A Quality-of-Experience Index for Streaming Video

Zhengfang Duanmu, *Student Member, IEEE*, Kai Zeng, Kede Ma, *Student Member, IEEE*, Abdul Rehman, and Zhou Wang, *Fellow, IEEE*

***Abstract*—With the rapid growth of streaming media applications, there has been a strong demand of quality-of-experience (QoE) measurement and QoE-driven video delivery technologies. Most existing methods rely on bitrate and global statistics of stalling events for QoE prediction. This is problematic for two reasons. First, using the same bitrate to encode different video content results in drastically different presentation quality. Second, the interactions between video presentation quality and playback stalling experiences are not accounted for. In this work, we first build a streaming video database and carry out a subjective user study to investigate the human responses to the combined effect of video compression, initial buffering, and stalling. We then propose a novel QoE prediction approach named Streaming QoE Index that accounts for the instantaneous quality degradation due to perceptual video presentation impairment, the playback stalling events, and the instantaneous interactions between them. Experimental results show that the proposed model is in close agreement with subjective opinions and significantly outperforms existing QoE models. The proposed model provides a highly effective and efficient meanings for QoE prediction in video streaming services.[1]**

***Index Terms*—Adaptive bitrate streaming, quality-of-experience, objective quality assessment, subjective quality assessment, streaming video, video stalling.**

I. 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. HTTP Live Streaming (HLS) [1], Silverlight Smooth Streaming (MSS) [2], HTTP Dynamic Streaming (HDS) [3], and Dynamic Adaptive Streaming over HTTP (DASH) [4] achieve decoder-driven rate adaptation by providing video streams in a variety of bitrates and breaking them into small HTTP file segments. The media information of each segment is stored in a manifest file, which is created at server and transmitted to client to provide the specification and location of each segment. Throughout the streaming process, the video player at the client adaptively switches among the available streams by selecting segments based on playback rate, buffer condition and instantaneous TCP throughput [5].

Due to the increasing popularity of video streaming services, users are continuously raising their expectations on better services. A recent survey [6] carried out to investigate the user preference on the type of video delivery services shows a dominating role of QoE in the user choice over the other categories such as content, timing, quality, ease-of-use, portability, interactivity, and sharing. Another study [7] shows that global premium content delivery networks lost \$2.16 billion of revenue due to poor quality video streams in 2012 and are expected to miss out on an astounding \$20 billion by 2017. The poor streaming experience has became a major threat to the video service ecosystem. Therefore, achieving optimal QoE of end viewers has been the central goal of modern video delivery services.

QoE for HTTP Adaptive Streaming (HAS) has been a rapidly evolving research topic and has attracted an increasing amount of attention from both industry and academia. As the humans are the ultimate receiver of videos in most applications, subjective evaluation is the most straightforward and reliable approach to evaluate the QoE of streaming videos. A comprehensive subjective user study has several benefits. First, it provides useful data to study human behaviors in evaluating perceived quality of streaming videos. Second, it supplies a test set to evaluate, compare and optimize streaming strategies. Third, it is useful to validate and compare the performance of existing objective QoE models. Although such subjective user studies provide reliable evaluations, they are inconvenient, time-consuming and expensive. Most importantly, they are not applicable in the real-time playback scheduling framework. Therefore, highly accurate, low complexity objective models are desirable to enable efficient design of quality-control and resource allocation protocols for media delivery systems. Over the past decade, substantial effort has been made to develop objective QoE models [8]–[22]. Most of them are designed for specific applications such as static video quality assessment or progressive video streaming. Furthermore, little work has been done to compare them with subjective data comprising a wide variety of video sequences.

In this work, we aim to design an objective QoE model that accounts for both the presentation quality variations and the impact of stalling experience in streaming videos. Our major contributions are threefold. First, we construct a video database dedicated to the combined effect of initial buffering, stalling and video compression on QoE, which is one of the first publicly available databases of its kind. Second, we investigate the interactions between video presentation quality and playback stalling. Our experiments show that the video presentation quality of the freezing frame exhibits interesting relationship, which has not been observed before, with the dissatisfaction level of the stalling event. Third, we formulate a joint video streaming QoE model that incorporates both the video presentation quality

Manuscript received March 8, 2016; revised July 21, 2016; accepted August 29, 2016. Date of publication September 12, 2016; date of current version January 13, 2017. The guest editor coordinating the review of this paper and approving it for publication was Dr. L. Skorin-Kapov

The authors are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1 Canada (e-mail: zduanmu@uwaterloo.ca; kzeng@uwaterloo.ca; k29ma@uwaterloo.ca; abdul.rehman@uwaterloo.ca; zhou.wang@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2016.2608329

[1]The subjective database is available online at https://ece.uwaterloo.ca/%7Ezduanmu/jstsp16qoe/. Preliminary results of Section III were submitted to the 23rd International Conference on Image Processing, USA, 2016.

and the influence of playback stalling. Experiments on the benchmark database show that the proposed model significantly outperforms existing QoE models. The instantaneous QoE prediction is ideal for the optimization of media streaming systems.

II. Related Work

A. Subjective QoE Studies

A significant number of subjective QoE studies have been conducted to understand the perceptual impact of different types of impairments on HAS. Two excellent surveys on QoE subjective study can be found in [23] and [24]. Here we only provide a brief overview: Pastrana *et al.* [25] made one of the first attempts to measure the impact of stalling in video streaming services. The study showed that QoE is influenced by both the *duration* and the *frequency* of stalling events and was confirmed by Qi *et al.* [26] and Moorthy *et al.* [27]. Among those findings, the most important one is that viewers tend to prefer videos that have less number of freeze events (even if they are relative longer) to videos that have a sequence of short freezes through time. Besides, Qi *et al.* [26] also found that a stalling of frame-level duration could not be perceived, and thus has no impact on QoE. Staelens *et al.* [28] extended Qi's research and conclude that isolated stallings up to approximately 400 ms is acceptable to the end-users. Moorthy *et al.* [27] investigated the trade-off between stalling and quality switching. While many studies [29], [30] assumed that stalling events are more annoying than quality switches, the results in [27] showed that few stalling events are not yielding worse quality than downward quality switches. Hoßfeld *et al.* [31] and Sackl *et al.* [32] found fundamental differences between initial delays and stalling. Unlike initial delay which is somewhat expected by today's consumers, stalling invokes a sudden unexpected interruption and distort the temporal video structure. Hence, stalling is processed differently by the human sensory system, *i.e.*, it is perceived much worse [33]. Garcia *et al.* [34] investigated the quality impact of the combined effect of initial loading, stalling, and compression for high definition sequences, from which they observed an additive impact of stalling and compression on perceived QoE. Besides the effect of video impairment itself, Seshadrinathan *et al.* [35] described a hysteresis effect in a recent study of time-varying video quality. In particular, an unpleasant viewing experience in the past tends to penalize the QoE in the future and affect the overall QoE.

Based on these subjective user studies, one may conclude that: 1) video presentation quality, duration and frequency of stalling are the key factors contributing towards the overall QoE; 2) Although very short stalling may not be perceived and thus has little impact on QoE, visible stalling events can severely degrade QoE; 3) Viewers are much more tolerant to initial buffering than stalling; 4) An unpleasant viewing experience in the past tends to penalize future QoE.

