This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

A Continuous QoE Evaluation Framework for Video Streaming over HTTP

Nagabhushan Eswara, *Student Member*, *IEEE*, Manasa K, Avinash Kommineni, Soumen Chakraborty, Hemanth P. Sethuram, Kiran Kuchi, Abhinav Kumar, *Member*, *IEEE*, Sumohana S. Channappayya, *Member*, *IEEE*

***Abstract*—A continuous evaluation of the end user's Quality-of-Experience (QoE) is essential for efficient video streaming. This is crucial for networks with constrained resources that offer time varying channel quality to its users. In Hyper Text Transfer Protocol (HTTP) based video streaming, the QoE is measured by quantifying the perceptual impact of distortions caused by rate adaptation or interruptions in playback due to rebuffering events. The resulting impact on the QoE due to these distortions has been studied individually in the literature. However, the QoE is determined by an interplay of these distortions, and therefore necessitates a combined study of them. To the best of our knowledge, there is no publicly available database that studies these distortions jointly on a continuous time basis. In this paper, our contributions are two-fold. Firstly, we present a database consisting of videos at Full High Definition and Ultra High Definition resolutions. We consider various levels of rate adaptation and rebuffering distortions together in these videos as experienced in a typical realistic setting. A subjective evaluation of these videos is conducted on a continuous time scale. Secondly, we present a QoE evaluation framework comprising a learning based model during playback and an exponential model during rebuffering. Further, we perform an objective evaluation of popular video quality assessment and continuous time QoE metrics over the constructed database. The objective evaluation study demonstrates that the performance of the proposed QoE model is superior to that of the objective metrics. The database is publicly available for download at http://www.iith.ac.in/∼lfovia/downloads.html.**

***Index Terms*—DASH, Full HD, HTTP video streaming, QoE, rate adaptation, rebuffering, recency effect, STSQ, subjective study, support vector regression, Ultra HD.**

I. Introduction

In recent years, Hyper Text Transfer Protocol (HTTP) based streaming has become a popular choice for video streaming as most of the content delivery networks are configured to support HTTP services. To cater to the needs of HTTP video streaming and interoperability of services, the Moving Picture Experts Group has developed Dynamic Adaptive Streaming over HTTP (DASH), a standard that defines the framework for streaming over HTTP [1].

In a typical client-server scenario of video streaming, when a video is transmitted from the server, the video data at the client is stored in a buffer before the playback is started. This allows the client to cache the video ahead of time when the network conditions are good and compensate for the times when the conditions turn bad. Network conditions vary due to several factors such as congestion due to increased traffic (which is quite common during peak hours), network coverage, channel variations due to user mobility etc. Under consistently poor network conditions, the video buffer runs out of data causing the playback to stall. This is known as a *rebuffering event*. Given the significant increase in the demand for online video services [2] relative to the growth of network infrastructure, the likelihood of the occurrence of rebuffering events has increased. The playback interruptions due to rebuffering events often result in a drop in the user's quality of experience (QoE). The DASH standard allows the client to adapt the video rate to combat network fluctuations so that an uninterrupted playback is maintained. It has been shown that the rate adaptation in DASH can lead to a time varying video quality which in turn can lead to a degradation in the user's QoE [3]. However, the joint impact of rate adaptation and rebuffering events on QoE has not been studied. This is the motivation for our work.

In this work, we design and perform an experiment to mimic the playback scenarios typical to video streaming. To study their joint impact on QoE, rate adaptation and rebuffering distortions are artificially introduced in Full High Definition (FHD) and Ultra High Definition (UHD) videos. A database composed of FHD and UHD videos with the aforementioned distortions is constructed and a subjective evaluation of these videos is conducted for QoE assessment on a continuous time scale. The QoE of the corresponding reference videos is also evaluated. Based on the subjective scores collected from the experiment, we design an objective framework to perform continuous QoE evaluation. The framework comprises a learning based QoE evaluation during playback and a parametric QoE evaluation during rebuffering. Using the framework, we propose a QoE prediction model by employing

This work was supported in part by the Department of Science and Technology, Government of India under the Start Up Research Grant (Young Scientists) scheme of Science and Engineering Research Board (SERB) for the Project "Performance evaluation of cellular networks in unlicensed spectrum co-existing with WiFi" (ref. no. YSS/2015/001859), the Project "Converged Cloud Communication Technologies" (ref. no. R-23011/03/2014) by the Ministry of Electronics and Information Technology, Government of India, and Intel India PhD Fellowship by Intel Corporation.

N. Eswara, M. K, S. S. Channappayya are with the Laboratory for Video and Image Analysis (LFOVIA), Department of Electrical Engineering, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India. (email: ee13p1004@iith.ac.in; ee12p1002@iith.ac.in; sumohana@iith.ac.in)

A. Kommineni, K. Kuchi and A. Kumar are with the Department of Electrical Engineering, Indian Institute of Technology Hyderabad, Kandi, Telangana 502285, India. (email: ee12b1014@iith.ac.in; kkuchi@iith.ac.in; abhinavkumar@iith.ac.in)

S. Chakraborty and H. P. Sethuram are with Intel Corporation, Bengaluru, Karnataka, India. (email: soumen.chakraborty@intel.com; hemanth.p.sethuram@intel.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

a Support Vector Regression (SVR) for prediction during playback and an exponential depreciation model for prediction during rebuffering. The proposed QoE model is shown to significantly outperform the QoE predictions from popular video quality assessment (VQA) and continuous time QoE metrics.

The rest of the paper is organized as follows. We review the relevant literature in Section II. The construction and the description of the database is presented in Section III. In Section IV, the subjective evaluation is explained. The proposed QoE framework and the prediction model are described in Section V. The performance evaluation of the proposed model is discussed in Section VI, followed by concluding remarks in Section VII.

II. Background

According to the International Telecommunications Union (ITU), QoE is defined as *the overall acceptability of an application or service, as perceived subjectively by the end user* [4]. QoE centric design has become a key aspect of the system design today as the primary consumers of many data intensive applications are directly the end users. For media delivery applications, the QoE based system design can play a particularly vital role in enhancing the overall performance of the system. Given the widespread application of HTTP streaming services, the assessment of the user's QoE becomes very important.

The key factors that influence the QoE of a user in a typical HTTP video streaming session are as follows: 1) source video quality, 2) time variation in the video quality, 3) initial buffering time (startup delay), 4) rebuffering (stalling) events and 5) video content. Source video quality, as the name suggests, refers to the quality of the video captured at the source. It is determined by several factors such as resolution, frame rate, encoding format etc. It is also affected by other distortions such as motion blur, camera shake etc. that are introduced at the time of video capture. In a HTTP video streaming framework such as DASH, the source videos are encoded at different rates and at different resolutions. These videos are broken down into short duration sequences called segments. This enables the video client to dynamically adapt across the rates, known as *rate adaptation*, according to the variations in the network. However, the rate adaptation causes a variation in the perceptual quality over time, known as the time varying quality. In order to assess the time varying quality effectively, one has to rely on robust VQA metrics. Traditional Image Quality Assessment (IQA) metrics used for VQA such as Peak Signal-to-Noise Ratio (PSNR) and Multi Scale - Structural Similarity Index Measure (MS-SSIM) [5] fail to capture temporal distortions present in the video. Full Reference VQA metrics such as MOVIE [6], STMAD [7] etc., Reduced Reference VQA metrics such as Video-RRED [8], No Reference VQA metrics such as Video-BLIINDS [9], VIIDEO [10] are not capable of assessing perceptual quality and QoE of the videos subjected to distortions such as rate adaptation (or time varying quality) and rebuffering (or frame freeze) that are typically encountered in adaptive streaming.

In [11], the LIVE Mobile Video Quality Assessment Database provides a subjective study of various distortions including frame freezes (rebuffering) and temporally varying compression rates (rate adaptation) similar to the ones encountered in online video viewing, and when viewed on handheld devices (mobile and tablet). The video sequences in the database contain artificially introduced distortions at different levels and a subjective evaluation was conducted on handheld devices to measure the quality on a continuous time scale. However, the rate adaptation and frame freeze distortions have been studied separately. In order to assess QoE, these two distortions must be considered jointly as the realistic viewing experience is determined by a combination of both these distortions. Further, the duration of the reference videos considered in the database is 15 seconds, which could be inadequate to study the long term effects of these distortions on users' QoE.

In [12], the relation between the QoE and the network Quality-of-Service (QoS) based on rebuffering events has been investigated. Startup delay, rebuffering duration and rebuffering frequency have been identified as the three key factors that impact QoE. Further, it is shown that the rebuffering frequency can potentially have a large influence on QoE. In [13], the authors propose a YouTube QoE model based on IQX (exponential Interdependency of Quality of service and quality of eXperience) hypothesis [14] to relate QoE and stallings (or rebufferings) and validate it through a subjective evaluation. The study brings out the fact that startup delays upto 10 seconds have negligible impact on QoE compared to rebuffering events. It also points out that the QoE is heavily influenced by the stalling length (or the rebuffering duration) and the number of stallings. However, the study does not consider the rate adaptation distortion for measuring the QoE. Also, the subjective study data is not available for further analysis.

In [15], the effects of initial buffering and rebuffering caused by video delivery impairments are investigated. The authors propose the Delivery Quality Score (DQS) model to predict the QoE. The DQS model employs a combination of raised cosine and ramp functions to measure the continuous time changes in QoE when subjected to initial buffering and rebuffering events. The model is validated using Mean Opinion Scores (MOS) obtained through Absolute Category Rating with Hidden Reference (ACR-HR) method of subjective evaluation. Since only a single opinion score was recorded from the subjects per video, the continuous time QoE scores are synthesized using their proposed model such that the score computed at the end of the video correlates with the subjective score. The DQS model is employed to synthesize the continuous time scores in [16] for estimating the time-varying quality for videos with rebuffering events. However, in both [15] and [16], the lack of ground-truth continuous time subjective scores leads to inadequate validation of the synthesized QoE scores although the predicted final QoE score is shown to correlate well with the overall subjective opinion score. A QoE study in the LIVE-Avvasi Mobile Video Database provides 180 videos derived from 24 reference videos by incorporating a wide variety of rebuffering events, along with the associated opinion

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

scores on the viewing experience obtained through a subjective evaluation [17], [18]. However, the availability of only a single Difference Mean Opinion Score for every video is not helpful in understanding the continuous time variation of the QoE prior to and post rebuffering. Also, the time varying quality due to rate adaptation is not considered for the assessment of QoE in both [15] and [17].

