A DYNAMIC SYSTEM MODEL OF TIME-VARYING SUBJECTIVE QUALITY OF VIDEO STREAMS OVER HTTP

Chao Chen, Lark Kwon Choi, Gustavo de Veciana, Constantine Caramanis
Robert W. Heath Jr. and Alan C. Bovik
The University of Texas at Austin
Department of Electrical and Computer Engineering
1 University Station C0803, Austin TX - 78712-0240, USA

ABSTRACT

Newly developed HTTP-based video streaming technology enables flexible rate-adaptation in varying channel conditions. The users' Quality of Experience (QoE) of rate-adaptive HTTP video streams, however, is not well understood. Therefore, designing QoE-optimized rate-adaptive video streaming algorithms remains a challenging task. An important aspect of understanding and modeling QoE is to be able to predict the up-to-the-moment subjective quality of video as it is played. We propose a dynamic system model to predict the time-varying subjective quality (TVSQ) of rate-adaptive videos that is transported over HTTP. For this purpose, we built a video database and measured TVSQ via a subjective study. A dynamic system model is developed using the database and the measured human data. We show that the proposed model can effectively predict the TVSQ of rate-adaptive videos in an online manner, which is necessary to be able to conduct QoE-optimized online rate-adaptation for HTTP-based video streaming.

***Index Terms*—** QoE, HTTP-based streaming, Time-varying subjective quality

1. INTRODUCTION

HTTP-based adaptive bitrate video streaming is an alternative to Real-Time Transport Protocol (RTP)-based methods because of its firewall-friendly property. Leading companies such as Apple, Microsoft and Adobe have embraced HTTP-based video streaming and have proposed protocols [1, 2, 3]. Furthermore, the Moving Picture Experts Group (MPEG) has issued an international standard for HTTP-based video streaming called Dynamic Adaptive Streaming over HTTP (DASH) [4].

In HTTP-based rate-adaptive streaming, a video is first partitioned into video chunks, each several seconds long. Each video chunk is then encoded into multiple representations at different bitrates. The client can select an appropriate representation of each video chunk to download, thereby adapting the downloading bitrate to its channel condition. Although HTTP-based streaming protocols provide flexibility in rate adaptation, designing rate control methods to optimize end-users' Quality of Experience (QoE) is difficult, since the relationship between the served bitrate and the users' viewing experience is not well understood.

One important indicator of QoE is the *time-varying subjective quality* (TVSQ) of the viewed videos. This is a time series or temporal record of one or more viewers' judgments of the quality of the video as it is being played and viewed.

This research was supported in part by Intel Inc. and Cisco Corp. under the VAWN program and by the National Science Fundation under grant 115.

In this paper, we propose a method to predict the TVSQ of videos streamed over HTTP. Predicting the TVSQ of a quality-varying video is challenging because the TVSQ depends on many elements of the video including spatial distortions, temporal artifacts, and variations in both of these [5, 6]. Our approach to estimate TVSQ is to begin by predicting the short-time subjective quality (STSQ) of videos. A STSQ predictor such as those in [5, 7, 8, 9, 10] operates by extracting perceptually relevant spatial and temporal features from videos then uses these to form predictions of local video quality. The basic premise of these models is that STSQ of videos is a relatively stationary phenomenon. These so-called Video Quality Assessment (VQA) models do not capture long-term variations in STSQ nor do they predict human behavioral responses to these variations.

Here, we propose a method to continuously predict TVSQ using a dynamic system model fed by an STSQ prediction engine (see Fig. 1). Quality-varying videos are partitioned into 1 second long video chunks and the STSQ of each chunk is predicted using the Video-RRED algorithm [10]. We use Video-RRED because of its excellent quality prediction performance and fast computational speed. The computed STSQs are then fed to our dynamic system model, which predicts TVSQ.

We built a database of quality varying video sequences that simulate quality fluctuations encountered in video streaming applications. We then conducted a subjective experiment to measure the TVSQs of these video sequences. We used this TVSQ database to determine the dynamic system model. Experimental results show that the proposed model reliably tracks the TVSQ of video sequences suffering from time-varying impairments. The estimated TVSQs can then be used to guide online rate-adaptation strategies towards maximizing the QoE of viewers.