B. Existing Objective QoE Models

The existing QoE models can be roughly categorized as follows:

1) Signal Fidelity Measurement
Objective VQA approaches tackle the QoE problem from a signal fidelity point of view to provide computational models that can automatically predict video presentation quality. In practice, for the sake of operational convenience, bitrate and Quantization Parameter (QP) are often used as the indicators of video presentation quality [1]–[4]. However, using the same bitrate or QP to encode different video content can lead to drastically different visual quality. In addition, different encoders operate at the same bitrate or QP but different operational or complexity modes can also cause large quality variations in the compressed video streams. In order to have a better estimation of the user perceived QoE, it is desired to assess the raw video. 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 easy to calculate and mathematically convenient, but unfortunately do not correlate well with perceived visual quality [36]. Research in perceptual VQA [37], [38] has been drawing significant attention in recent years, exemplified by the success of the structural similarity index (SSIM) [8], the multi-scale structural similarity index (MS-SSIM) [9], motion-based video integrity evaluation index (MOVIE) [10], video quality metric (VQM) [11] and SSIMplus [12]. State-of-the-art VQA models employ human visual system features in quality assessment, and thus provide perceptually more meaningful prediction. Nevertheless, all of 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 issues that could significantly degrade user QoE. How modern VQA models can be used in the context of HAS is still an open problem.
2) QoE Prediction via Quality-of-Service (QoS)
The philosophy behind this type of approach is that there exists an causal relationship between generic QoS problems (e.g, loss, delay, jitter, reordering and throughput limitations) and generic QoE problems (e.g., glitches, artifacts and excessive waiting time) [39]. Therefore, QoE can be easily quantified once the mapping function between QoS and QoE is known. Thanks to the reliability of TCP, HAS is immune to glitches and artifacts introduced by packet drop. Thus, most existing research in this direction are dedicated to $stalling\,experience\,quantification$. Watanabe *et al.* [13] attempt to quantify streaming video QoE based on playback stallings. They observed a logarithmic relationship between the global length of stalling events and QoE. Mok *et al.* [40] associated the length and frequency of stalling to QoE with a linear function. Hoßfeld *et al.* [14], [15], [39] demonstrated the superiority of exponential mapping functions in many streaming applications. Although the global QoS statistics-based QoE models are computationally efficient, they ignore the importance of temporal factors. Rodriguez *et al.* [17] consider the pattern of jitter and local content importance

by subjective training of the content. Yeganeh *et al.* [18] quantify the stalling experience with a raised cosine function and the recovery of satisfaction level during the playback state with a linear model. Deepti *et al.* [19] employ a Hammerstein-Wiener model using the stalling length, the total number of stalling events, the time since the previous stall, and the inverse stalling density as the key features to predict the instantaneous experience at each moment. The stalling experience quantification approach is only adequate in the progressive download services because it is unable to measure the experience loss of video quality. However, in HAS, a source video is always encoded into multiple representations, which have different presentation quality.

3) Hybrid Approach
Apparently both video presentation quality and stalling experience quantification capture important aspects in QoE. Unfortunately, very few approaches incorporate the two aspects into a unified model. 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. Singh *et al.* [20] tried to solve this problem by training a random neural network [41] using QP, frequency, average and maximum duration of stalling events as input features. Xue *et al.* [21] estimated the video presentation quality by QP and weighted the impact of stalling by packet bit count as an indicator of motion complexity. Both algorithms define video presentation quality as a function of QP, which has been proven to be a poor perceptual quality indicator.

Despite the demonstrated success, most existing QoE predictors either underestimate the effect of perceptual video presentation quality or simply equate it to bitrate or QP. More importantly, one common assumption of all these approaches is that there is no interaction between video presentation quality and stalling experience, which has not been systematically examined.

III. Subjective Quality-of-Experience User Study of Streaming Videos

To the best of our knowledge, current publicly available databases are dedicated to either video presentation quality that is affected by compression, channel transmission losses, scaling, or the impact of stalling in terms of its occurring position, duration, and frequency. However, QoE of streaming video should be a joint effect of the video presentation quality and playback stalling. Although the combined effect of stalling and video bitrate has been investigated by Garcia *et al.* [34], the study suffers from the following problems: (1) the dataset is of insufficient size (6 source sequences); (2) bitrate is not a good indicator of video presentation quality as discussed in the Section II-B; and (3) the database is not publicly available. Therefore, our goal is to develop a dedicated database to study the interaction between stalling effect and presentation quality for video streaming.

TABLE I
Information of Reference Videos

Index	Name	Frame Rate	Description
a	Animation	25	animation, high motion
b	Biking	50	human, outdoor
c	BirdsOfPrey	30	natural, static
d	ButterFly	25	natural, outdoor
e	CloudSea1	24	architecture, static
f	CloudSea2	24	outdoor, high motion
g	CostaRica1	25	natural, static
h	CostaRica2	25	natural, static
i	Football1	25	human, high motion
j	Football2	25	human, high motion
k	Football3	25	human, high motion
l	Forest1	25	natural, static
m	Forest2	25	natural, outdoor
n	MTV	25	human, indoor
o	Ski	30	outdoor, high motion
p	Squirrel	25	animation, outdoor
q	Transformer1	24	human, static
r	Transformer2	24	human, architecture
s	Basketball1	25	human, high motion
t	Basketball2	25	human, high motion

A. Video Database and Subjective User Study

A video database, named streaming video QoE database, of 20 pristine high-quality videos of size 1920×1080 are selected to cover diverse content, including humans, plants, natural scenes, architectures and computer-synthesized sceneries. All videos have the length of 10 s [42]. The detailed specifications of those videos are listed in Table I and a screenshot from each video is included in Fig. 1. Using aforementioned sequences as the source, each video is encoded into three bitrate levels (500 Kbps, 1500 Kbps, 3000 Kbps) with x264 encoder to cover different quality levels. The choices of bitrate levels are based on commonly-used parameters for transmission of HD videos over networks. A 5-s stalling event is simulated at either the beginning or the middle point of the encoded sequences. The stalling indicator was implemented as a spinning wheel. In total, we obtain 200 test samples that include 20 source videos, 60 compressed videos, 60 initial buffering videos, and 60 mid-stalling videos.

The subjective testing experiment is setup as a normal indoor home settings with ordinary illumination level, with no reflecting ceiling walls and floors. All videos are displayed at their actual pixel resolution on an LCD monitor at a resolution of 2560×1600 pixel with Truecolor (32 bit) at 60 Hz. The monitor is calibrated in accordance with the recommendations of ITU-T BT.500 [43]. A customized graphical user interface is used to render the videos on the screen with random order and to record the individual subject ratings on the database. The study adopts a single-stimulus quality scoring strategy. A total of 25 naïve subjects, including 13 males and 12 females aged between 22 and 30, participate in the subjective test. Visual acuity (*i.e.*, Snellen test) and color vision (*i.e.*, Ishihara) are confirmed from each subject before the subjective test. A training session was performed before the data collection, during which, 4 videos (of 1. pristine quality video, 2. 500 Kbps encoded video, 3. video with