To address the problem of QoE assessment, a subjective evaluation of videos having time varying quality has been conducted in [3]. Based on the scores collected, the Time Varying Subjective Quality (TVSQ) metric has been proposed to predict the continuous time perceptual quality. TVSQ employs Short Time Subjective Quality (STSQ) to measure the current video quality and a Hammerstein-Wiener model to capture the memory effects and nonlinearities involved in predicting the current perceptual quality [3]. TVSQ is effective in measuring the time varying perceptual quality due to rate adaptation. However, it does not consider the impact of the rebuffering events on the user's QoE. An enhanced version of TVSQ known as enhanced TVSQ (eTVSQ) has been proposed in [19] to jointly account for the rebuffering distortion and the rate adaptation distortion. However, the proposed STSQ depreciation methodology employed during rebuffering for predicting the perceived quality lacks substantiation, and hence cannot be used without adequate validation.

In [20], a QoE-oriented transcoding approach to enhance the quality of mobile 3D video streaming is proposed. The proposed approach is based on learning pre-controlled QoE patterns of 3D video contents. A no-reference framework for capturing video QoE inside the network core known as MintMOS is presented in [21]. MintMOS estimates the QoE of a video stream in network transit and offers suggestions to improve it based on a k-dimensional QoE space. However, the measured QoE using MintMOS is limited by the k-parameters chosen while constructing the QoE space. Since the constructed QoE space is limited, the QoE prediction accuracy outside this space is unknown. It must be noted that in both [20] and [21], only the overall QoE of the streamed video is measured and not the continuous time QoE. A comprehensive study of the streaming strategies has been explored in [22]. Towards addressing some of the open questions on improving QoE, the work studies the effect of aspects such as rate adaptation levels, switching strategies, content, chunk duration and recency effects on the QoE. The authors provide insights and inferences on each of these aspects based on multiple laboratory and crowdsourced subjective studies. However, the study does not address the impairments caused on QoE due to rebuffering. Also, the continuous time QoE variations are not studied since only an overall opinion score was recorded per video from the subjects. In [23], the authors study the dependency between the presentation quality and playback stallings based on a subjective study. Results indicate that there is a significant drop in the user quality when there are stallings. It is also inferred that the degradation in quality is very severe when stalling occurs in the middle of the video rather than at the beginning. The study finds that the impact of stallings on the user quality is not independent and there is a significant influence of the presentation quality as well.

While most on-demand content delivery services offer videos at FHD resolution, the video delivery at UHD resolution is increasingly being adopted. In [24], it is reported that UHD will not only find its applicability in large screen presentations such as cinema halls and other public venues but may soon become a part of our personal spaces in mobile and non-mobile environments. UHD resolution is expected to considerably improve the benefits of the users by bringing a stronger sensation of reality or presence, higher transparency to the real world, more content information and so on. In [25], it is statistically shown that there is a significant difference in the viewing strategies while viewing HD and UHD images. It is demonstrated that the UHD resolution images grab visual attention more than HD images and subjects tend to fixate at a few attentive regions in the images more intently when viewing in UHD. An analysis of the subjective quality assessment of 4K-UHD encoded by HEVC for broadcasting services is presented in [26]. [27] considers an objective quality comparison of 4K-UHD and upscaled 4K-UHD videos using AVC and HEVC compressions. Thus, it is meaningful to consider QoE assessment of videos at FHD and UHD resolutions.

Subjective studies involving UHD videos are gaining popularity owing to the increased consumption of videos at UHD resolution. In [28], a UHD video database consisting of 15 videos of 10 seconds each is presented. A set of videos at 4K resolution has been made freely available by Netflix [29]. However, the 4K resolution employed here is the Digital Cinema Initiatives standard resolution 4096×2160, which is slightly different from the standard UHD resolution 3840×2160 with an aspect ratio of 16:9. Further, the available 4K videos are not long enough to study for the QoE with rate adaptation and rebuffering distortions.

All QoE subjective studies conducted thus far have employed either HD (1280×720) or lower resolution video sequences. Further, most of the QoE subjective studies are performed by considering the overall opinion scores from the subjects. However, for monitoring streaming sessions in real time and optimizing the end user QoE dynamically, the QoE must be evaluated on a continuous time scale. To the best of our knowledge, there are no studies on the continuous time QoE assessment at FHD and UHD resolutions till date. Also, there is no publicly available database that studies the joint impact of rate adaptation and rebuffering distortions on continuous time QoE. There is a need for a continuous time QoE evaluation metric that jointly considers both these distortions. These shortcomings motivated us to create a new database of FHD and UHD videos that is dedicated to the understanding of QoE in the presence of rate adaptation and rebuffering distortions.

We first present our Laboratory FOr Video and Image Analysis (LFOVIA) Video QoE database along with a subjective QoE evaluation. We also present a QoE evaluation framework based on the subjective scores and then propose a model for QoE prediction based on this framework. With a continuous increase in the demand for videos at higher resolutions and a push towards the design of QoE centric systems for video streaming, we believe that our database will be useful for the development of sophisticated QoE prediction models. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

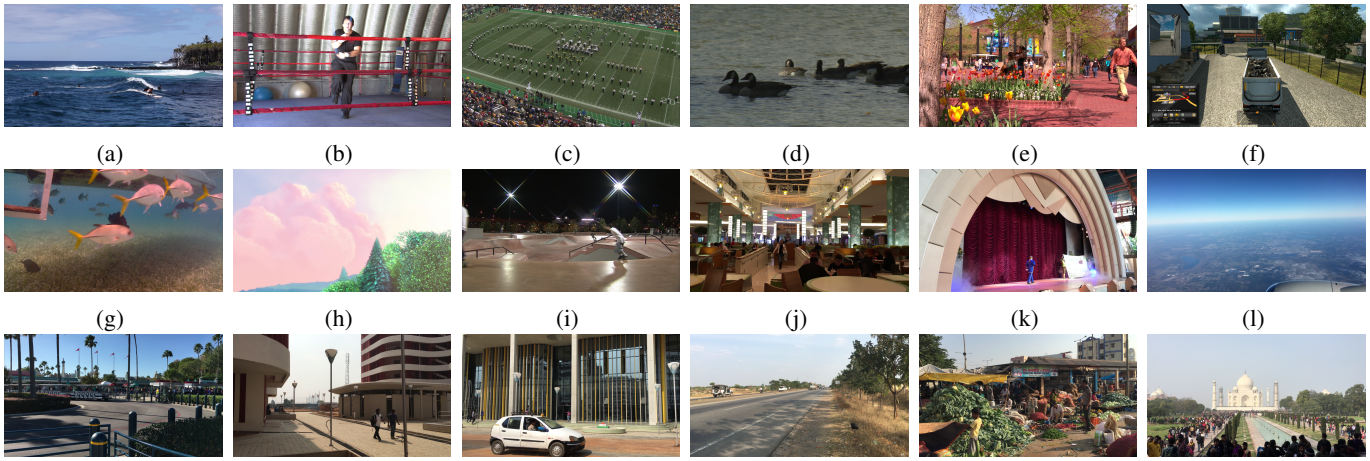


Fig. 1: Representative frames from FHD reference videos (a)-(i) and UHD reference videos (j)-(r).

details of the database are discussed in Section III.

III. Database Description

The proposed LFOVIA-QoE database consists of 54 videos out of which 18 are reference (undistorted) and 36 are distorted videos. The reference videos consist of 9 videos each at FHD and UHD resolutions. For every reference video, 2 distorted videos are generated, resulting in a total of 36 distorted videos. We first discuss the details of the source videos that are used for constructing reference videos followed by an explanation on the generation of distorted videos. The resolution of the video sequences used in the study is described in the next subsection.

A. Resolution

The FHD resolution videos in the proposed database are taken from various sources that are either in the public domain or are freely available for academic research and development. Two of the FHD reference videos are taken from Twitch [29], a subsidiary unit of Amazon, one from the Blender Foundation [30] and the rest from Institute for Telecommunication Sciences (ITS) which is the research and engineering laboratory of the National Telecommunications and Information Administration (NTIA), an agency of the United States Department of Commerce [31]. The UHD videos have been generated by us and the content has been acquired at various locations including the campus of Indian Institute of Technology (IIT) Hyderabad, India, a local vegetable market, the Taj Mahal and Disney World using an Apple iPhone 6S 4K-UHD camera under default settings. The video shooting involved 8-bit YUV 420 format and the videos are encoded at a rate of about 50 Mbps in Apple's QuickTime video format (MOV).

B. Duration

The reference video duration is another important factor in the QoE study. Hence, it needs to be carefully chosen depending on the objective of the study. The study in [32] suggests that a video duration of 10 seconds is sufficient for overall QoE assessment. The duration of the videos considered in the QoE study in [13] are 30 and 60 seconds. In [17], the video clips ranged between 29 and 134 seconds in duration, post introduction of rebuffering impairments. The study in [3] employs a relatively longer video duration of 300 seconds. A continuous time evaluation on short video clips does not allow enough room for studying the impact of rate adaptation and rebuffering distortions on QoE. On the other hand, too long a video duration could pose the risk of the user learning the distortion patterns, thereby affecting the true QoE assessment. Thus, the video duration should be neither too short nor too long for continuous evaluation in order to ensure a fair QoE assessment. Hence, in our study, we consider reference videos with a uniform playback duration of 120 seconds across all the resolutions.

The videos hosted by ITS-NTIA are of short durations (between 14s-30s) and involve two frame rates of 25 and 30 frames per second (fps). The Twitch videos are around

TABLE I: Nomenclature of the reference videos in the database.

