	SSIMplus+proposed
SRCC	0.3154	0.6715	0.7492	0.8177	**0.9009**	0.7928	0.8807	0.8024	0.9007
PLCC	0.3133	0.6663	0.7391	0.8432	0.9015	0.8193	0.8776	0.8350	**0.9026**

benchmark. In all of our tests, all video sequences were employed (i.e., both videos with and without stalling events are included).

Fig. 3 shows the scatter plots of the MOS prediction results for each presentation VQA quality with (shown in the first row) and without (shown in the second row) inorporating the proposed method. The corresponding SRCC and PLCC results are given in Table 1. We have three observations here. First, the proposed model significantly outperforms its baseline presentation VQA model. Second, a higher compactness in the scatter plots is achieved by applying the proposed model because of the proper penalties given to the videos with stalling. Finally, the best performance is obtained by combining the proposed method with SSIMplus [14] VQA model.

In Table 1, we have also included the re-buffering based FTW model for comparison. Apparently, the proposed scheme outperforms the FTW model, which does not use a presentation quality measure and does not properly account for the relationship between presentation quality and playback stallings.

5. CONCLUSIONS AND FUTURE WORK

We have presented a subjective study to understand human visual QoE of streaming video and proposed an objective model to characterize the perceptual QoE. Our work represents one of the first attempts to bridge the gap between the presentation VQA and stalling-centric models in QoE prediction. The subjective experiment reveals some interesting relationship between the impact of stalling and the instantaneous presentation quality. The experiments also demonstrate that the proposed model is simple in expression and effective in performance.

Future research may be extended in many directions. First, a comprehensive subject-rated database that consists of more stalling patterns and video quality variations is desired to better understand the behaviours of human viewers and to examine the performance of existing objective QoE methods. Second, how to quantify the influence of the semantics of stalling position, and how to incorporate it into QoE models should be studied. Third, how to quantify the quality switching experience in adaptive video streaming needs to be exploited.

6. REFERENCES

[1] comScore, "comscore releases june 2015 u.s. desktop online video rankings," July 2015. [Online]. Available: https://www.comscore.com/Insights/Market-Rankings.

[2] Apple Inc., "HTTP Live Streaming Technical Overview 2013." [Online]. Available: https://developer.apple.com/library/ios/documentation/ networkinginternet/conceptual/streamingmediaguide.

[3] A. Zambelli, "Smooth streaming technical overview." [Online]. Available: http://www.iis.net/learn/media/on-demand-smooth-streaming.

[4] Adobe Systems Inc., "HTTP Dynamic Streaming 2013." [Online]. Available: http://www.adobe.com/products/hds-dynamic-streaming.html.

[5] DASH Industry Forum, "For Promotion of MPEG-DASH 2013." [Online]. Available: http://dashif.org.

[6] Cisco Inc., "survey: Cisco ibsg youth survey," Cisco IBSG Youth Focus Group Sessions, 2010.

[7] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The handbook of video databases: design and applications*, Sept. 2003.

[8] Z. Wang and A. C. Bovik, "Modern image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing*, vol. 2, no. 1, pp. 1–156, 2006.

[9] ——, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[10] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. on Signals, Systems, and Computers*, 2003.

[12] K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Processing*, Feb 2010.

[13] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting*, vol. 50, no. 3, pp. 312–322, Sept 2004.