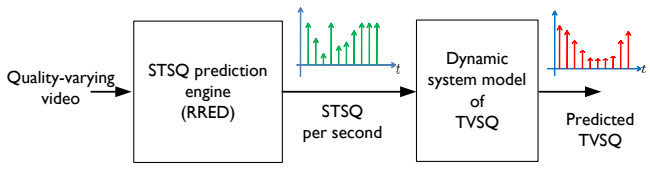

Fig. 1. Proposed paradigm for time-varying subjective quality estimation.

TVSQ estimation is an important research topic in the realm of visual quality assessment [11, 12, 13, 14, 15, 16]. Pearson *et al.* studied the relationship between STSQ and TVSQ for packet videos transmitted over ATM networks [11]. In [12], Tan *et al.* proposed an algorithm to estimate TVSQ. Its performance was evaluated on a database of three videos, on which the encoding data rates were

adapted over a slow time scale of 30–40 seconds. In [13], a first order infinite impulse response (IIR) filter was used to predict TVSQ based on per-frame distortions, which were predicted by spatial and temporal features extracted from the video. This method was shown to track the dynamics of TVSQ on low bit-rate videos. In [14], an adaptive IIR filter was proposed to model TVSQ. Since the main objective of [14] is to predict the overall subjective quality of a long video sequence using the predicted TVSQ, performance of this model was not validated against measured TVSQ. In [15], the authors studied a temporal pooling strategy that maps STSQ to the overall visual quality using a model of visual hysteresis. As an intermediate step, STSQ is first mapped to TVSQ, then the overall quality is predicted as a time-averaged TVSQ. Although this pooling strategy yields good predictions of the overall video quality, the model for TVSQ is a non-casual system, which contradicts the fact that TVSQ at a moment only depends on previous viewing experiences. Instead, their method seeks to capture the hysteresis effect in the final overall video quality prediction by incorporating the forward-time effects of the hysteresis effect. In [16], a convolutional neural network was employed to map features extracted from each video frame to TVSQ. The predicted TVSQs were shown to achieve a good correlation with measured TVSQ values on constant bitrate videos.

Unlike existing TVSQ prediction methods [11, 12, 13, 14, 15, 16], our proposed TVSQ prediction method is designed for HTTP-based video streaming. Newly proposed HTTP-based video streaming protocols like DASH provide the flexibility to adapt video bitrates over finer time-scales, e.g. 2-4 seconds, whereas prior models have mainly targeted videos on which the encoding rate is fixed [13][16] or changing slowly [12]. In [14] and [15], estimated TVSQ is used as an intermediate result in an overall video quality prediction process. The performances of these models were not validated against measured TVSQ. Towards filling this gap, we have designed and built a video quality database specifically configured to enable the development of TVSQ prediction models of HTTP-based video streams. The encoding bitrates of the videos in the new database varies randomly over time scales of several seconds to simulate quality-varying videos streamed over HTTP. The experimental results show that our new TVSQ prediction method effectively captures the TVSQ of the videos in the database.

Here, we introduce some of the key notation. The function $\mathrm{q}^{\mathrm{V}}[t]$ will denote the STSQ at the t^{th} second of a video sequence. The function $\mathrm{q}^{\mathrm{P}}[t]$ denotes the TVSQ following the playout of the first t seconds of the video. The notation $\big(\mathrm{x}\big)_{t_1:t_2}$ denotes the time series $\big(\mathrm{x}[t_1], \cdots, \mathrm{x}[t_2]\big)$.

2. SUBJECTIVE STUDY

In this section, we present the details of the database construction and the design of the subjective experiments.

2.1. Database Construction

We built a database of quality-varying videos and we measured TVSQ using a Single Stimulus Continuous Quality Evaluation (SSCQE) method [17]. We created quality-varying video sequences in four steps as follows:

1. We constructed a 250 second long reference video sequence by concatenating 25 short videos selected from the new LIVE Mobile Video Quality Assessment Database [18]. These short videos are each 10 seconds long, having spatial resolution of 720p (1280×720) and frame rate 30fps.
2. We encoded the video sequence into 21 compressed versions having different bitrates. To achieve a wide range of video quality exemplars, the encoding bitrates were chosen to range from hundreds of Kbps to several Mbps.
3. We partitioned every compressed version into one second long video chunks and predicted the Differential Mean Opinion Score (DMOS) of STSQ using the RRED index [10]. DMOS scores range from 0 to 100 and higher value indicates worse quality. To represent STSQ more naturally, so that higher numbers indicate better STSQ, we used a Reversed DMOS (RDMOS) given by $\mathrm{RDMOS} = 100 - \mathrm{DMOS}$. Broadly, a RDMOS score of less than 30 on the LIVE database indicates bad quality, while scores higher than 70 indicate good quality. In the following, we denote by $\mathrm{q}_\ell[t]$ the STSQ of the t^{th} chunk in the ℓ^{th} compressed version.
4. We constructed quality-varying videos by concatenating the video chunks selected from different compressed versions.