Video	Name	Resolution	Frame Rate	Format
(a)	Surfers	FHD	30	YUV422
(b)	Boxing	FHD	30	YUV422
(c)	Football	FHD	30	YUV422
(d)	Goose Park	FHD	30	YUV422
(e)	Tulip	FHD	25	YUV422
(f)	Euro Truck	FHD	60	YUV420
(g)	Under Water	FHD	30	YUV422
(h)	Big Buck Bunny	FHD	24	YUV420
(i)	Skating	FHD	30	YUV422
(j)	Restaurant	UHD	30	YUV420
(k)	Theme Show	UHD	30	YUV420
(l)	Flight	UHD	30	YUV420
(m)	Disney World	UHD	30	YUV420
(n)	Hostel	UHD	30	YUV420
(o)	Dining Block	UHD	30	YUV420
(p)	Highway	UHD	30	YUV420
(q)	Vegetable Market	UHD	30	YUV420
(r)	Taj Mahal	UHD	30	YUV420

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

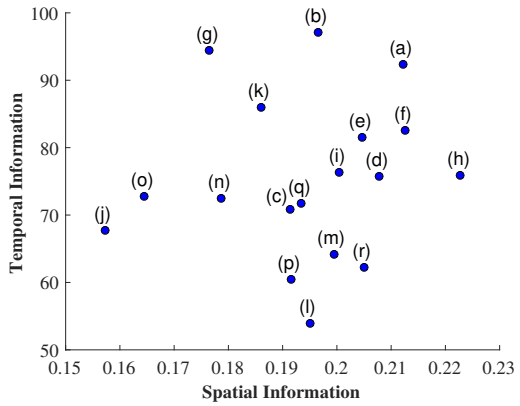


Fig. 2: Spatial and Temporal Information indices [33]. Text above each point in the plot indicates the reference video as per the Table I.

1 minute in duration and are gaming videos with a frame rate of 60 fps. The video taken from the Blender Foundation is the animated video Big Buck Bunny which is close to 10 minutes in duration with a frame rate of 24 fps. All UHD videos are close to 1 minute in duration with a uniform frame rate of 30 fps. Given that a uniform playback duration of 120 seconds is required, we next explain the construction of the reference videos from the source videos.

C. Reference Video Construction

The reference videos at each resolution are constructed by concatenating source videos having different durations. Concatenating short videos to form longer video sequences to study the time varying quality has been employed in [3]. During concatenation, it was made sure that the resolution, the frame rate and the decoded video format are preserved. For the animated video sequence Big Buck Bunny, the approximately 10 minutes duration reference video was clipped to 120 seconds from the beginning. All the reference video sequences are maintained in raw YUV video format.

Fig. 1 shows one representative video frame from each reference video sequence of the database. The first nine videos are at FHD resolution while the rest are at UHD resolution. Table I shows the nomenclature and the attributes of these videos. The videos for the study are selected so as to encompass a wide variety of content from nature, wildlife, outdoor, marine, sports, animation and gaming. The availability of a rich variety of content in the reference videos ensures that the videos spur and retain interest in subjects as they watch, thereby allowing them to respond well to the distortions. Fig. 2 shows the Spatial Information (SI) and the Temporal Information (TI) indices of the reference videos computed over the luminance component as recommended in [33].

To study the impact of rate adaptation and rebuffering on the user's QoE, we artificially introduced these distortions into the reference videos to generate distorted video sequences. The strategy for introducing rebuffering distortion is explained in the following subsection.

TABLE II: Distribution of rebuffering frequency and duration across the resolutions.

	Rebuffering Duration (seconds)					
Rebuffering Frequency (rebufferings per min.)	1	2	3	5	7	9
0.5	UHD	FHD	UHD	FHD	UHD	FHD
1	FHD	UHD	FHD	UHD	FHD	UHD
2	UHD	FHD	UHD	FHD	UHD	FHD
3	FHD	UHD	FHD	UHD	FHD	UHD
4	UHD	FHD	UHD	FHD	UHD	FHD
5	FHD	UHD	FHD	UHD	FHD	UHD

D. Rebuffering

According to the viewer experience report in [34], playback interruptions are reported to be the biggest challenge to combat as they severely degrade the user's viewing experience. To understand the impact of playback interruptions due to rebuffering, the following two aspects are considered in the study: (1) *Rebuffering Frequency* defined as the average number of rebuffering instances per minute of the playback, and (2) *Rebuffering Duration* measured in seconds.

In our study, we employ various rebuffering frequencies ranging from a minimum rebuffering frequency of 0.5 rebuffering events per minute to a maximum of 5 rebuffering events per minute. Based on an analysis over a large data of user ratings, it has been reported that the drop in the user QoE begins to saturate after a rebuffering duration of approximately 7-8 seconds [34]. Hence, in our study, we consider a maximum rebuffering duration of 9 seconds. Further, in order to examine whether the users' quality judgments are sensitive to short rebuffering durations, we employ rebuffering durations from 1 to 3 seconds with a granularity of 1 second. For rebuffering durations larger than 3 seconds, we employ a granularity of 2 seconds.

We consider a rebuffering grid of distortions consisting of different rebuffering frequencies on one axis and rebuffering durations on the other as shown in the Table II. The distortions are evenly distributed across the resolutions to ensure a fair evaluation of the QoE. This even distribution also allows for the study of distortions at a specific resolution by subsampling distorted videos from the grid at that resolution. For every distinct combination of rebuffering frequency and duration, the rebuffering events are introduced at random locations of the playback in the reference video sequence using a uniform distribution. Thus, each distorted video consists of a {rebuffering frequency, rebuffering duration} pair as its attributes. Given that 36 distorted videos need to be generated from a set of 18 reference videos, we randomly assigned 2 distorted videos per reference video. Next, we explain the strategy employed for the rate adaptation.

E. Rate Adaptation

The rate adaptation is usually preferred to be smooth in order to minimize any undesired perceptual annoyance that is caused to the user [35]. In this work, we perform rate adaptation using the strategy described as follows.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

TABLE III: Rate to resolution mapping for rate adaptation.

Rate (kbps)	Resolution (pixels)
300	640×360 (SD 360p)
600	640×360 (SD 360p)
1200	1280×720 (HD 720p)
2400	1280×720 (HD 720p)
4800	1920×1080 (FHD 1080p)
9600	3840×2160 (UHD 2160p)

TABLE IV: *FFmpeg* encoder settings.

Profile	High
Preset	Medium
Video Codec	x264
Encoder	H.264
Container Format	.mp4
Resampling Method	Lanczos

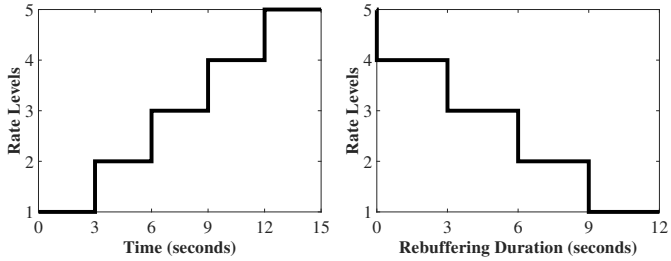
The constant rates used for encoding the reference videos to create various video representations for rate adaptation are mentioned in Table III. The selection of these rates are based on the rate-quality relation that has been observed to follow the Weber-Fechner Law (WFL) [36], [37]. The WFL states that the perceived intensity (perceived quality) is proportional to the logarithm of the magnitude of the stimulus (rate). Hence, the encoding rates considered in Table III for defining video representations are scaled by a factor of 2.

The selection of an appropriate resolution for any given encoding rate is important for rate adaptation. Encoding a video at an arbitrarily low rate while maintaining its original resolution can introduce significant encoding artifacts that are visually annoying. This effect is more pronounced when the original video resolutions are higher. Hence, in order to minimize these artifacts, higher resolution videos are generally downsampled to a suitable lower resolution before encoding at lower rates. At a given video rate, acceptable levels of video quality can be achieved only by a certain set of low resolutions. This set in general is video specific and is decided by several factors such as content, motion information, frame rate etc. Table III summarizes the rate-resolution mapping that is employed in the study. Note that while downsampling, the resolutions are chosen such that they adhere to standard formats and preserve the aspect ratio of the original resolution, which is 16:9 in our case.

Video encoding is performed using *FFmpeg* under default settings [38] as illustrated in the Table IV. After encoding at the downsampled resolution, the encoded videos are upsampled back to the original resolution so that a uniform resolution is maintained throughout the video. Resolution changes are achieved using the Lanczos resampling method [27]. The generated distorted video sequences are maintained in raw YUV video format.

The video rate considered at the beginning of playback is 2400 kbps for FHD and 4800 kbps for UHD videos, respectively. The rate adaptation can be categorized into two types: (1) Upward Rate Adaptation (URA) and (2) Downward Rate Adaptation (DRA).

1) Upward Rate Adaptation: In case no rebuffering event occurs for a duration of T_{URA} seconds, the video rate is switched up by one rate level. The rate up-switch keeps occurring every T_{URA} seconds until the maximum video rate (video representation) is reached or a rebuffering event occurs as illustrated in Fig. 3a. This period of T_{URA} seconds is allowed to mimic the scenario of checking whether the network conditions have considerably improved before switching up the rate. For instance, if a FHD video is currently being played at 2400 kbps and does not experience any rebuffering for T_{URA} seconds, then the network conditions are assumed to have improved and hence the video is switched up to 4800 kbps after T_{URA} seconds. This rate is maintained until the occurrence of a rebuffering event, in which case the following downward rate adaptation is performed.



(a) Upward Rate Adaptation (b) Downward Rate Adaptation

Fig. 3: Illustration of rate adaptation.

2) Downward Rate Adaptation: Downward rate adaptation is triggered whenever a rebuffering event occurs. The video rate chosen post rebuffering, i.e., after the playback is resumed, is determined by the rebuffering duration. The video rate is switched down by one rate level for every rebuffering duration of T_{DRA} seconds as illustrated in Fig. 3b. The DRA will be performed as long as the video is in the state of rebuffering and exits as soon as the rebuffering ends. The DRA continues until the post rebuffering playback rate becomes equal to the minimum available rate, which is 300 kbps in our case. Note that there cannot be any playback during DRA as the video is in the state of rebuffering. Thus, the DRA only facilitates the video rate to be chosen after the rebuffering ends so that the playback can be resumed at the newly adapted video rate. For instance, if a video is currently being played at 1200 kbps and enters rebuffering for a duration less than T_{DRA} seconds, then the video rate is switched down to 600 kbps. If the rebuffering duration is greater than T_{DRA} seconds, then the video rate after the playback resumes will be 300 kbps.

We have employed a uniform URA and DRA with a constant T_{URA} = T_{DRA} = 3 seconds [39]. Fig. 4 shows the frames of a distorted video from the database prior to rebuffering, during rebuffering and post rebuffering along with their respective PSNR values. A drop in the video quality because of the DRA due to a rebuffering event can be noted in terms of the decreased PSNR from Fig. 4. In the next section, we describe the subjective evaluation of the videos in the database.