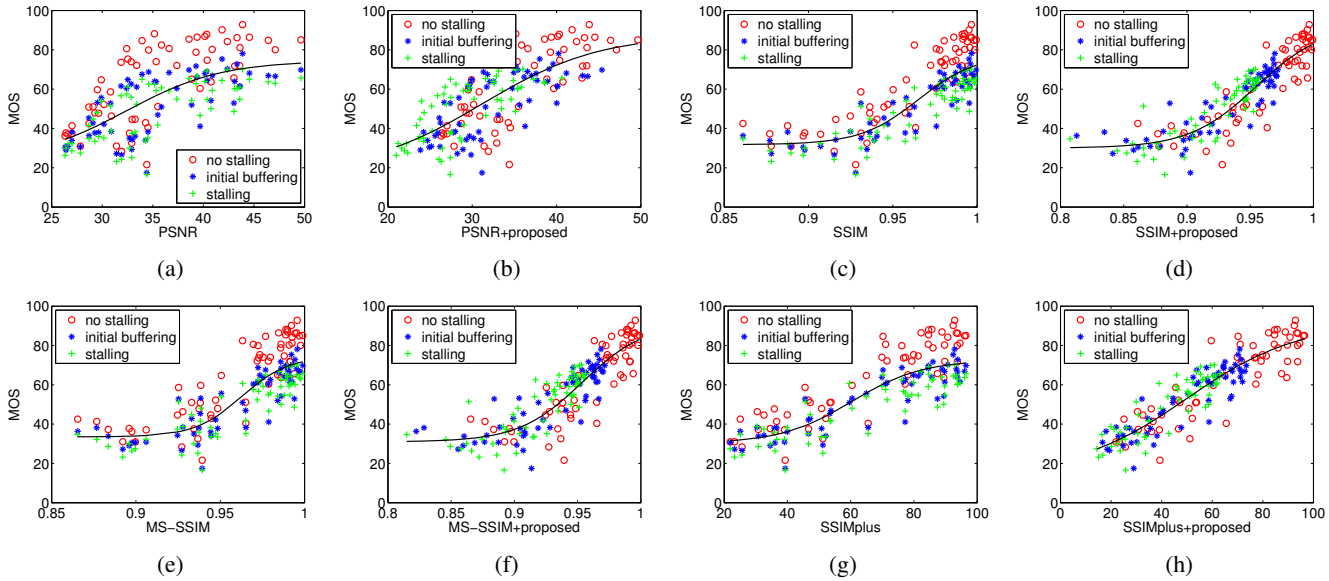


Fig. 3: Predicted QoE vs. MOS.

[14] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video quality-of-experience assessment," in *Proc. SPIE*, vol. 9394, 2015, pp. 939 406–939 406–11.

[15] T. Hoßfeld, M. Seufert, M. Hirth, T. Zinner, P. Tran-Gia, and R. Schatz, "Quantification of YouTube QoE via crowdsourcing," in *Proc. IEEE Int. Sym. Multimedia*, Dec 2011, pp. 494–499.

[16] O. Oyman and S. Singh, "Quality of experience for HTTP adaptive streaming services," *IEEE Comm. Magazine*, vol. 50, no. 4, pp. 20–27, April 2012.

[17] T. Hoßfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: between the devil and the deep blue sea," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2012, pp. 1–6.

[18] A. Sackl, S. Egger, and R. Schatz, "Where's the music? comparing the QoE impact of temporal impairments between music and video streaming," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2013, pp. 64–69.

[19] H. Yeganeh, R. Kordasiewicz, M. Gallant, D. Ghadiyaram, and A. Bovik, "Delivery quality score model for Internet video," in *Proc. IEEE Int. Conf. Image Proc.*, Oct 2014, pp. 2007–2011.

[20] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A time-varying subjective quality model for mobile streaming videos with stalling events," in *Proc. SPIE*, 2015.

[21] M. Garcia, D. Dytko, and A. Raake, "Quality impact due to initial loading, stalling, and video bitrate in progressive download video services," in *Proc. IEEE Int. Conf. Multimedia and Expo*, 2014, pp. 129–134.

[22] R. R. Pastrana-Vidal and J.-C. Gicquel, "A no-reference video quality metric based on a human assessment model," in *Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2007.

[23] J. Xue, D.-Q. Zhang, H. Yu, and C. W. Chen, "Assessing quality of experience for adaptive HTTP video streaming," in *Proc. IEEE Int. Conf. Multimedia and Expo*, July 2014, pp. 1–6.

[24] Y. Liu, Z. G. Li, and Y. C. Soh, "Region-of-interest based resource allocation for conversational video communication of H. 264/AVC," *IEEE Trans. Circuits and Systems for Video Tech.*, 2008.

[25] ITU-R BT.500-12, "Recommendation: Methodology for the subjective assessment of the quality of television pictures," Nov. 1993.

[26] M. Fiedler, T. Hoßfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, no. 2, pp. 36–41, March 2010.

[27] K. Seshadrinathan and A. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, May 2011, pp. 1153–1156.

[28] H. Ebbinghaus, *Memory: A contribution to experimental psychology*. Teachers college, Columbia university, 1913.