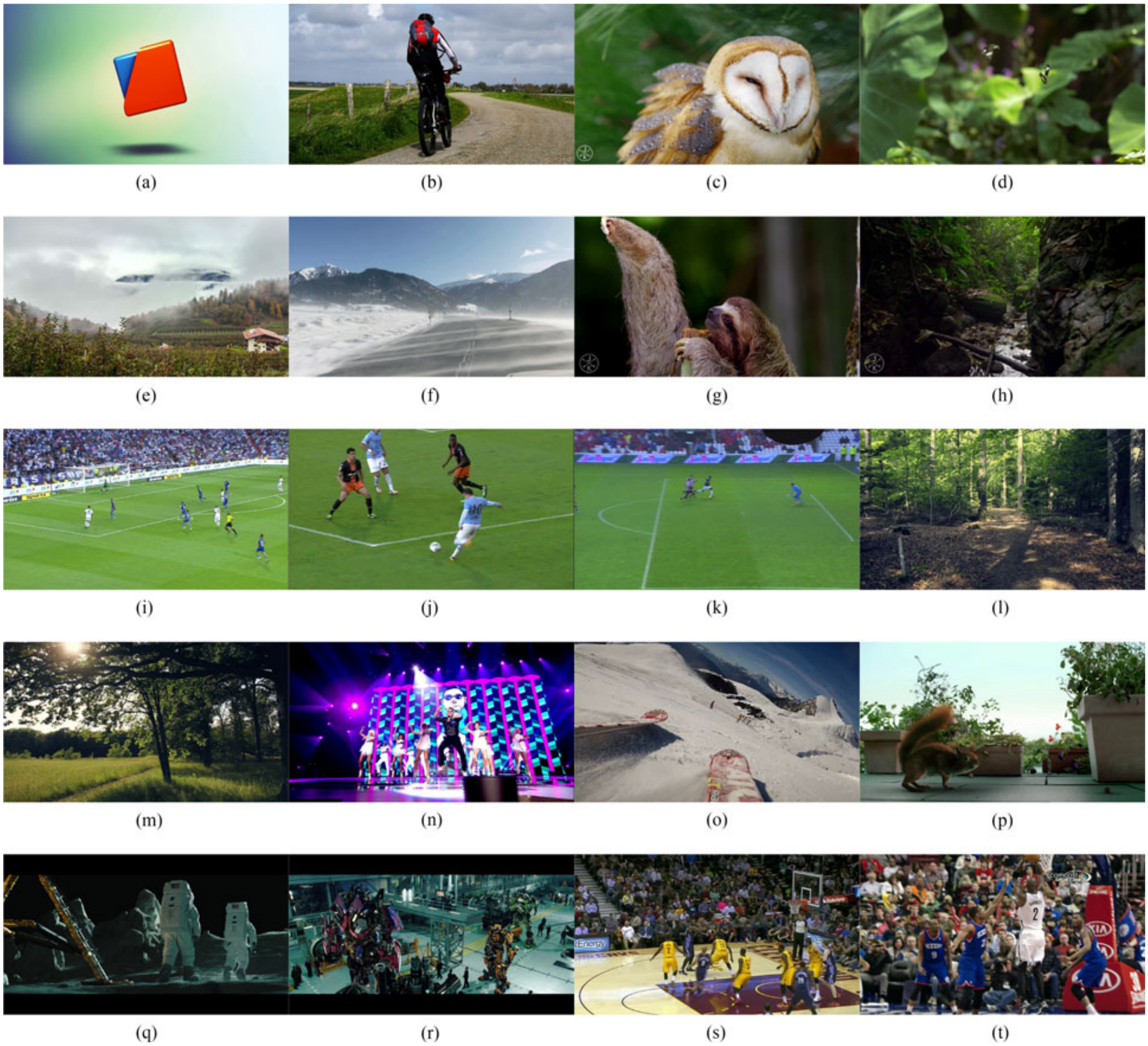


Fig. 1. Subjective test sequences.

initial buffering, and 4. video with stalling) were presented to the subjects. We used the same methods to generate the videos used in the training and testing sessions. Therefore, subjects knew what distortion types would be expected before the test session, and thus learning effects are kept minimal in the subjective experiment. Subjects were instructed with sample videos to judge the overall visual quality considering both picture distortion artifacts and video freezes as quality degradation factors. The subjects are allowed to move their positions to get closer or farther away from the screen for better observation. For each subject, the whole study takes about one and half hour, which is divided into three sessions with two 7-min breaks in-between. In order to minimize the influence of fatigue effect, the length of a session was limited to 25 min. The choice of a 100-point continuous scale as opposed to a discrete 5-point ITU-R Absolute Category Scale (ACR) has advantages: expanded range, finer distinctions between ratings, and demonstrated prior efficacy [44].

A common dilemma in every subjective video quality experiment is how much instruction should be given to the subjects. In practice, humans are often attracted by video content rather than quality variations. But to collect quality scores, certain instruction has to be given to the subjects in order to obtain their opinions on video quality. On the other hand, if too much instruction is given, the subjects may be over-educated to give "clean" but unrealistic scores. In our study, to give uniform instruction to all subjects, and to investigate the interactions between presentation quality and delay/stalling, we find it necessary to inform the subjects about what types of quality degradations they should expect to see. Other than that, no further specifications are given.

Since the break between successive test sessions is considerably short, alignment on the subjective scores is not performed. In other words, raw subjective scores are used in the subsequent analysis. After the subjective user study, two outliers are removed based on the outlier removal scheme suggested in [43]. The final quality score for each individual image is computed as the average of subjective scores, namely the mean opinion score (MOS), from all valid subjects. Considering the MOS as the "ground truth", the performance of individual subjects can

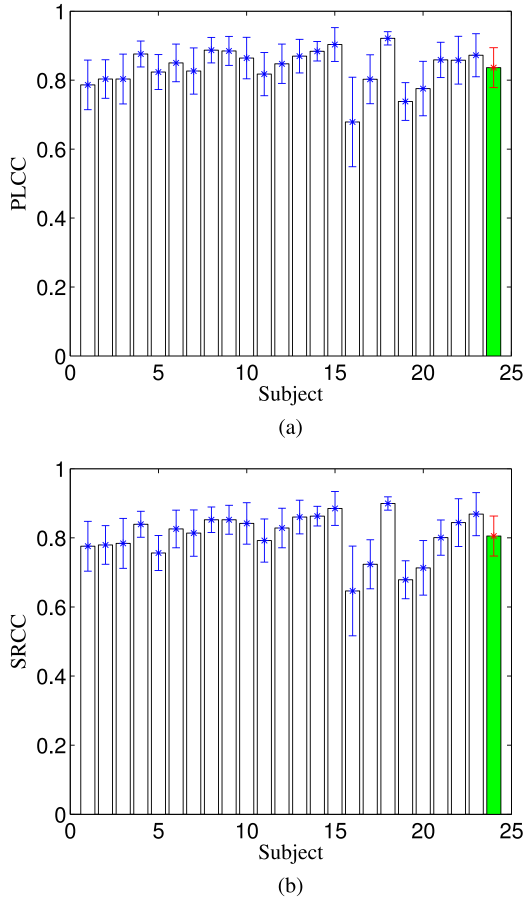


Fig. 2. PLCC and SRCC between individual subject rating and MOS. Rightmost column: Performance of an average subject.

be evaluated by calculating the correlation coefficient between individual subject ratings and MOS values for each image set, and then averaging the correlation coefficients of all image sets. The Pearson linear correlation coefficient (PLCC) and Spearman's rand-order correlation coefficient (SRCC) are employed as comparison criteria, whose range is from 0 to 1 and higher values indicate better performance. They can be computed for each subject and their values for all subject are depicted in Fig. 2. It can be seen that each individual subject performs well in terms of predicting MOSs. The average performance across all individual subjects is also given in the rightmost column in Fig. 2. This provides a general idea about the performance of an average subject. Therefore, we conclude that considerable agreement is observed among different subjects on the perceived quality of the test video sequences.