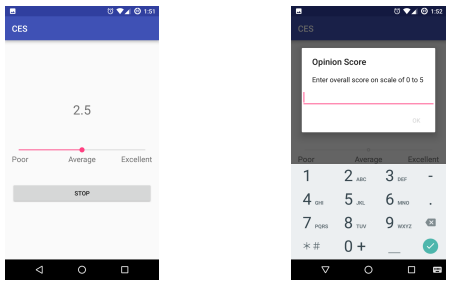
IV. Subjective Evaluation

A subjective evaluation of the videos in the proposed database was conducted at LFOVIA, IIT Hyderabad [40]. The

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

(a) Pre-rebuffering: PSNR = 40.0847 dB. (b) Rebuffering Frame (c) Post-rebuffering: PSNR = 30.6447 dB.

Fig. 4: Illustration of rebuffering and rate adaptation.



(a) Slider. (b) Overall Score.

Fig. 5: CES Android app.

Fig. 5a shows the evaluation slider used to give continuous time QoE scores in the range [0, 5]. Fig. 5b shows the window used to give the overall QoE score at the end of each video.

subjective study setup and the evaluation procedure conducted were as per ITU recommendations [41]. We employed a continuous time Single Stimulus Continuous Quality Evaluation procedure with Hidden Reference (SSCQE-HR) for evaluating videos. The SSCQE procedure involves recording the opinion scores from subjects for each video on a continuous scale with the help of a moving slider. The HR procedure involves recording the opinion scores for the reference videos along with the distorted videos without the subjects being aware of the reference videos being shown.

We developed an Android based Continuous Evaluation Slider (CES) application (app) to record the subjective scores whose user interface is shown in Fig. 5. The app consists of a slider which the subjects can move or slide continuously using their thumb in accordance with their opinion scores as they watch the video. As the slider is moved, the user ratings were continuously recorded by the app in the background in real time based on the position of the slider. The usage of CES eliminated any latency between the subjective ratings and the recorded scores. The frequency at which the rating is recorded in CES can be set according to the requirement. We set this frequency to 2 Hz, implying that there were two ratings recorded every second. The opinion scores ranged between 0 and 5 with 0 being the worst and 5 the best. The initial slider position was always set to the center with an opinion score of 2.5 before the rating process was started in order to avoid subjects getting biased towards either of the two extremes. At the end of the video, the subjects were asked to give an opinion score for the overall QoE in the range [0, 5]. Then, both the continuous QoE and the overall QoE scores were linearly scaled up to the range [0, 100] for further processing.

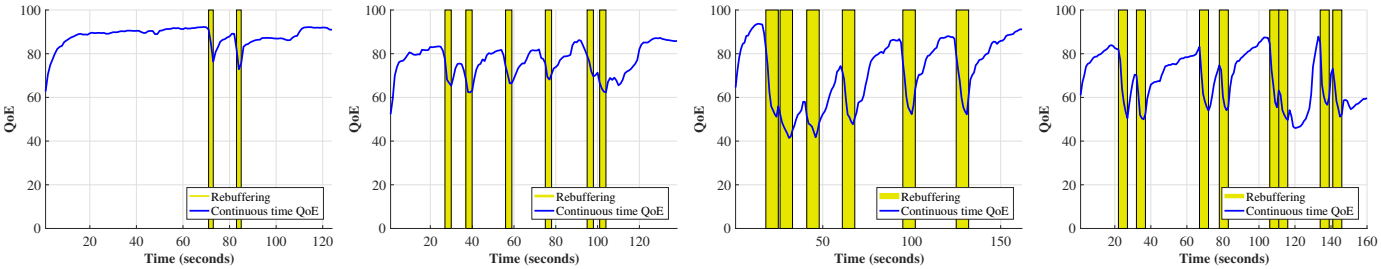
All the videos in the database were evaluated by 21 subjects. All the subjects were naïve and had not participated in any such study before. The subjects' age ranged between 20 and 40. The subjects were mostly students from IIT Hyderabad. All the subjects who participated in the study declared no vision related problems. The entire video evaluation took about 3 hours per subject. The evaluation was conducted in multiple sessions with sufficient breaks in between in order to avoid any induced bias or the effects of fatigue. For some subjects, the evaluation sessions were conducted on different days while adhering to ITU recommendations [41].

The subjects were given clear instructions on the evaluation procedure and CES app usage. They were asked to press the START button as soon as the video playback started in order to begin the recording of the opinion scores. Further, they were instructed to move the slider in accordance with the scores that they wanted to give. As soon as the video ended, they were instructed to press the STOP button. They were then asked to give an overall QoE score for the video that they just watched before proceeding to the next video.

Before the actual evaluation session started, subjects underwent a training session where a few sample videos unrelated to the videos in the proposed database were shown until they became comfortable watching videos under test conditions and got a good handle on using the CES app for evaluating videos. The training session lasted for about 10-15 minutes. In case of sessions that spanned across multiple days, the subjects were made to undergo a training routine to recall the evaluation procedure, which lasted for less than 10 minutes.

Both distorted and reference videos were arranged in a random order before they were played to the subjects. It was also ensured that no two video sequences corresponding to the same reference video sequence were played consecutively in order to avoid any induced bias or memory affecting the opinion scores when the same video content is played back-to-back. The video sequences were displayed on a 49 inch LG 4K-UHD TeleVision (TV) under standard conditions as specified by ITU [41]. Since the TV does not support the playback of raw videos, the videos were encoded at a very high rate of 100 Mbps (close to 10 times the maximum encoding rate used in the study) using *FFmpeg* before rendering them on the TV. An encoding rate of 100 Mbps ensured that the perceptual quality of the video after encoding is the same as that of the corresponding raw video, for all the video sequences in the proposed database.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology



{(r), 1, 2} {(i), 3, 3} {(h), 3, 7} {(g), 4, 5}

Fig. 6: Illustration of continuous time QoE evolution. Caption below each subfigure indicates {reference video, rebuffering frequency, rebuffering duration} attributes of the respective video where the reference video is as per the Table I.

In the next subsection, we describe the methodology employed for aggregating the subjective scores.

A. Aggregation of Subjective Scores

Given that the sampling frequency was set to 2 Hz, two scores were sampled every second. We averaged these two scores to produce one score per second. We refer to this duration of one second as one time slot. Since the recording of video scores was not synchronized with the playback, there were mismatches in terms of the number of continuous time scores collected and the video playback duration. In case the number of continuous time scores exceeded the actual duration by 3 or more seconds, the corresponding videos were noted and the subjects were re-shown the same video in a random order in another session. This situation arose due to an early START press or a late STOP press or both. The same procedure was repeated if the number of continuous time scores were less than the playback duration. This situation arose due to either a late START press or an early STOP press or both. The cases where the subjects' scores were accepted are when the number of continuous time scores are exactly equal to the playback duration (in seconds) or when the number of continuous time scores did not exceed the playback duration by more than two seconds. In case of one extra score, one score at the end was discarded, whereas, in case of two extra scores, one in the beginning and one at the end were discarded.

B. Outlier Removal

The subjective scores given by each subject were screened with an outlier removal procedure. This procedure checks whether the opinion scores given by a subject deviate drastically from the majority of the subjects' scores. We adopted the subject outlier removal procedure as specified in [41] based on 95% Confidence Interval (CI) of the continuous time QoE scores. In our study, we found 5 outlier subjects and excluded their scores from further analysis. Post outlier removal, continuous time QoE scores and overall QoE scores collected from 16 valid subjects for both reference and distorted videos were then averaged to derive continuous time QoE scores and an overall QoE score respectively for each video. Further, we derived 95% CI bounds for the averaged continuous time QoE scores.

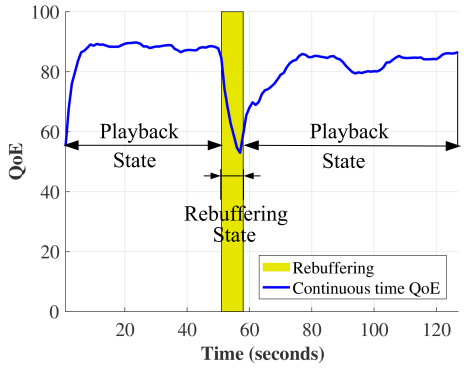


Fig. 7: Illustration of playback and rebuffering states for a distorted video with the attribute {(k), 0.5, 7}.

A framework for the continuous time QoE evaluation based on continuous time QoE scores is presented in the next section.

V. QoE Evaluation Framework and Modeling

In this section, we present a QoE evaluation framework and then propose a QoE prediction model using the framework. The continuous time QoE evolution for four distorted video sequences arbitrarily chosen from the LFOVIA-QoE database are illustrated in Fig. 6. The caption below each subfigure represents the {reference video, rebuffering frequency, rebuffering duration} attribute pair for the video illustrated. Based on the observations from the continuous time QoE plots, the QoE evolution in time can be categorized into the following video client states: (1) Playback state and (2) Rebuffering state. For one of the distorted videos in the database, these two states are illustrated in Fig. 7 and are explained as follows.

1) Playback State: This state is characterized by the periods where a video is playing back and there are no interruptions due to rebuffering. In this state, the video client performs URA as described in Section III-E if the video rate is not the highest and stays at the highest rate once reached as depicted in Fig. 7. At the highest rate, the user's QoE is expected to be high owing to the rendering of the highest rate video representation and a small variation in the QoE is mostly attributed to the video content and related characteristics.

2) Rebuffering State: This state is constituted by the periods where there are rebuffering events causing the playback to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

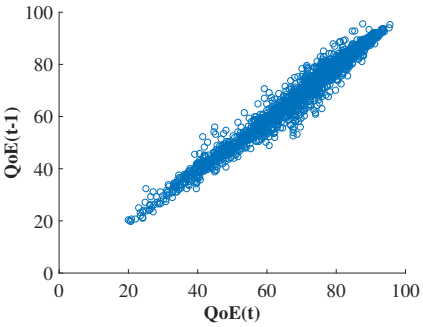


Fig. 8: Scatterplot of the current QoE versus the previous QoE: PLCC = 0.9891.

stall. A significant drop in the QoE is observed in the rebuffering state as depicted in Fig. 7. The exit of rebuffering is followed by a URA in the playback state. The URA tries to recover the fall in the QoE that occurred due to the rebuffering.