Next, we explain how video chunks are selected from the compressed versions to construct quality-varying videos.

2.2. Constructions of Quality-varying Videos

We seek to be able to predict the TVSQ using the STSQ. To better understand the relationship between TVSQ and previous STSQ, we need to broadly sample the space of STSQ and observe the resulted TVSQ. Thus, we constructed quality-varying videos such that the STSQs of video chunks vary randomly across time.

Specifically, we first generated a target STSQ sequence $\{\mathrm{q}^{\mathrm{tgt}}[t] : t = 1, \cdots, 250\}$. The target STSQ $\mathrm{q}^{\mathrm{tgt}}[t]$ was fixed at a constant value over every 4 second interval. Across the 4 second intervals, $\mathrm{q}^{\mathrm{tgt}}[t]$ was designed to vary as an i.i.d random process, whose marginal distribution is $\mathcal{N}(50, 16^2)$ clipped to the range $[0, 100]$. Then, we chose the t^{th} chunk of the $\ell_t^{*\mathrm{th}}$ compressed version, where $\ell_t^* = \arg\min_\ell \big|\mathrm{q}^{\mathrm{tgt}}[t] - \mathrm{q}_\ell[t]\big|$, as the t^{th} chunk of the constructed video. Therefore, for the constructed video, we have $\mathrm{q}^{\mathrm{V}}[t] = \mathrm{q}_{\ell_t^*}[t]$. Since the STSQs of the compressed videos in the database finely partition the scale of RDMOS, we have $\mathrm{q}^{\mathrm{V}}[t] \approx \mathrm{q}^{\mathrm{tgt}}[t]$.

We note that, in a subjective experiment, there is always a delay or latency between a change in video quality and a subject's response. We designed the $\mathrm{q}^{\mathrm{tgt}}[t]$ to be fixed over 4 second intervals, which are comfortably longer than the subjects' latency and short enough to simulate quality variation in adaptive video streaming. We also note that the Video-RRED algorithm is calibrated to predict the STSQ values of the LIVE video database, which is distributed as the normal distribution $\mathcal{N}(50, 10^2)$ clipped in the range of $[0, 100]$. Therefore, to sample the space of STSQ uniformly, we design $\mathrm{q}^{\mathrm{tgt}}[t]$ as obeying the same normal distribution.

We created five 250-second quality-varying videos. Together with the uncompressed video sequence, the database contains 1500 seconds of videos.

2.3. Subjective Experiment

We conducted a subjective study to measure the TVSQs of the quality-varying videos in our database. The study was completed at the LIVE Subjective Testing Lab at The University of Texas at Austin. Sixteen subjects participated in the study. One of the quality varying videos was used as the training sequence. The other four quality-varying videos were used as test sequences. The uncompressed video sequence was also included in the test as a hidden reference.

We developed a user interface for the subjective study using the Matlab XGL toolbox [19]. Video sequences were displayed to the viewers on a 58 inch Panasonic HDTV plasma monitor at a viewing distance of about 4 times the picture height. During the play of each video, a continuous scale sliding bar was shown at the screen bottom. The subject could move the bar via a mouse to feedback his/her TVSQ. The position of the bar was sampled and recorded in real time as each frame was displayed (30 fps).

2.4. Data Processing

Denote by $c_{i,j}[t]$ the TVSQ score assigned by the i^{th} subject to the t^{th} chunk of the j^{th} quality-varying video. Let $c_i^{\text{ref}}[t]$ denote the TVSQ score assigned to the reference video. We offset the impact of video content on the TVSQs using $c_{i,j}^{\text{offset}}[t] = 100 - \left(c_i^{\text{ref}}[t] - c_{i,j}[t]\right)$. Let $\mathrm{T} = 250$ be the length of the test videos and $\mathrm{N} = 4$ be the number of test videos. We computed the Z-scores of TVSQ, denoted by $z_{i,j}[t]$, using