B. Subjective Data Analysis

One of the main objectives of this subjective experiment is to investigate whether the impact of the stalling events are independent of the video presentation quality. If the answer is yes, then regardless of the video presentation quality, stallings will have the same impact on the overall QoE scores. Assuming an additive relationship between stalling and video presentation quality as in [34], we are expecting a near constant MOS drop

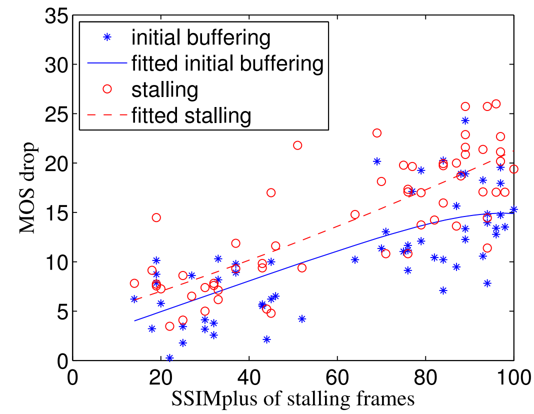


Fig. 3. SSIMplus of stalling frames versus MOS drop.

across different video presentation quality when a stalling event occurs in the middle of the sequences.

Fig. 3 shows a scatter plot of the instantaneous quality of the freezing frame predicted by SSIMplus [12] 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 video presentation quality at the freezing frame. We further performed a statistical significance test as follows. Denoting the SSIMplus score of the initial buffered/stalling frame, and the MOS drop of the test video with initial buffering/stalling as random variables X_1/X_2 and Y_1/Y_2, we specify the null hypotheses H_1/H_2 as that X_1/X_2 is uncorrelated with Y_1/Y_2. The test statistic is $t = \frac{r\sqrt{N-2}}{1-r^2}$, where r and N are the correlation coefficient and the number of samples, respectively. The resulting test statistic is used to compute the P-values by referring to a t-distribution with $N-2$ degrees of freedom. Since the P-values (6.32×10^{-8} for initial buffering and 6.87×10^{-13} for stalling) are much smaller than the significance level 0.05, we reject the null hypotheses in favor of the alternatives. The results suggest that there is sufficient evidence at the 0.05 significance level to conclude that there is a linear relationship in the population between the SSIMplus score (estimation of the presentation quality) of the initial buffered/stalling frame and the QoE drop. This phenomenon was not observed in previous studies. 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.

C. Performance of Existing Objective QoE Models

Using the above database, we test the performance of four existing VQA models, including PSNR, SSIM [8], MS-SSIM [9] and SSIMplus [12] and four state-of-the-art QoE models [15], [17], [21], [40]. The implementations for the VQA models are obtained from the original authors and we implement four QoE models since they are not publicly available. For the purpose of fairness, all models are tested using their default parameter settings. In order to compare the performance of VQA and stalling-based QoE models, the quality of video without stalling

TABLE II
COMPARISON OF THE EXISTING QOE METHODS

	Stalling		Presentation quality	
QoE models	Regression function	Influencing factors	Regression function	Influencing factors
FTW [15]	exponential	stalling length, # of stalling	N/A	N/A
Mok's [40]	linear	stalling length, stalling frequency, initial buffering length	N/A	N/A
VsQM [17]	exponential	average stalling length per segment, # of stalling per segment, period per segment	N/A	N/A
Xue's [21]	logarithmic	stalling length, # of stalling, bit count of the stalling segment	linear	QP

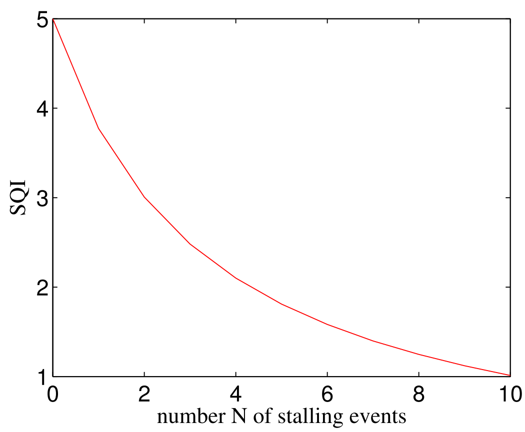


Fig. 4. SQI at different number of stalling events.

are estimated by VQA and the result is applied to the same video with stalling events. For the hybrid model in [21], the model parameter c is not given in the original paper. We set $c = 0.05$ such that the model achieves its optimal performance on the current database. A comparison of the four QoE models is shown in Table II. Three criteria are employed for performance evaluation by comparing MOS and objective QoE. Some of the criteria are included in previous tests carried out by the video quality experts group [45]. Other criteria are adopted in previous study [46]. These evaluation criteria are: 1) PLCC after a nonlinear modified logistic mapping between the subjective and objective scores [46]; 2) SRCC; 3) Mean absolute error (MAE) after the non-linear mapping. Among the above metrics, PLCC and MAE are adopted to evaluate prediction accuracy, and SRCC is employed to assess prediction monotonicity [45]. A better objective VQA measure should have higher PLCC and SRCC while lower MAE values. Figs. 6–8 summarize the evaluation results, which is somewhat disappointing because state-of-the-art QoE models do not seem to provide adequate predictions of perceived quality of streaming videos. Even the model with the best performance is only moderately correlated with subjective scores. These test results also provide some useful insights regarding the general approaches used in QoE models. First, the hybrid model [21] significantly outperforms QoS-QoE correlation models. This suggests that the importance of video presentation quality in QoE should not be underestimated. Second, even though modern VQA models cannot capture the experience loss of stalling, most of them performs reasonably well on the streaming video QoE database. These observations suggest a hybrid model that equips VQA methods as the video quality predictor would be more promising in QoE estimation.

IV. OBJECTIVE QUALITY-OF-EXPERIENCE MODEL OF STREAMING VIDEOS

Motivated by the observation and analysis provided in the previous section, we develop a unified QoE prediction model named Streaming QoE Index (SQI) by incorporating the video presentation quality and the impact of initial buffering and stalling events. In particular, we consider QoE as a combined experience of video presentation quality, stalling experience and their interaction.

A. 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. Currently, the development of the next-generation ISO base media file format that incorporates time-varying video quality metric is ongoing [47]. The manifest or metadata file 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