Given the distinct QoE evolution in each of the two states, the QoE evaluation can be performed distinctly in each of them. Thus, we propose a QoE evaluation framework consisting of two sub-frameworks corresponding to the two states as follows.

1. Learning based QoE evaluation in the playback state and
2. Parametric QoE evaluation in the rebuffering state.

Using this framework, we propose a model for QoE prediction that is described in detail in the following subsections.

A. Playback state: Learning based QoE Evaluation

In this subsection, we describe a learning based QoE evaluation in the playback state using SVR [42]. However, any other supervised learning approach can also be used in the proposed framework. The dataset is divided into two sets, the training set for SVR training and the test set for SVR testing. Identification of the right set of features is a crucial task for SVR learning, which can otherwise lead to poor prediction performances. The selection of features as well as the construction of feature and label vectors for training are discussed next.

1) Selection of Features: Given that there are no rebuffering events in the playback state, the current QoE is largely affected by the current time slot's video quality, also known as the STSQ [3]. Here, the time slot is as defined in Section IV-A. A scatterplot of the current time slot QoE versus the previous time slot QoE for all the distorted videos in the database is shown in Fig. 8. It is observed from Fig. 8 that there exists a high degree of correlation (measured in terms of Pearson Linear Correlation Coefficient (PLCC)) between the QoE in the current and the previous time slots. Hence, we consider the following features for training the SVR: 1) STSQ and 2) Previous time slot QoE.

The STSQ can be measured using any of the robust VQA algorithms. In our QoE prediction model, we employ Video Reduced Reference Entropic Differencing (Video-RRED) [8] as the VQA metric for measuring the current time slot video quality. This is motivated by the usage of Video-RRED for STSQ in [3]. Note that since Video-RRED is a reduced reference VQA metric, the availability of necessary features from its pristine reference video is required for computing the STSQs of any given distorted video. However, the STSQs can be precomputed and made available to facilitate continuous time QoE computation. Next, we explain the feature vector construction for SVR.

2) Construction of Feature Vector: Depending on the number of rebuffering events, there could be more than one playback state in a given training video that is noncontiguous in time. Hence, features corresponding to each playback state are collected and concatenated to form a *playback state feature vector* for each training video.

Let P^i be the number of playback states in the training video i. Let L^i_j be the length of the j^{th} playback state of the i^{th} training video expressed in terms of the number of time slots. Let $\mathbf{f}^i_{jk}$ represent the k^{th} time slot feature vector in the playback state j of the i^{th} training video. We denote the number of features in the feature vector as N_F (N_F = 2 in our case). Note that the feature vector in a given time slot comprises the STSQ of that time slot and the previous time slot's QoE. A set of feature vectors corresponding to each playback state is constructed and all such sets of the training video i are concatenated to form a *playback state feature* vector $\mathbf{F}^i$ as

$$\mathbf{F}^i = [\underbrace{\mathbf{f}^i_{11}\ \mathbf{f}^i_{12}\ \ldots\ \mathbf{f}^i_{1L^i_1}}_{\text{playback state - 1}}\ \underbrace{\mathbf{f}^i_{21}\ \mathbf{f}^i_{22}\ \ldots\ \mathbf{f}^i_{2L^i_2}}_{\text{playback state - 2}}\ \cdots\ \underbrace{\mathbf{f}^i_{P^i1}\ \mathbf{f}^i_{P^i2}\ \ldots\ \mathbf{f}^i_{P^iL^i_{P^i}}}_{\text{playback state - } P^i}], \tag{1}$$

where, the feature set vector $\mathbf{f}^i_{jk}$ is given by

$$\mathbf{f}^i_{jk} = [f^i_{jk,1}\ f^i_{jk,2}\ \cdots\ f^i_{jk,N_F}]^T.$$

Let N_{TrV} be the number of videos in the training set. The vectors $\mathbf{F}^i\ \forall\ i \in \{1, 2, \cdots, N_{TrV}\}$, given in (1), are then concatenated to form the *playback feature* vector denoted by $\mathbf{F}^{TrV}$ as follows:

$$\mathbf{F}^{TrV} = [\mathbf{F}^1\ \mathbf{F}^2\ \cdots\ \mathbf{F}^{N_{TrV}}]. \tag{2}$$

3) Construction of Label Vector: The labels used in SVR training are the continuous time QoE scores obtained from the subjective evaluation described in Section IV. The construction of the *playback state QoE* and the *playback QoE* vectors are similar to those described for the corresponding feature vectors.

Let q^i_{jk} be the k^{th} time slot's QoE score in the playback state j of the i^{th} training video. A vector of QoE scores corresponding to each playback state is constructed and all such vectors of the training video i are concatenated to form a *playback state QoE vector* $\mathbf{Q}^i$ as

$$\mathbf{Q}^i = [\underbrace{q^i_{11}\ q^i_{12}\ \ldots\ q^i_{1L^i_1}}_{\text{playback state - 1}}\ \underbrace{q^i_{21}\ q^i_{22}\ \ldots\ q^i_{2L^i_2}}_{\text{playback state - 2}}\ \cdots\ \underbrace{q^i_{P^i1}\ q^i_{P^i2}\ \ldots\ q^i_{P^iL^i_{P^i}}}_{\text{playback state - } P^i}].$$

The vectors $\mathbf{Q}^i\ \forall\ i \in \{1, 2, \cdots, N_{TrV}\}$ are then concatenated to form a *playback QoE* vector $\mathbf{Q}^{TrV}$ as

$$\mathbf{Q}^{TrV} = [\mathbf{Q}^1\ \mathbf{Q}^2\ \cdots\ \mathbf{Q}^{N_{TrV}}]. \tag{3}$$

The vectors $\mathbf{F}^{TrV}$ and $\mathbf{Q}^{TrV}$ constitute the training features and the corresponding training labels, respectively. These

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

features and labels are used for SVR training. In the next subsection, we explain the parametric QoE evaluation in the rebuffering state.

B. Rebuffering state: Parametric QoE Evaluation

We propose a parametric QoE evaluation for characterizing the drop in QoE in the rebuffering state. Given the significant fall in the QoE across all the rebuffering states of all the distorted videos, we hypothesize the QoE to fall exponentially with time. This hypothesis for the rebuffering state is motivated by the IQX hypothesis defined in [14]. Therefore, in our QoE prediction model, we present an Exponential QoE Depreciation (EQD) scheme parametrized by the parameter λ to model the fall in QoE during rebuffering. However, any other parametric modeling approach can be used to model the same in the proposed framework. In the EQD model, the QoE is assumed to depreciate exponentially with time until the exit of rebuffering. In the rebuffering state, the QoE prediction at time slot t is computed from the QoE in the previous time slot $t-1$ using the recursive relation defined as follows:

$$\text{QoE}(t) = e^{-\lambda}\text{QoE}(t-1), \tag{4}$$

where λ is the QoE depreciation parameter and $\lambda \geq 0$. From the opinion scores, it is found that there is a dependency of the parameter λ on the QoE prior to the rebuffering event. We define the following two terms for the convenience of the subsequent description.

1. *Pre-rebuffering QoE* score (Q_{PrB}): The continuous time QoE score just before the onset of rebuffering.
2. *Post-rebuffering QoE* score (Q_{PoB}): The continuous time QoE score at the end of rebuffering.

In order to obtain λ for any given Q_{PrB} and predict QoE in the rebuffering state using (4), we model the relation between λ and Q_{PrB} as the following linear function:

$$\lambda = p_1 Q_{PrB} + p_0, \tag{5}$$

where p_0 and p_1 are the linear coefficients.

To determine the coefficients p_0 and p_1, we employ the following procedure. Let B^i be the number of rebuffering states in the training video i. Let D^i_j be the rebuffering duration of j^{th} rebuffering state in the i^{th} training video. Let $Q^i_{PrB_j}$ be the pre-rebuffering QoE score of j^{th} rebuffering state in the i^{th} training video. Let q^i_{jk} represent the k^{th} time slot QoE score in the rebuffering state j of the i^{th} training video. A vector of QoE scores prefixed by its associated pre-rebuffering QoE score corresponding to each rebuffering state is constructed and all such vectors in the training video i are concatenated to form a *rebuffering state QoE* vector $\mathbf{R}^i$ as

$$\mathbf{R}^i = [\underbrace{Q^i_{PrB_1} q^i_{11} \ldots q^i_{1D^i_1}}_{\text{rebuffering state - 1}} \underbrace{Q^i_{PrB_2} q^i_{21} \ldots q^i_{2D^i_2}}_{\text{rebuffering state - 2}} \cdots \underbrace{Q^i_{PrB_{B^i}} q^i_{B^i 1} \ldots q^i_{B^i D^i_{B^i}}}_{\text{rebuffering state - } B^i}].$$

An exponential fit is performed using *nlinfit* function in MATLAB[1] to obtain the depreciation (or the decay) parameter λ^i_j for each rebuffering state j in $\mathbf{R}^i$ to form a *rebuffering state depreciation parameter* vector $\mathbf{\Lambda}^i$ as

$$\mathbf{\Lambda}^i = [\lambda^i_1 \ \lambda^i_2 \ \cdots \ \lambda^i_{B^i}]. \tag{6}$$

The corresponding pre-rebuffering QoE scores for each rebuffering state are gathered to construct a *pre-rebuffering state QoE* vector $\mathbf{Q^i_{PrB}}$ as follows:

$$\mathbf{Q^i_{PrB}} = [Q^i_{PrB1} \ Q^i_{PrB2} \ \cdots \ Q^i_{PrBB^i}]. \tag{7}$$

The vectors $\mathbf{Q^i_{PrB}}$ and $\mathbf{\Lambda}^i \ \forall \ i \in \{1, 2, \cdots, N_{TrV}\}$, given in (6) and (7), are then concatenated to form the *rebuffering depreciation parameter* and the *pre-rebuffering QoE* vectors denoted by $\mathbf{\Lambda}^{TrV}$ and $\mathbf{Q^{TrV}_{PrB}}$ respectively as follows:

$$\mathbf{\Lambda}^{TrV} = [\mathbf{\Lambda}^1 \ \mathbf{\Lambda}^2 \ \cdots \ \mathbf{\Lambda}^{N_{TrV}}] \text{ and} \tag{8}$$

$$\mathbf{Q^{TrV}_{PrB}} = [\mathbf{Q^1_{PrB}} \ \mathbf{Q^2_{PrB}} \ \cdots \ \mathbf{Q^{N_{TrV}}_{PrB}}]. \tag{9}$$

A linear regression is performed between the vectors $\mathbf{Q^{TrV}_{PrB}}$ and $\mathbf{\Lambda}^{TrV}$ using MATLAB to obtain the linear coefficients p_0 and p_1 in (5). Given that $Q_{PoB} < Q_{PrB}$, λ is always positive. Thus, a larger Q_{PrB} results in a higher λ and a steeper fall in

TABLE V: Summary of the notations used.