$$m_i = \frac{1}{\mathrm{N}}\frac{1}{\mathrm{T}}\sum_{j=1}^{\mathrm{N}}\sum_{t=1}^{\mathrm{T}} c_{i,j}^{\text{offset}}[t],$$

$$\sigma_i^2 = \frac{1}{\mathrm{NT}-1}\sum_{j=1}^{\mathrm{N}}\sum_{t=1}^{\mathrm{T}}\left(c_{i,j}^{\text{offset}}[t] - m_i\right)^2,$$

and

$$z_{i,j}[t] = \frac{c_{i,j}^{\text{offset}}[t] - m_i}{\sigma_i}.$$

Then we computed $\bar{z}_j[t]$ and $\eta_j[t]$ as the average and standard deviation of $\{z_{i,j}[t],\, i = 1\cdots 16\}$, respectively. We found that the values of $\bar{z}_j[t]$ all lie in the range $[-4, 4]$. Therefore we map $\bar{z}_j[t]$ to the range $[0, 100]$ using $\mathrm{q}_j^{\mathrm{P}}[t] = \left(\frac{\bar{z}_j[t]+4}{8}\right)\cdot 100$. Correspondingly, the 95% confidence interval of $\mathrm{q}_j^{\mathrm{P}}[t]$ is computed as $\mathrm{q}_j^{\mathrm{P}}[t] \pm \epsilon_j[t]$, where $\epsilon_j[t] = \left(\frac{1.96\eta_j[t]+4}{8}\right)\cdot 100$.

In the sequel, with some abuse of notation, we suppress the subscript j and denote by $\mathrm{q}^{\mathrm{P}}[t]$ and $\epsilon[t]$ the TVSQ and its confidence interval, respectively.

3. SYSTEM MODEL IDENTIFICATION

We employ a Hammerstein-Wiener (HW) model to estimate the TVSQs of quality-varying videos. As shown in Fig. 2, the core of the HW model is the Output-Error (OE) model (see. [20]), which is intended to capture the hysteresis inherent in human behavioral responses to tempral quality variations. At the input and output of the HW model, two memoryless non-linear functions are employed to model non-linearities in the human response. The OE model is a linear dynamic system, which has the following form:

$$\mathrm{v}[t] = \sum_{d=1}^{\mathrm{d_b}} b_d\, \mathrm{u}[t-d] + \sum_{d=1}^{\mathrm{d_f}} f_d\, \mathrm{v}[t-d] \tag{1}$$

where $\mathrm{u}[t]$ and $\mathrm{v}[t]$ are the input and output of the OE model, respectively. The coefficients $\mathbf{b} = (b_1, \cdots, b_{\mathrm{d_b}})^{\mathsf{T}}$ and $\mathbf{f} = (f_1, \cdots, f_{\mathrm{d_f}})^{\mathsf{T}}$ are model parameters to be determined. We have found that if the input and output static functions are chosen as generalized sigmoid functions [21], then the proposed HW model can predict TVSQ accurately. Thus, we set the input and output functions to be

$$\mathrm{u}[t] = \beta_3 + \beta_4 \frac{1}{1+\exp\left(-(\beta_1 \mathrm{q}^{\mathrm{V}}[t] + \beta_2)\right)}, \tag{2}$$

and

$$\widehat{\mathrm{q}^{\mathrm{P}}}[t] = \gamma_3 + \gamma_4 \frac{1}{1+\exp\left(-(\gamma_1 \mathrm{v}[t] + \gamma_2)\right)}, \tag{3}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_4)^{\mathsf{T}}$ and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_4)^{\mathsf{T}}$ are model parameters and $\widehat{\mathrm{q}^{\mathrm{P}}}$ is the estimated TVSQ. We let $\boldsymbol{\theta} = \left(\mathbf{b}^{\mathsf{T}}, \mathbf{f}^{\mathsf{T}}, \boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}}\right)^{\mathsf{T}}$ be the parameters of the proposed HW model, and $\widehat{\mathrm{q}^{\mathrm{P}}}$ can be regarded as a function both of time t and parameter $\boldsymbol{\theta}$. Thus, in the following, we explicitly rewrite $\widehat{\mathrm{q}^{\mathrm{P}}}$ as $\widehat{\mathrm{q}^{\mathrm{P}}}(t, \boldsymbol{\theta})$. Next, we show how to estimate the model parameter $\boldsymbol{\theta}$.