$$P_n = P_{i-1}. \tag{2}$$

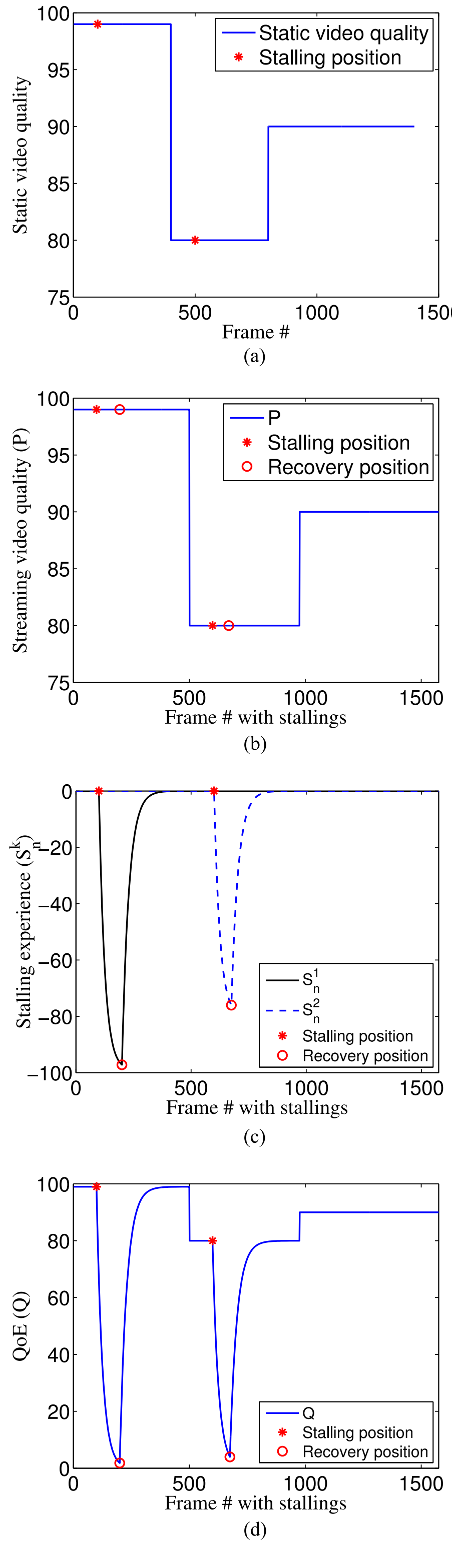


Fig. 5. An illustrative example of and channel responses at each frame. (a) Video presentation quality of the static video at each frame. '*' indicates the position of stalling. (b) Video presentation quality of the streaming video during playback at each frame. '*' indicates the position of stalling 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.

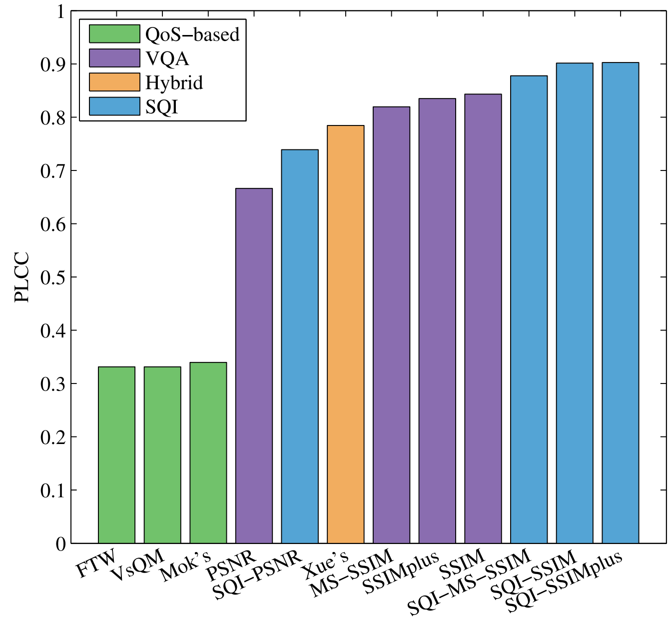


Fig. 6. PLCC of QoE models on streaming video QoE database.

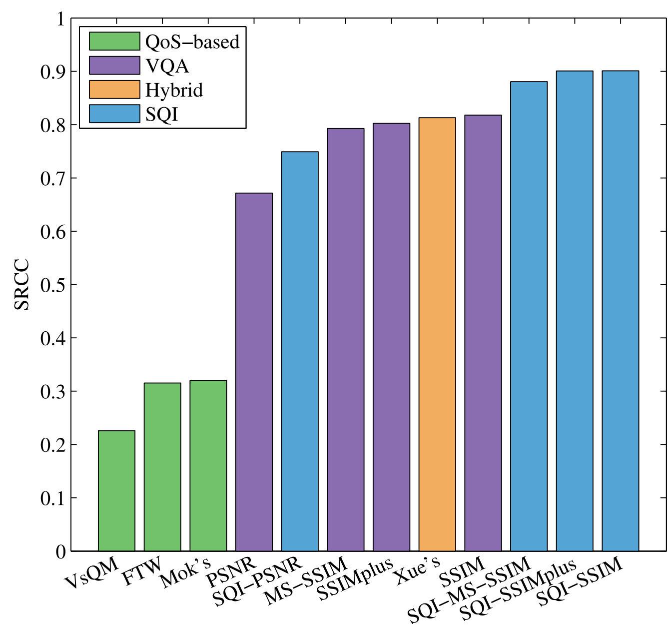


Fig. 7. SRCC of QoE models on streaming video QoE database.

B. *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.

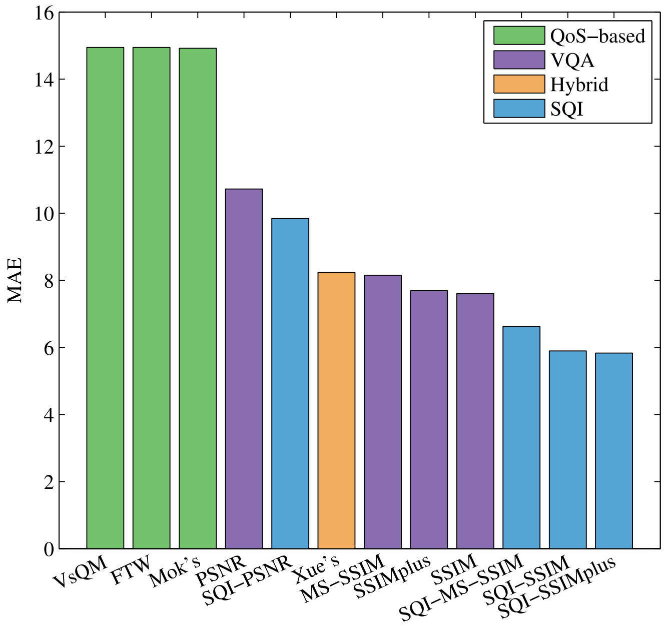


Fig. 8. MAE of QoE models on streaming video QoE database.

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 study on the impact of waiting time on user experience in queuing services [48] has a long history from both an economic and a psychological perspective, and has been recently extended to quantify the relationship between QoE and QoS in adaptive streaming [39]. The exponential decay function has been successfully used in previous studies [14], [15], [39]. 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 [14], [15], [39]. Third, QoE also depends on a behavioural hysteresis "after effect" [35]. 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 [49]

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

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

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

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: $\{T_0^{\text{init}}, T_1^{\text{init}}\}$ for initial delay and $\{T_0, T_1\}$ for other playback stallings, respectively. We also assume 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.

An important fact we have learned from the previous subjective study [27] is that the frequency of stalling negatively correlates with QoE for a streaming video of constant quality, sufficient length, and a fixed total length of stalling L. Although not explicitly defined in the expression, it can be shown that the effect of stalling frequency can be captured by the proposed model with a deliberate parameter selection. To see that, we first adopt the aforementioned test condition in [27] and assume $P_n = C$, where C is a positive constant. Then, the end-of-process QoE of the proposed model is fully determined by experience loss of stalling, which becomes a function of stalling frequency only. When the total length of stalling L is fixed and assume equal length of each individual stall, then the length of each stall is L/N, and the stalling frequency is inverse proportional to the total number of stalls N. Thus, we only need to check whether the cumulated QoE drop over all time

$$G(N) = \int_{-\infty}^{\infty} S(t)dt,\ \text{for}\ l_k = \frac{L}{N}, k = 1, 2, ..., N \tag{6}$$

is monotonically decreasing with respect to N. By substituting (4) and (5) into (6), we can simplify the expression as

$$G(N) = C\,(T_1 - T_0)\left\{N \exp\left[-\left(\frac{L}{NT_0}\right)\right] - N\right\} - CL \quad \text{for } N \geq 1, T_0 > 0, T_1 > 0, L > 0. \tag{7}$$

Let $g(x) = x\exp\{-(\frac{L}{xT_0})\} - x$, it is not hard to verify $\frac{dg(x)}{dx} < 0, \forall x \geq 1$. Therefore, the model is able to implicitly account for the effect of stalling frequency as long as $T_1 > T_0$.