Symbol	Description
B^i	Number of rebuffering states in the i^{th} training video
D^i_j	Duration of the j^{th} rebuffering state of the i^{th} training video
$\mathbf{f}^i_{jk}$	k^{th} time slot feature vector in the playback state j of the i^{th} training video
$\mathbf{f}^i(t)$	Feature vector at time slot t in the playback state of the i^{th} test video
$\mathbf{F}^i$	Playback state feature vector of the i^{th} training video
$\mathbf{F}^{TrV}$	Playback feature vector for training
L^i_j	Length of the j^{th} playback state of the i^{th} training video
λ	Exponential depreciation parameter
$\mathbf{\Lambda}^i$	Rebuffering state depreciation parameter vector of the i^{th} training video
$\mathbf{\Lambda}^{TrV}$	Rebuffering depreciation parameter vector for training
N_F	Number of features in the feature set
N_{TrV}	Number of training videos
N_{TeV}	Number of test videos
P^i	Number of playback states in the i^{th} training video
q^i_{jk}	k^{th} time slot QoE score in any state (playback or rebuffering) j of the i^{th} training video
$\hat{q}^i(t)$	Predicted QoE at time slot t of the i^{th} test video
$\mathbf{Q}^i$	Playback state QoE vector of the i^{th} training video
$\mathbf{Q}^{TrV}$	Playback QoE vector for training
Q_{PrB}	Pre-rebuffering QoE score
Q_{PoB}	Post-rebuffering QoE score
$Q^i_{PrB_j}$	Pre-rebuffering QoE score prior to j^{th} rebuffering state of the i^{th} training video
$\mathbf{Q^i_{PrB}}$	Pre-rebuffering state QoE vector of the i^{th} training video
$\mathbf{Q^{TrV}_{PrB}}$	Pre-rebuffering QoE vector for training
$\mathbf{R}^i$	Rebuffering state QoE vector of the i^{th} training video

[1]MATLAB is the registered trademark of The MathWorks, Inc.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

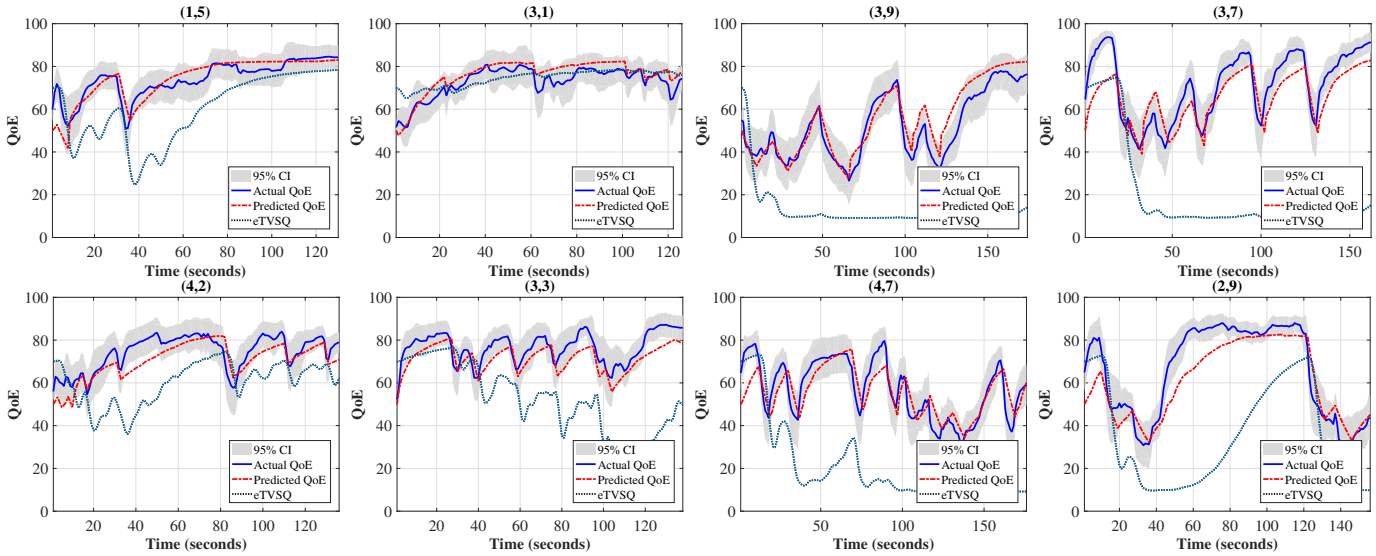


Fig. 9: QoE prediction performance of the proposed model with a training:test ratio of 80:20. The title of each of the plots indicates the attribute pair "(rebuffering frequency, rebuffering duration)" of the respective video.

the QoE. Table V shows the summary of the notations used. The QoE prediction performance of SVR and EQD models is discussed in the next section.

VI. Performance Evaluation

The 36 distorted videos in the database were divided into training and test sets with a training:test ratio of 80:20. Accordingly, 28 videos constituted the training set and 8 videos the test set respectively. The SVR was trained using the feature vector $\mathbf{F}^{TrV}$ and the corresponding label vector $\mathbf{Q}^{TrV}$ as described in (2) and (3), respectively. SVR training and testing was performed using radial basis as the kernel function [42]. The STSQ scores required for training were precomputed using Video-RRED [8]. Since the time slot has a granularity of 1 second, the STSQ scores were computed for each video segment of 1 second in duration. During the test phase, the previous QoE required in the very first time slot of the QoE prediction was initialized to a value of 50, which is the mean value of the QoE range [0, 100]. In the training phase, the linear coefficients p_1 and p_0 required to compute the depreciation parameter λ in the rebuffering state are determined using $\mathbf{\Lambda}^{TrV}$ (8) and $\mathbf{Q}_{\mathbf{PrB}}^{TrV}$ (9) as explained in Section V-B.

The SVR and EQD trained models were then tested individually on each of the videos in the test set. Let N_{TeV} be the number of test videos. Let $\mathbf{f}^i(t)$ represent the feature vector at time slot t in the playback state of the i^{th} test video. During playback, the feature vector $\mathbf{f}^i(t)$ is used as the input to the SVR for predicting the QoE $\hat{q}^i(t)$ in each time slot t as explained in Section V-A. In case of a rebuffering event, the depreciation parameter λ was computed using (5). In the rebuffering state, $\hat{q}^i(t)$ is predicted using (4). The predictions were performed individually on all test video sequences $i \in \{1, 2, \cdots, N_{TeV}\}$.

Training and testing were performed over several random permutations of the training and the test set videos. The QoE prediction performance on the test videos in one such permutation is illustrated in Fig. 9. The prediction performance of the objective metric eTVSQ is also plotted. For the assessment of the QoE prediction performance, we consider the following three measures: (1) Pearson Linear Correlation Coefficient (PLCC), (2) Spearman Rank Order Correlation Coefficient (SROCC) and (3) Root Mean Squared Error (RMSE). While PLCC and SROCC quantify how well the predicted QoE tracks the actual subjective QoE, RMSE indicates the closeness between the predicted and the actual scores. The median of each of these measures across different permutations of the training and the test set is used to quantify the performance of the proposed model.

It is observed from Fig. 9 that the SVR based model using Video-RRED performs well in predicting the QoE. It can be noted that the predicted continuous time QoE is able to track the actual continuous time QoE reasonably well at most places even though the prediction is smoothened at few places. The difference in the performance between the predicted QoE and the actual QoE can be attributed to the performance capability of STSQ, which in turn depends on the VQA method used for computing STSQs. The more robust the VQA metric, the better is the STSQ, and hence, the predicted QoE. With better VQA metrics that perform well across videos at FHD and UHD resolutions, it is expected that the QoE prediction performances can be improved. Further, it can be observed that the eTVSQ performs poorly in predicting the QoE. This is due to a severe STSQ depreciation strategy employed in [19]. It is clear from these plots that such a strategy is poor and ineffective. Thus, the prediction performance of the objective metric eTVSQ is inferior to that of the proposed model.

The QoE prediction performance of the proposed method was also analyzed for various training:test ratios as follows – 50:50 {18,18}, 60:40 {22,14} and 70:30 {25,11}. Here, the quantities a and b in {a,b} represent the number of training and test videos involved in that particular training:test

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

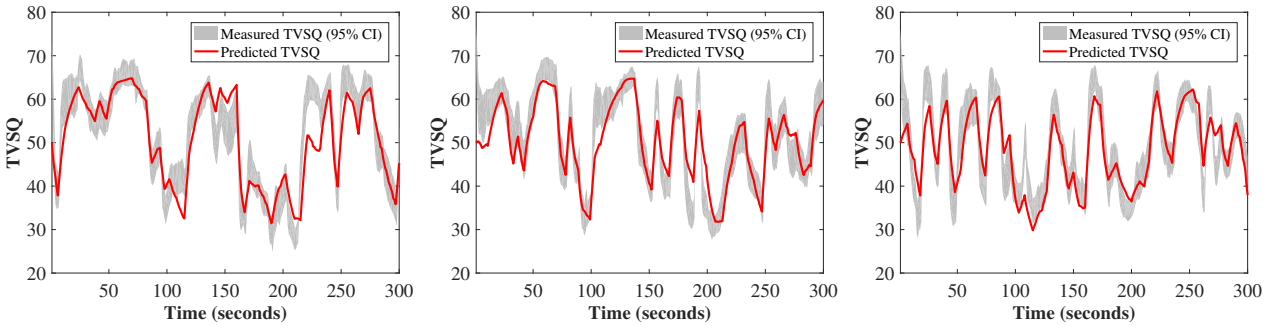


Fig. 10: TVSQ prediction performance of the proposed playback model on the LIVE QoE database [3].