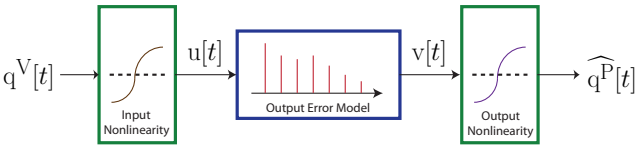

Fig. 2. Hammerstein-Wiener model for TVSQ prediction.

3.1. Model Parameter Estimation

To estimate parameter $\boldsymbol{\theta}$, we minimize the error between the measured TVSQ and the predicted TVSQ. Following [16], we use outage rate (OR) as the error metric, which is defined as the frequency that the predicted TVSQ deviates more than twice the confidence interval of the measured TVSQ. More specifically, we define OR as

$$\mathrm{J}(\boldsymbol{\theta}) = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathbb{1}\left(\left|\widehat{\mathrm{q}^{\mathrm{P}}}(t,\boldsymbol{\theta}) - \mathrm{q}^{\mathrm{P}}[t]\right| > 2\epsilon[t]\right), \tag{4}$$

where $\mathbb{1}(\cdot)$ is the indicator function. To minimize $\mathrm{J}(\boldsymbol{\theta})$, we employ a gradient-descent algorithm to determine the model parameters. The gradient of the indicator function $\mathbb{1}(|x| > 2\epsilon)$ in (4), however, is zero almost everywhere and thus the conventional gradient-descent algorithm cannot be applied. To resolve this difficulty, we approximate the indicator function $\mathbb{1}(|x| > 2\epsilon)$ with

$$\mathrm{U}_{\nu,\epsilon}(x) = \mathrm{h}(x, \nu, -2\epsilon) + (1 - \mathrm{h}(x, \nu, 2\epsilon)), \tag{5}$$

where $\mathrm{h}(x, a, b) = 1/\left(1 + \exp(-a(x+b))\right)$ is the logistic function. Note that as $\nu \to \infty$, $\mathrm{U}_{\nu,\epsilon}(x)$ converges to $\mathbb{1}\left(|x| > 2\epsilon\right)$. The error metric $\mathrm{J}(\boldsymbol{\theta})$ can thus be approximated by $\mathrm{J}(\boldsymbol{\theta}) = \lim_{\nu\to\infty} \mathrm{J}_{\nu}^{\text{apx}}(\boldsymbol{\theta})$, where

$$\mathrm{J}_{\nu}^{\text{apx}}(\boldsymbol{\theta}) = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} \mathrm{U}_{\nu,\epsilon[t]}\left(\widehat{\mathrm{q}^{\mathrm{P}}}(t,\boldsymbol{\theta}) - \mathrm{q}^{\mathrm{P}}[t]\right). \tag{6}$$

The iterative algorithm used for model parameter identification is described in Algorithm 1. In each iteration, a gradient-descent algorithm is applied to minimize $\mathrm{J}_{\nu}^{\text{apx}}(\boldsymbol{\theta})$ by moving $\boldsymbol{\theta}$ along the negative gradient of $\mathrm{J}_{\nu}^{\text{apx}}(\boldsymbol{\theta})$. At the end of each iteration, the parameter

Algorithm 1 Parameter optimization algorithm

Inputs: $\mathrm{q}^{\mathrm{V}}[t]$, $\mathrm{q}^{\mathrm{P}}[t]$, $\boldsymbol{\theta}^{(0)}$, and $\nu = 0.8$

1: **while** $\mathrm{J}\left(\boldsymbol{\theta}^{(\mathrm{i})}\right) - \mathrm{J}\left(\boldsymbol{\theta}^{(\mathrm{i}+1)}\right) \geq \delta$ **do**
2: $\quad \mathrm{i} := \mathrm{i} + 1$
3: $\quad \boldsymbol{\theta}^{(\mathrm{i}+1)} = \arg\min_{\boldsymbol{\theta}} \mathrm{J}_{\nu}^{\text{apx}}(\boldsymbol{\theta})$ via gradient-descent
4: $\quad \nu := 1.2\nu$
5: **end while**

ν is increased by a factor 1.2. The algorithm terminates when the decreases in $\mathrm{J}(\boldsymbol{\theta})$ between two iterations falls below a threshold δ. In our implementation, we set $\delta = 1 \times 10^{-5}$.