In addition, we have also learned from previous subjective study [14] that the impact of stalling tends to saturate with the increase of the number of stalling events at a constant quality setting. Interestingly, with the independent and additive assumption, SQI is still able to predict that the overall QoE has an exponential-like response for each addition stalling event. To understand this, let us denote the video presentation quality of each frame/segment, the length of static video in seconds, the duration of each stalling events, the number of stalling events, and the overall QoE by P_n, T, T_s, N, and Q, respectively. In [14], the authors performed their subjective study with a constant quality setting, *i.e.*, $P_n = P$. According to (2), the video presentation quality that caused by the stalling events changes from $P_n = P, \forall n \in [0, T]$ to $P_n = P, \forall n \in [0, T + NT_s]$. According to (5), the overall stalling experience is $NS^k(T_s), \forall k \in [1, N]$. Thus, the overall QoE can be represented as $Q = \frac{(T+NT_s)P + NS_k(T_s)}{T+NT_s}$. We plot Q with respect to N on a 5-point absolute category rating (ACR) scale in Fig. 4, where it can be observed that the influence of each additional stalling event follows an exponential-like decreasing pattern in SQI.

In real-world applications, to measure the impact of stalling at individual frames, we convert the continuous function in (5) into its discrete form by sampling the function at each discrete time instance n:

$$S_n = S\left(\frac{n}{f}\right). \tag{8}$$

C. Overall QoE

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

$$Q_n = P_n + S_n. \tag{9}$$

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

$$A_n = \frac{(n-1)A_{n-1} + Q_n}{n} \tag{10}$$

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. 5.

TABLE III
SQI PARAMETERS

Parameter	Description
T_0	rate of dissatisfaction in stalling event
T_1	strength of memory in stalling event
T_0^{init}	rate of dissatisfaction in initial buffering event
T_1^{init}	strength of memory in initial buffering event
P_0	expectation on initial quality of the video

D. Implementation Details

Throughout the paper, the proposed SQI uses the following parameter settings: $T_0^{\text{init}} = 2$, $T_1^{\text{init}} = 0.5$, $T_0 = 1$, $T_1 = 1.2$ and $P_0 = 0.8 \cdot |(V(\cdot)|$, where $|V(\cdot)|$ is the dynamic range of adopted VQA kernel, *e.g.* PSNR ranges from 0 to infinity (in the actual computation, we set the range of PSNR to 0–50); SSIM and MS-SSIM range from -1 to 1; and SSIMplus ranges from 0 to 100. These values are somewhat arbitrary, but we find that in our current experiments, the performance of the SQI is fairly insensitive to variations of T_0^{init}, T_1^{init}, T_0 and T_1 at least within an order of magnitude of the parameter values. P_0 is rather insensitive from $0.5|(V(\cdot)|$ (Xue's [21] selection) to $|(V(\cdot)|$. The parameters are summarized in the Table III. Note that the initial buffering parameters do not have to satisfy the stalling frequency because it cannot occur more than once in one session. In real-world applications, the proposed scheme may include two step computations on the client side. First, stalling events are detected in the video player. A straightforward way to detect stalling events is to inspect the player progress every x milliseconds, *e.g.* 50. If the player has not advanced as much as it is expected to, then we can infer a stalling has occurred. By taking the difference between the expected progress and actual progress, the duration and frequency of stalling can be measured reliably. In the second step, which is only necessary in the applications that require an end-of-process score, is the computation of the QoE cumulation. Both steps demand minimum computation and can be updated in real time. Moreover, the instantaneous QoE prediction is a valuable property for many applications such as live streaming quality monitoring and adaptive streaming decision making.

V. VALIDATION

To the best of our knowledge, there is no other subject-rated publicly available video database that have a combination of compression distortion, initial buffering, and stalling events. Thus, we validate SQI model using the streaming video QoE database described in Section III and compare its performance against eight existing objective QoE models. Among the eight QoE models, four VQA algorithms including PSNR, SSIM [8], MS-SSIM [9] and SSIMplus [12], are employed as the frame-level video presentation quality measures. They also provide useful baseline comparisons. PLCC, SRCC and MAE are calculated to evaluate the performance of all QoE models. The performance comparison results are provided in Figs. 6–8, respectively. It can be seen that the proposed method delivers the

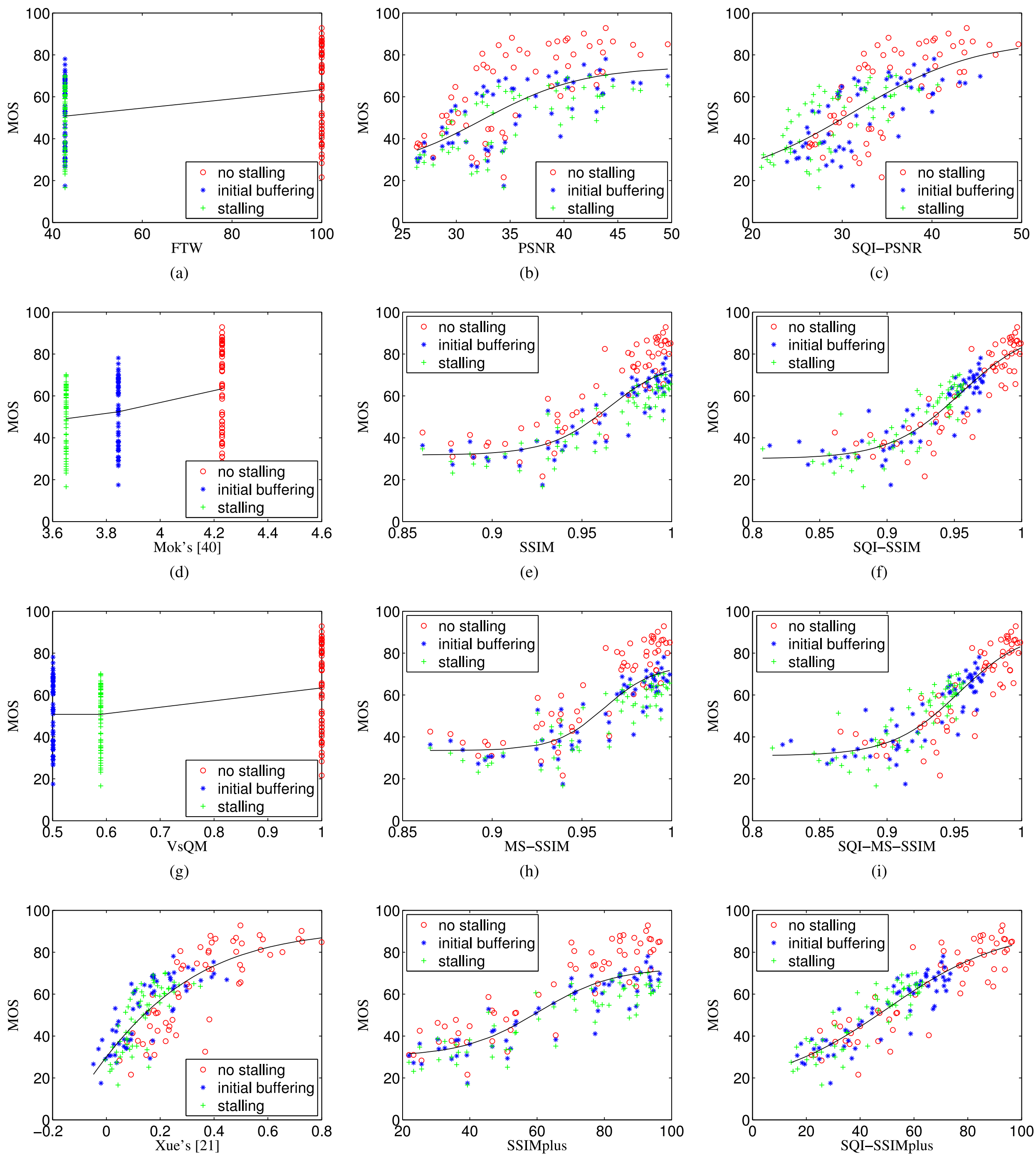


Fig. 9. Predicted QoE versus MOS.

best performance in predicting subjective QoE on the streaming video QoE database with both compression and frame-freeze impairment.