TABLE VI: SVR, EQD and the overall performances of the proposed QoE model under various training:test ratios.

Training:Test Ratio	States	PLCC	SROCC	RMSE
50:50	Playback	0.7422	0.5944	10.5675
	Rebuffering	0.8235	0.8265	8.2389
	Overall	**0.8126**	**0.7190**	**10.2705**
60:40	Playback	0.7545	0.6128	10.4105
	Rebuffering	0.8293	0.8308	7.9930
	Overall	**0.8204**	**0.7283**	**10.1140**
70:30	Playback	0.7614	0.6189	10.3693
	Rebuffering	0.8331	0.8323	7.8052
	Overall	**0.8242**	**0.7323**	**10.0674**
80:20	Playback	0.7701	0.6296	10.3028
	Rebuffering	0.8376	0.8319	7.5822
	Overall	**0.8290**	**0.7377**	**9.9971**

TABLE VII: Performance of the proposed playback model on the LIVE QoE database [3].

Model	PLCC	SROCC	RMSE
Proposed (80:20)	0.8461	0.8383	5.2786

split, respectively. The individual prediction performances of SVR (playback state) and EQD (rebuffering state) schemes and the overall performance of the proposed QoE model in terms of median PLCC, SROCC and RMSE under different training:test ratios are presented in the Table VI. It must be noted from the Table VI that the performance of the proposed model is consistent across various training:test ratios that are considered.

To further demonstrate the effectiveness of the proposed method, we tested the proposed SVR model on the LIVE QoE database [3]. The LIVE QoE database consists of 15 video sequences having time varying quality distortions along with continuous time TVSQ scores. The SVR was trained and tested with a split ratio of 80:20. The prediction performance of the model on the test set videos is shown in Fig. 10. The results are tabulated in Table VII. A high prediction performance on the LIVE QoE database suggests that the proposed SVR playback model is effective for QoE prediction.

A. Objective Evaluation

The distorted videos in the database were also evaluated for QoE using some of the popular objective I/VQA and QoE metrics. Particularly, we conducted the QoE evaluation using the following metrics: PSNR [11], MS-SSIM [5], Feature Similarity Index Measure (FSIM) [43], Video-RRED [8], TVSQ [3], and eTVSQ [19]. While PSNR, MS-SSIM and FSIM are spatial quality metrics, Video-RRED is a spatio-temporal VQA metric. TVSQ and eTVSQ are QoE metrics. All quality evaluations made by the metrics were mapped using a four-parameter, monotonic logistic function as outlined by Video Quality Experts Group in [44] to predict the subjective QoE [45], [46]. Of all the metrics considered for comparison, only eTVSQ has the ability to measure both rate adaptation and rebuffering distortions. The other metrics can measure only the rate adaptation distortion. Therefore, we considered only the playback states of the distorted videos for QoE assessment using the proposed model to ensure a fair comparison. The performance of these metrics is tabulated in Table VIII. The performance of eTVSQ against the proposed QoE model is tabulated in Table IX.

TABLE VIII: Performance of popular objective metrics against the proposed model in the playback state with a training:test ratio of 80:20.

Metric	PLCC	SROCC	RMSE
PSNR [11]	0.4947	0.2467	11.0965
MS-SSIM [5]	0.6298	0.4225	9.8898
FSIM [43]	0.5313	0.3447	10.8034
Video-RRED [8]	0.4973	0.3508	11.2885
TVSQ [3]	0.1124	0.1559	12.8401
eTVSQ [19]	0.5841	0.5761	31.6943
Proposed (80:20)	**0.7701**	**0.6296**	**10.3028**

TABLE IX: Performance comparison of eTVSQ and the proposed model with a training:test ratio of 80:20.

Metric	PLCC	SROCC	RMSE
eTVSQ [19]	0.6178	0.6189	30.9696
Proposed (80:20)	**0.8290**	**0.7377**	**9.9971**

Clearly, a QoE model that yields a high PLCC and SROCC and a low RMSE is desired. It is observed from Table VIII that the popular metrics such as PSNR, MS-SSIM, FSIM do not perform as well as the proposed model in measuring the QoE in the playback state. This is because these metrics do not account for the time variation in the quality and rebuffering distortions. Although the TVSQ intends to capture the time variation in quality due to rate adaptation, it fails to predict the QoE effectively as there is no in-built mechanism to account for the distortions caused by rebuffering. This also indicates that the measurement of rebuffering distortion is vital to QoE assessment together with the rate adaptation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

In Table IX, we compare the performance of eTVSQ with the proposed QoE model over both playback and rebuffering states, jointly. The relatively high PLCC and SROCC performances and a low RMSE performance of the proposed model in comparison with eTVSQ indicates better QoE prediction using the proposed model. The proposed model is also compared with the method described in [23]. Although a quantitative comparison could not be made due to lack of ground truth continuous time subjective scores, we performed a qualitative comparison and our findings resonate with some of theirs. An important finding is the dependence of the QoE in the rebuffering state on the pre-rebuffering QoE (as described in (4)). It is also inferred from the comparison that the drop in QoE is higher in case of higher pre-rebuffering QoE as the subjects' quality expectations are also higher (as described in (5)).

The high performance of the proposed QoE model in terms of PLCC, SROCC and RMSE justifies our hypothesis that the user's QoE behavior can be distinctly modeled in each of the two states: the playback state and the rebuffering state. However, the prediction in the two states are not independent. The prediction in the rebuffering state depends on the QoE depreciation parameter, which is in-turn derived based on the pre-rebuffering QoE in the playback state before the onset of rebuffering. Similarly, the QoE prediction in the playback state after the exit of rebuffering depends on the post-rebuffering QoE. Thus, the proposed QoE model has effectively captured the joint impact of rate adaptation and rebuffering distortions on the user QoE. Further, it can be inferred that the presented continuous time QoE evaluation framework is promising for the development of future QoE prediction models.

B. Recency Effect Analysis

We investigated the impact of recently watched video segments on the overall QoE also termed as the recency effect [47]. We computed the average of the continuous time QoE scores from the end of the video playback in the backward direction with different window sizes varying from 1 second to 120 seconds. The maximum window size was limited to 120 seconds as the durations of the distorted videos are different due to varying rebuffering durations. These average QoE scores are then correlated with the subjective overall QoE scores across all the distorted videos. The PLCC plot of the same is shown in Fig. 11. It should be noted that the most recent QoE has a significantly high correlation of about 0.7. The influence of the past played content on the current QoE reduces significantly beyond 1 second of the playback in the backward direction. Thus, it can be concluded that the most recent experience has a much larger influence on the current QoE and contributes significantly to the overall QoE of the user.

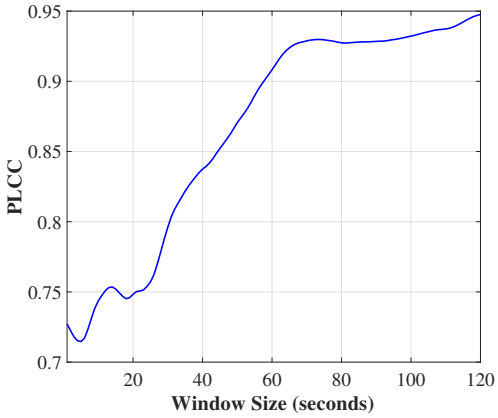


Fig. 11: Illustration of the recency effect on the overall QoE.

VII. Conclusions

We presented a QoE database consisting of FHD and UHD videos subjected to rate adaptation and rebuffering distortions with a carefully selected range of rebuffering frequencies and durations. We discussed the subjective QoE evaluation to study the effect of these distortions jointly on a continuous time scale. We presented an objective continuous time QoE evaluation framework consisting of learning based and parametrized QoE evaluation in the playback and the rebuffering states, respectively. Using the framework, we proposed a QoE prediction model comprising SVR and EQD schemes for the playback and the rebuffering states, respectively. The performance of the proposed model was demonstrated and compared against popular objective I/VQA and QoE metrics. We found that the proposed QoE model outperforms all the considered I/VQA and continuous time QoE assessment techniques. A superior performance of the proposed model suggests that the QoE behavior can be modeled distinctly in the playback and the rebuffering states even though the predictions in the two states are not independent. The proposed framework not only allows us to perform QoE prediction using the desired VQA method for SVR in the playback state, but also provides the flexibility to facilitate QoE prediction in the two states using alternative modeling techniques. Finally, the recency effect analysis indicates that the most recent experience has an influence of about 70% on the overall QoE.

Acknowledgment

The authors would like to thank Dr. Nandini Ramesh Sankar of IIT Hyderabad for the suggestions that helped in improving the presentation quality of the paper.

References

[1] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE MultiMedia*, vol. 18, pp. 62–67, Apr. 2011.
[2] "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021 white paper," Feb. 2017.
[3] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, and A. C. Bovik, "Modeling the time-varying subjective quality of http video streams with rate adaptations," *IEEE Trans. Image Process.*, vol. 23, pp. 2206–2221, May 2014.
[4] ITU, "Quality of experience requirements for iptv services," *Recommendation ITU-T G.1080*, Dec. 2008.
[5] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, vol. 2, pp. 1398–1402, Nov. 2003.
[6] K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.*, vol. 19, pp. 335–350, Feb. 2010.

[7] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *Proc. IEEE Int. Conf. on Image Processing*, pp. 2505–2508, Sep. 2011.

[8] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, pp. 684–694, Apr. 2013.

[9] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, pp. 1352–1365, Mar. 2014.

[10] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, pp. 289–300, Jan. 2016.

[11] 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, pp. 652–671, Oct. 2012.

[12] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang, "Measuring the quality of experience of http video streaming," in *Proc. IFIP/IEEE Int. Symp. on Integrated Network Management (IM 2011) and Workshops*, pp. 485–492, May 2011.

[13] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, *Data Traffic Monitoring and Analysis: From Measurement, Classification, and Anomaly Detection to Quality of Experience*, ch. Internet Video Delivery in YouTube: From Traffic Measurements to Quality of Experience, pp. 264–301. Springer Berlin Heidelberg, 2013.

[14] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Network*, vol. 24, pp. 36–41, Mar. 2010.

[15] H. Yeganeh, R. Kordasiewicz, M. Gallant, D. Ghadiyaram, and A. C. Bovik, "Delivery quality score model for internet video," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2007–2011, Oct. 2014.