3.2. Model Order Selection

According to the parameterizations of the HW model in (1), (2), and (3), models of lower order are special cases of the model of higher order. Therefore, in principle, the higher the model order, the better performance one can achieve. A large model order, however, may result in over-fitting the data when the model parameters are learned, which could subsequently degrade the model's performance. To select an appropriate order for the HW model, we employed the Minimum Description Length (MDL) criterion, which is widely used in the realm of model identification, machine learning, and hypothesis testing [22][20]. For simplicity, we set $\mathrm{d_b} = \mathrm{d_f} - 1$ in (1) and test the models with different $\mathrm{d_f}$. The description length of a $(\mathrm{d_f} - 1, \mathrm{d_f})$-order model is defined in [20] as

$$\mathsf{L}^{\mathrm{des}}(\mathrm{d_f}) = \mathrm{J}(\boldsymbol{\theta}^*_{\mathrm{d_f}}) \left(1 + (2\mathrm{d_f} - 1)\frac{\log(\mathrm{T} - \mathrm{d_f})}{\mathrm{T} - \mathrm{d_f}}\right), \qquad (7)$$

where $\boldsymbol{\theta}^*_{\mathrm{d_f}}$ is the model parameter of the $(\mathrm{d_f} - 1, \mathrm{d_f})$-order model determined through Algorithm 1. The first multiplicative term in (7), which is defined in (4) as the OR, represents the ability of a model to describe the test data. The second multiplicative term reflects the complexity of the model. Thus, the definition of (7) balances the accuracy and complexity of the model. In Fig. 3, we plot the description lengths of the proposed models under different configurations of $\mathrm{d_f}$. It is seen that the minimum description length is achieved at $\mathrm{d_f} = 25$. Therefore, we set $\mathrm{d_b} = 24$ and $\mathrm{d_f} = 25$.

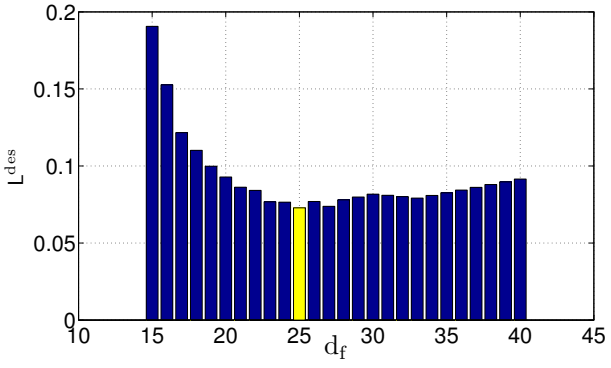

Fig. 3. Description length vs. model order.

4. PERFORMANCE EVALUATION

We employ a leave-one-out (LOO) cross-validation protocol to test the proposed TVSQ prediction model. Each time, we pick one video from the four videos of the database as the validation video and train the model parameters on the other three videos. This procedure is repeated such that each video in the database is used once as the validation video.

In Fig. 4, we plot the estimated TVSQ and the 95% confidence interval of the measured TVSQ. It is seen that the proposed model can effectively track the variation of measured TVSQ of the four quality-varying videos in our database. In Table. 1, we show the

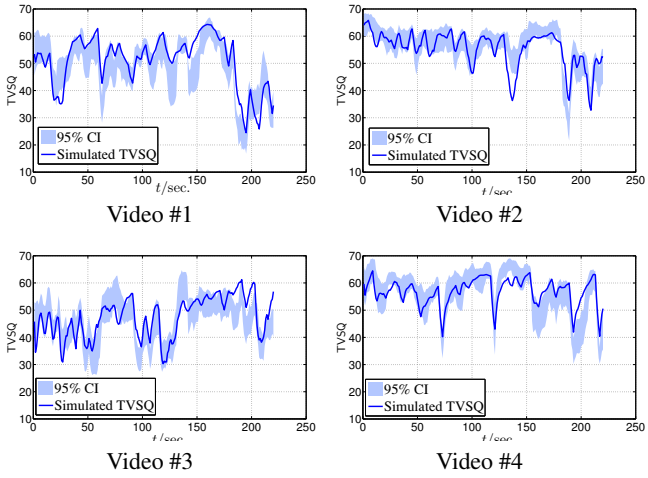

Fig. 4. The estimated TVSQ and the 95% confidence interval (CI) of the measured TVSQ.