Fig. 9 shows the scatter plots of the MOS prediction results for each QoE model. The existing QoE models, presentation VQA quality with and without incorporating the proposed methods are listed in the first, second and third columns, respectively. We have two observations here. First, the proposed SQI models significantly outperform their baseline presentation VQA models. It is obvious that a higher compactness in the scatter plots is achieved by applying the proposed model, which adds proper penalties for initial buffering and stalling in addition to the presentation quality impairment. Second, the best performance is obtained by combining the proposed method with the SSIMplus [12] VQA model.

To ascertain that the improvement of the proposed model is statistically significant, we carry out a statistical significance analysis by following the approach introduced in [46]. First, a

TABLE IV
STATISTICAL SIGNIFICANCE MATRIX BASED ON F-STATISTICS ON THE STREAMING VIDEO QOE DATABASE

	FTW [15]	Mok's [40]	VsQM [17]	Xue's [21]	PSNR	SSIM [8]	MS-SSIM [9]	SSIMplus [12]	SQI-PSNR	SQI-SSIM	SQI-MS-SSIM	SQI-SSIMplus
FTW [15]	-	-	-	0	0	0	0	0	0	0	0	0
Mok's [40]	-	-	-	0	0	0	0	0	0	0	0	0
VsQM [17]	-	-	-	0	0	0	0	0	0	0	0	0
Xue's [21]	1	1	1	-	1	-	-	-	1	0	0	0
PSNR	1	1	1	0	-	0	0	0	-	0	0	0
SSIM [8]	1	1	1	-	1	-	-	-	1	0	-	0
MS-SSIM [9]	1	1	1	-	1	-	-	-	1	0	0	0
SSIMplus [12]	1	1	1	-	1	-	-	-	1	0	-	0
SQI-PSNR	1	1	1	0	-	0	0	0	-	0	0	0
SQI-SSIM	1	1	1	1	1	1	1	1	1	-	-	-
SQI-MS-SSIM	1	1	1	1	1	-	1	-	1	-	-	-
SQI-SSIMplus	1	1	1	1	1	1	1	1	1	-	-	-

A symbol "1" means that the performance of the row model is statistically better than that of the column model, a symbol "0" means that the row model is statistically worse, a symbol "-" means that the row and column models are statistically indistinguishable.

nonlinear regression function is applied to map the objective quality scores to predict the subjective scores. We observe that the prediction residuals all have zero-mean, and thus the model with lower variance is generally considered better than the one with higher variance. We conduct a hypothesis testing using F-statistics. Since the number of samples exceeds 50, the Gaussian assumption of the residuals approximately hold based on the central limit theorem [50]. The test statistic is the ratio of variances. The null hypothesis is that the prediction residuals from one quality model come from the same distribution and are statistically indistinguishable (with 95% confidence) from the residuals from another model. After comparing every possible pairs of objective models, the results are summarized in Table IV, where a symbol '1' means the row model performs significantly better than the column model, a symbol '0' means the opposite, and a symbol '-' indicates that the row and column models are statistically indistinguishable. It can be observed that most existing QoE models are statistically indistinguishable from each other, while the proposed model is statistically better than all other methods on the streaming video QoE database.

It can be observed from the experiments that the QoS-based QoE models [15], [17], [40] do not perform well on the database. The major reason is that QoS-based models (i.e., FTW [15], Mok's [40], and VsQM [17]), do not take the presentation quality of the videos into consideration except for their bitrates. A common "mistake" is to equate bitrate with quality, or assume a constant bitrate implies a constant presentation quality. This is highly problematic because videos coded at the same bitrate but of different content could have drastically different presentation quality. This is often the most dominant QoE factor, and in many cases all other factors (such as stalling) become only secondary. Indeed, this is quite apparent from our test results, where even PSNR, a very crude presentation quality measure that does not take into account any initial buffering or stalling at all, performs significantly better than QoS-based methods that ignore presentation quality. By contrast, the proposed method not only builds upon the most advanced presentation quality model (*e.g.*, SSIMplus, which has been shown to be much better than PSNR and other VQA measures), but moves one step further by capturing the interactions between video presentation quality and the impact of stalling.

VI. 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 SQI model is simple in expression and effective in performance.

Future research may be carried out in many directions. First, although we have tried our best to construct a database that comprise as many content type as possible, the experiment is by no means exhaustive. A comprehensive subject-rated database that consists of more content types, stalling patterns and video quality variations is desired to better understand the behaviors 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 and its possible interactions with other QoE influencing factors needs to be exploited. Fourth, how to integrate the QoE model into the adaptive streaming decision making engine for optimal playback control is another challenging problem that is worth further investigations.

REFERENCES

[1] Apple Inc., "HTTP live streaming technical overview 2013," [Online]. Available: https://developer.apple.com/library/ios/documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html , Accessed Feb. 15, 2015.
[2] A. Zambelli, "Smooth streaming technical overview," [Online]. Available: http://www.iis.net/learn/media/on-demand-smooth-streaming/smoothstreamingtechnical-overview, Accessed Feb. 15, 2015.
[3] Adobe Systems Inc., "HTTP dynamic streaming 2013," [Online]. Available: http://www.adobe.com/products/hds-dynamic-streaming.html, Accessed Feb. 15, 2015.

[4] DASH Industry Forum, "For promotion of MPEG-DASH 2013," [Online]. Available: http://dashif.org, Accessed Feb.12, 2015.

[5] T. Stockhammer, "Dynamic adaptive streaming over http–: Standards and design principles," in *Proc. ACM Conf. Multimedia Syst.*, 2011, pp. 133–144.

[6] "Cisco IBSG Youth Focus Group, Cisco IBSG youth survey," [Online]. Available: http://www.cisco.com/c/dam/en_us/about/ac79/docs/ppt/Video_Disruption_SP_Strategies_IBSG.pdf, Accessed Jan. 12, 2015.

[7] Conviva Inc., "Viewer experience report," 2013. [Online]. Available: http://www.conviva.com/conviva-customer-survey-reports/ott-beyond-entertainment-csr/, Accessed Jan. 15, 2015.

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

[9] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. IEEE Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, vol. 2, pp. 1398–1402.

[10] K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 335–350, Feb. 2010.

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

[12] A. Rehman, K. Zeng, and Z. Wang, "Display device-adapted video Quality-of-Experience assessment," in *Proc. SPIE*, Feb. 2015, vol. 9394, pp. 939 406–939 406.

[13] K. Watanabe, J. Okamoto, and T. Kurita, "Objective video quality assessment method for evaluating effects of freeze distortion in arbitrary video scenes," in *Proc. Electron. Imag. Int. Soc. Opt. Photon.*, Jan. 2007, pp. 64 940–64 940.