[16] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A time-varying subjective quality model for mobile streaming videos with stalling events," *Proc. SPIE*, vol. 9599, pp. 959911–959911–8, 2015.

[17] D. Ghadiyaram, A. C. Bovik, H. Yeganeh, R. Kordasiewicz, and M. Gallant, "Study of the effects of stalling events on the quality of experience of mobile streaming videos," in *Proc. IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, pp. 989–993, Dec. 2014.

[18] H. Yeganeh, R. Kordasiewicz, M. Gallant, D. Ghadiyaram, and A. C. Bovik, "Live mobile stall video database," [Online]. Available: http://live.ece.utexas.edu/research/LIVEStallStudy/index.html.

[19] N. Eswara, S. Channappayya, A. Kumar, and K. Kuchi, "etvsq based video rate adaptation in cellular networks with α-fair resource allocation," in *Proc. IEEE Int. Conf. on Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, Apr. 2016.

[20] Y. Liu, S. Ci, H. Tang, Y. Ye, and J. Liu, "Qoe-oriented 3d video transcoding for mobile streaming," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 8, pp. 42:1–42:20, Oct. 2012.

[21] M. Venkataraman and M. Chatterjee, "Inferring video qoe in real time," *IEEE Network*, vol. 25, pp. 4–13, Jan. 2011.

[22] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnstrm, and N. Garca, "Perceptual quality of http adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE J. Sel. Areas Commun.*, vol. 34, pp. 2141–2153, Aug. 2016.

[23] K. Zeng, H. Yeganeh, and Z. Wang, "Quality-of-experience of streaming video: Interactions between presentation quality and playback stalling," in *Proc. IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2405–2409, Sep. 2016.

[24] ITU, "The present state of ultra-high definition television," *Rep. ITU-R BT.2246-5*, Jul. 2015.

[25] H. Nemoto, P. Hanhart, P. Korshunov, and T. Ebrahimi, "Impact of ultra high definition on visual attention," in *Proc. ACM Int. Conf. on Multimedia*, pp. 247–256, 2014.

[26] S. H. Bae, J. Kim, M. Kim, S. Cho, and J. S. Choi, "Assessments of subjective video quality on hevc-encoded 4k-uhd video for beyond-hdtv broadcasting services," *IEEE Trans. Broadcast.*, vol. 59, pp. 209–222, Jun. 2013.

[27] M. Cheon and J. S. Lee, "Objective quality comparison of 4k uhd and up-scaled 4k uhd videos," in *Proc. IEEE Int. Symp. on Multimedia (ISM)*, pp. 78–81, Dec. 2014.

[28] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The sjtu 4k video sequence dataset," in *Proc. Int. Workshop on Quality of Multimedia Experience (QoMEX)*, pp. 34–35, Jul. 2013.

[29] [Online]. Available: https://media.xiph.org/.

[30] [Online]. Available: https://peach.blender.org/.

[31] [Online]. Available: http://www.cdvl.org/.

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

[33] ITU, "Subjective video quality assessment methods for multimedia applications," *Recommendation ITU-T P.910*, Apr. 2008.

[34] [Online]. Available: http://www.conviva.com/conviva-viewer-experience-report/vxr-2015/.

[35] S. Akhshabi, A. C. Begen, and C. Dovrolis, "An experimental evaluation of rate-adaptation algorithms in adaptive streaming over http," in *Proc. of the Second Annu. ACM Conf. on Multimedia Systems*, MMSys '11, pp. 157–168, ACM, 2011.

[36] P. Reichl, S. Egger, R. Schatz, and A. D'Alconzo, "The logarithmic nature of qoe and the role of the weber-fechner law in qoe assessment," in *Proc. IEEE Int. Conf. on Communications (ICC)*, pp. 1–5, May 2010.

[37] C. Chen, X. Zhu, G. de Veciana, A. C. Bovik, and R. W. Heath, "Rate adaptation and admission control for video transmission with subjective quality constraints," *IEEE J. Sel. Topics Signal Process.*, vol. 9, pp. 22–36, Feb. 2015.

[38] [Online]. Available: http://www.ffmpeg.org/.

[39] C. Liu, I. Bouazizi, M. M. Hannuksela, and M. Gabbouj, "Rate adaptation for dynamic adaptive streaming over http in content distribution network," *Signal Processing: Image Communication*, vol. 27, no. 4, pp. 288 – 311, 2012.

[40] [Online]. Available: http://www.iith.ac.in/∼lfovia/.

[41] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation ITU-R BT. 500-13*, Jan. 2012.

[42] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, 2006.

[43] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, pp. 2378–2386, Aug. 2011.

[44] Final Report from the Video Quality Experts Group on the Validation of Objective Quality Metrics for Video Quality Assessment Phase I, Video Quality Experts Group (VQEG), 2000 [Online]. Available:www.its.bldrdoc.gov/media/8212/frtv_phase1_final_report.doc.

[45] A. K. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. C. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, pp. 587–599, Apr. 2010.

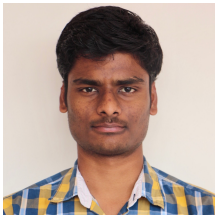
[46] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.*, vol. 19, pp. 1427–1441, Jun. 2010.

[47] D. E. Pearson, "Viewer response to time-varying video quality," in *Proc. SPIE, Human Vision and Electronic Imaging III*, vol. 3299, pp. 16–25, Jul. 1998.

Nagabhushan Eswara (S'16) received the B.E degree in Electronics and Communication Engineering from Visvesvaraya Technological University, India in 2010 and the M.Tech degree in Information and Communication Technology from the Indian Institute of Technology Jodhpur, India, in 2013. He is currently pursuing the Ph.D. degree at the Indian Institute of Technology Hyderabad, India. He is a recipient of Intel India PhD Fellowship 2014. His research interests include quality of experience analysis and modeling with applications to wireless multimedia communication systems, image and video quality assessment, machine learning and deep learning for computer vision.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCSVT.2017.2742601, IEEE Transactions on Circuits and Systems for Video Technology

Manasa K received the B.Tech. degree from Pondicherry Engineering College, India, in 2012 and the Ph. D. degree in Electrical Engineering from Indian Institute of Technology Hyderabad, India in 2017. She is currently a Postdoctoral Researcher at Conduent Labs India (Xerox Research Center). Her research interests include image and video quality assessment, statistical modeling of videos to measure quality, modeling quality of experience, quality of service, and user data (Big data) analytics. She was a recipient of the Excellence in Research Award at IIT Hyderabad in 2017 and the Above and Beyond at PEC in 2011.



Avinash Kommineni received the Bachelors degree in Electrical Engineering from Indian Institute of Technology Hyderabad, Hyderabad, India in 2016. He is currently pursuing his Masters in Computer Science and Engineering from State University of New York at Buffalo. His research interests include computer vision and machine learning.

Soumen Chakraborty received his B.Tech and M.Tech (dual) degree in electrical engineering from the Indian Institute of Technology (IIT) Kharagpur, India, in 2005. He is currently working as a Principal Engineer/Modem architect at Intel Corporation, Bengaluru, India. He joined Intel Corporation in 2014. Before joining Intel he has worked in Broadcom Corporation, Beceem Communications and Samsung R&D for 12 years in the field of Wireless systems and Image processing. His research interest includes image and video quality assessment, modem architecture, and wireless communication technologies. He is Member of the Board of studies at NIT, Trichy and represents Intel at Telecommunications Standards Development Society, India. He also serves as a Principal Investigator for PhD fellowships funded by Intel in a few Indian Universities. He holds more than 25 issued patents and has several pending patents in these fields.

Hemanth P. Sethuram received B.E. (Electronics and Communications) and M.Tech (Industrial Electronics) degrees from University of Mysore in 1993 and 1996, respectively. He is currently working as Systems Architect at Intel Corporation, India. He joined Intel Corporation in 2011. Before joining Intel, he has worked at Bharat Electronics, Philips, Stream Processors Inc. and ST Ericsson for 15 years in the fields of compression, network streaming and component frameworks for multimedia. His research interests include image and video quality assessments, systems architecture for multimedia and wireless communications systems and programming language technologies. He also serves as a Principal Investigator for PhD fellowships funded by Intel in a few Indian Universities.

Kiran Kuchi received the B.Tech. degree in electronics and communications engineering from Sri Venkateswara University College of Engineering, Tirupati, India, in 1995 and the M.S. and Ph.D. degrees in electrical engineering from The University of Texas at Arlington, Arlington, TX, USA, in 1997 and 2006, respectively.

During 2000-2008, he was with Nokia Research, Irving, TX, where he contributed to the development of Global System for Mobile Communication/EDGE, WiMax, and Long-Term Evolution systems. During 20082011, he was with the Centre of Excellence in Wireless Technology, where he led fourth-generation research and standardization efforts. He was also an adjunct faculty with the Department of Electric Engineering, Indian Institute of Technology Madras, Chennai, India. He is currently an Associate Professor with the Department of Electrical Engineering, Indian Institute of Technology Hyderabad, Hyderabad, India. He is the holder of 20 U.S. patents. His current research interests include physical-layer algorithms and the development of prototypes for fifth-generation systems.

Dr. Kuchi is the Chairman of the Wireless Systems Study Group-1 of the Telecommunications Standards Development Society, India.

Abhinav Kumar (S'04-M'14) received the BTech degree in electrical engineering and MTech degree in information and communication technology, and the PhD degree in electrical engineering from the Indian Institute of Technology, Delhi, in 2009 and 2013, respectively. From September to November, 2013, he was a research associate in the Indian Institute of Technology, Delhi. From December 2013 to November 2014, he was a postdoctoral fellow at the University of Waterloo, Canada. Since November 2014, he has been with Indian Institute of Technology Hyderabad, India, as an assistant professor. His research interests are in the different aspects of wireless communications and networking. He is a member of the IEEE.

Sumohana S. Channappayya (S'01-M'08) received the B.E. degree from the University of Mysore, India, in 1998, the M.S. degree in electrical engineering from the Arizona State University, Tempe, in 2000, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin, in 2007. He is currently Associate Professor of Electrical Engineering with IIT Hyderabad, where he directs the Laboratory for Video and Image Analysis (LFOVIA). His research interests include image and video quality assessment, multimedia communication, and biomedical imaging. He is a recipient of the Excellence in Teaching Award at IIT Hyderabad (2013).