OR, linear correlation coefficient (LCC), and Spearman's rank correlation coefficient (RCC) of the estimated TVSQ versus the measured TVSQs. As compared with the TVSQ model presented in [16], which achieves a OR of 5.6%, the proposed model achieves a lower average OR of 3.4%. The LCC of the proposed TVSQ model is 0.827, which is worse than the LCC achieved by the model presented in [16], which is 0.92. It should be noted, however, that the comparison is unfair since the model in [16] was tested on constant bitrate videos, whose quality only varies in a smaller range. Most existing rate-adaptive video streaming algorithms are designed to maximize STSQ. Using our TVSQ database, we find that the LCC and RCC between STSQ and TVSQ are 0.414 and 0.357, respectively. Clearly, simply optimizing for STSQ will not necessarily maximize the TVSQ of users. The proposed model in this paper helps to understand the TVSQ of quality-varying videos and thus has the potential to be useful for designing TVSQ-optimized video streaming algorithms.

Table 1. Performance of the proposed model on the test videos

Video	#1	#2	#3	#4
OR	2.8%	3.2%	4.8%	2.8%
LCC	0.843	0.841	0.790	0.832
RCC	0.900	0.838	0.820	0.852

5. CONCLUSION AND FUTURE WORK

We proposed a dynamic system model for on-line TVSQ prediction. The accuracy of the model is validated on a database of four long-duration video sequences. The predicted TVSQ of the model correlates well with the measured TVSQ in subjective studies. For better model identification and validation, we are going to build a larger video database and measure TVSQ on a larger set of subjects.

6. REFERENCES

[1] Microsoft Corporation, "IIS smooth streaming technical overview," http://www.microsoft.com/en-us/download/default.aspx, Sept. 2009.

[2] R. Pantos and E. W. May, "HTTP live streaming," *IETF Internet Draft, work in progress*, Mar. 2011.

[3] Adobe Systems, "HTTP dynamics streaming," http://www.adobe.com/products/hds-dynamic-streaming.html, Mar. 2012.

[4] MPEG Requirements Group, "*ISO/IEC FCD* 23001-6 Part 6: Dynamics adaptive streaming over HTTP (DASH)," http://mpeg.chiariglione.org/working_documents/mpeg-b/dash/dash-dis.zip, Jan. 2011.

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

[6] M. Zink, O. Künzel, J. Schmitt, and R. Steinmetz, "Subjective impression of variations in layer encoded videos," in *11th Int'l Wkshp Qual Serv.* Springer, 2003, pp. 137–154.

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

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

[9] P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *18th IEEE Int'l Conf. Image Process. (ICIP)*, Sept. 2011, pp. 2505 –2508.

[10] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circ. Sys. Video Technol., to appear*, 2012.

[11] D. E. Pearson, "Viewer response to time-varying video quality," *Proc. SPIE*, vol. 3299, no. 1, 1998.

[12] K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Process.*, vol. 70, pp. 279 – 294, July 1998.

[13] M. A. Masry and S. S. Hemami, "A metric for the continuous quality evaluation of video with severe distortions," *Journal of Signal Process. Image Commun.*, vol. 19, no. 2, pp. 133 – 146, Feb. 2004.

[14] M. Barkowsky, B. Eskofier, R. Bitto, J. Bialkowski, and A. Kaup, "Perceptually motivated spatial and temporal integration of pixel based video quality measures," in *Welcome to Mobile Content Quality of Experience*, Mar. 2007, pp. 1 – 7.

[15] K. Seshadrinathan and A. C. Bovik, "Temporal hysteresis model of time-varying subjective video quality," in *IEEE Int'l Conf. Acoust. Speech Signal Process.*, May. 2011.

[16] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," *IEEE Trans. Neural Net.*, vol. 17, no. 5, pp. 1316 –1327, Sept. 2006.

[17] ITU-R Recommendation BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," http://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.500-13-201201-I!!PDF-E.pdf, Jan. 2012.

[18] A. Moorthy, L. Choi, A. C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: subjective, behavioral and objective studies," *IEEE Select. Topics Signal Process.*, vol. 6, no. 6, pp. 652 –671, Oct. 2012.

[19] Center for Perceptual Systems, The University of Texas at Austin, "The XGL Toolbox," http://svi.cps.utexas.edu/software.shtml, 2012.

[20] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1986.

[21] J. A. Nelder, "The fitting of a generalization of the logistic curve," *Biometrics*, vol. 17, no. 1, pp. pp. 89–110, 1961.

[22] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Info. Theory*, vol. 44, no. 6, pp. 2743 –2760, Oct. 1998.