[14] 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.

[15] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet video delivery in YouTube: From traffic measurements to quality of experience," in *Data Traffic Monitoring Anal.*, Jan. 2013, pp. 264–301.

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

[17] D. Rodriguez, J. Abrahao, D. Begazo, R. Rosa, and G. Bressan, "Quality metric to assess video streaming service over TCP considering temporal location of pauses," *IEEE Trans. Consumer Electron.*, vol. 58, no. 3, pp. 985–992, Aug. 2012.

[18] 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.

[19] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A time-varying subjective quality model for mobile streaming videos with stalling events," in *Proc. SPIE*, Sep. 2015, vol. 9599, pp. 959 911–959 911.

[20] K. D. Singh, Y. Hadjadj-Aoul, and G. Rubino, "Quality of experience estimation for adaptive HTTP/TCP video streaming using H. 264/AVC," in *Proc. IEEE Int. Conf. Consum. Commun. Netw.*, Jan. 2012, pp. 127–131.

[21] 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 Expo.*, Jul. 2014, pp. 1–6.

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

[23] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tut.*, vol. 17, no. 1, pp. 469–492, Sep. 2014.

[24] M.-N. Garcia *et al.*, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. IEEE Int. Conf. Quality Multimedia Exp.*, Sep. 2014, pp. 141–146.

[25] R. R. Pastrana-Vidal, J. C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic frame dropping impact on quality perception," in *Proc. SPIE*, Jan. 2004, vol. 5292, pp. 182–193.

[26] Y. Qi and M. Dai, "The effect of frame freezing and frame skipping on video quality," in *Proc. IEEE Int. Conf. Intell. Inform. Hiding Multimedia Signal Process.*, Dec. 2006, pp. 423–426.

[27] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652–671, Oct. 2012.

[28] N. Staelens *et al.*, "Assessing quality of experience of IPTV and video on demand services in real-life environments," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 458–466, Dec. 2010.

[29] A. Floris, L. Atzori, G. Ginesu, and D. Giusto, "QoE assessment of multimedia video consumption on tablet devices," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 1329–1334.

[30] L. Atzori, A. Floris, G. Ginesu, and D. D. Giusto, "Quality perception when streaming video on tablet devices," *J. Visual Commun. Image Representation*, vol. 25, no. 3, pp. 586–595, Apr. 2014.

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

[32] 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. Quality Multimedia Expo.*, Jul. 2013, pp. 64–69.

[33] E. Sebastian and R. Alexander, *Quality and Quality of Experience*. New York, NY, USA: Springer-Verlag, Jan. 2014.

[34] 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 Expo.*, Sep. 2014, pp. 129–134.

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

[36] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[37] Z. Wang, H. R. Sheikh, and A. C. Bovik, "Objective video quality assessment," in *The Handbook of Video Databases: Design and Applications*. Boca Raton, FL, USA: CRC Press, Sep. 2003.

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

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

[40] R. Mok, E. Chan, and R. Chang, "Measuring the quality of experience of HTTP video streaming," in *Proc. IFIP/IEEE Int. Symp. Integrated Netw. Manage.*, May 2011, pp. 485–492.

[41] E. Gelenbe, "Random neural networks with negative and positive signals and product form solution," *Neural Comput.*, vol. 1, no. 4, pp. 502–510, Aug. 1989.

[42] P. Fröhlich, S. Egger, R. Schatz, M. Mühlegger, K. Masuch, and B. Gardlo, "QoE in 10 s: Are short video clip lengths sufficient for quality of experience assessment?" in *Proc. IEEE Int. Conf. Quality Multimedia Expo.*, Jul. 2012, pp. 242–247.

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

[44] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1427–1441, Jun. 2010.

[45] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Apr. 2000. [Online]. Available: http://www.vqeg.org/

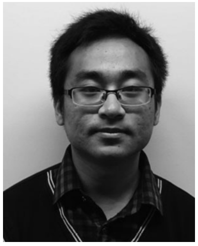
[46] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[47] *Information technology MPEG systems technologies: Carriage of timed metadata metrics of media in ISO base media file format*, ISO/IEC 23001-10, Sep. 2015.

[48] P. Kumar, M. U. Kalwani, and M. Dada, "The impact of waiting time guarantees on customers' waiting experiences," *Marketing Sci.*, vol. 16, no. 4, pp. 295–314, Nov. 1997.

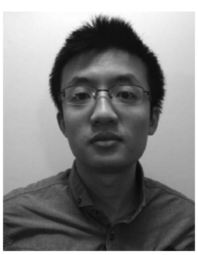
[49] H. Ebbinghaus, *Memory: A Contribution to Experimental Psychology*. New York, NY, USA: Teachers college, Columbia University, Oct. 1913.

[50] D. C. Montgomery, *Applied Statistics and Probability for Engineers,* 6th ed. New York, NY, USA: Wiley, 2013.



Zhengfang Duanmu (S'15) received the B.A.Sc degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2015, where he is currently working toward the M.A.Sc degree in electrical and computer engineering. His research interest lies in perceptual image processing and quality of experience.

Kai Zeng received the Ph.D. degree in electrical and computer engineering from University of Waterloo, Waterloo, ON, Canada, in 2013, where he is currently a Post-Doctoral Fellow. His research interests include computational video and image communication and processing. Dr. Zeng received the IEEE Signal Processing Society student travel grant in 2010 and 2012, and the prestigious 2013 Chinese Government Award for Outstanding Students Abroad.



Kede Ma (S'13) received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2012, and the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, where he is currently working toward the Ph.D. degree in electrical and computer engineering. His research interest lies in perceptual image processing.

Abdul Rehman received the Masters degree in communications engineering from Technical University Munich, Germany, and the Ph.D. degree in information and communication systems from the University of Waterloo, Canada. He is currently the President & CEO of SSIMWave, a company, he co-founded in 2013, dedicated to delivering excellence in visual quality-of-experience. He leads the development of SSIMWaves state-of-the-art video QoE measurement and optimization products geared towards the media, communication, and entertainment industry. His research interests include image and video processing, coding and quality assessment, and multimedia communications.

Zhou Wang (S'99–M'02–SM'12–F'14) received the Ph.D. degree from the University of Texas at Austin, Austin, TX, USA, in 2001. He is currently a Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research interests include image processing, coding, and quality assessment; computational vision and pattern analysis; multimedia communications; and biomedical signal processing. He has more than 100 publications in these fields with more than 30 000 citations (Google Scholar). Dr. Wang serves as a Senior Area Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING (2015-present), and an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2016-present). Previously, he served as a Member of IEEE Multimedia Signal Processing Technical Committee (2013–2015), an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING (2009–2014), *Pattern Recognition* (2006-present) and IEEE SIGNAL PROCESSING LETTERS (2006–2010), and a Guest Editor of IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING (2013–2014 and 2007–2009), *EURASIP Journal of Image and Video Processing* (2009–2010), and Signal, Image and Video Processing (2011–2013). He is a Fellow of Canadian Academy of Engineering, and have received the 2015 Primetime Engineering Emmy Award, the 2014 NSERC E.W.R. Steacie Memorial Fellowship Award, the 2013 IEEE Signal Processing Magazine Best Paper Award, the 2009 IEEE Signal Processing Society Best Paper Award, the 2009 Ontario Early Researcher Award, and the ICIP 2008 IBM Best Student Paper Award (as Senior Author).