Continuous Prediction of Streaming Video QoE Using Dynamic Networks

Christos G. Bampis, Zhi Li, and Alan C. Bovik

***Abstract*—Streaming video data accounts for a large portion of mobile network traffic. Given the throughput and buffer limitations that currently affect mobile streaming, compression artifacts and rebuffering events commonly occur. Being able to predict the effects of these impairments on perceived video quality of experience (QoE) could lead to improved resource allocation strategies enabling the delivery of higher quality video. Toward this goal, we propose a first of a kind *continuous* QoE prediction engine. Prediction is based on a nonlinear autoregressive model with exogenous outputs. Our QoE prediction model is driven by three QoE-aware inputs: An objective measure of perceptual video quality, rebuffering-aware information, and a QoE memory descriptor that accounts for recency. We evaluate our method on a recent QoE dataset containing continuous time subjective scores.**

***Index Terms*—Subjective quality of experience (QoE), video quality assessment (VQA), video streaming.**

I. Introduction

MOBILE streaming video occupies a dominant portion of global network traffic. Since network throughput can be volatile and hard to predict, video compression artifacts and rebuffering events often occur. For example, when the available bandwidth is unable to satisfy the playout rate on the client side, the client will either ask for a video segment encoded at a lower bitrate or (if the available bandwidth is small and the client's buffer is empty) stop the playout (rebuffering). Either can lead to unpleasant losses of perceived quality of Experience (QoE). Clearly, being able to predict perceived QoE could enable the design of perceptually driven resource allocation strategies that minimize these effects.

In streaming applications, it is the client side that is best informed regarding the streaming bitrate and rebuffering events, and this is where QoE prediction is most relevant. The server side could assist this process by precalculating the video quality values during encoding, and by sending them to the client in the manifest file. The client would then make a decision, e.g., to stream at a lower bitrate, or to interrupt the playback.

While QoE prediction is easily motivated, it remains a difficult task. Modeling the perception of video distortions is a complex problem [1] that is exacerbated by a variety of time-dependent behavioral factors, such as recency [2], which significantly affects the perceived QoE [3]–[6].

A variety of retrospective and continuous time QoE prediction models have been advanced. Three broad approaches may be identified: objective video quality prediction, rebuffering evaluation, and more general models.

The first approach focuses on video distortions such as compression and packet loss. A wide variety of video quality assessment (VQA) models is available, including those that require a pristine reference for comparison, such as [7]–[16] and those that do not [17]–[21]. The second approach focuses on the effects of rebuffering events. The number, location, and frequency of rebuffering events can significantly affect perceived QoE [22]–[26]. However, these studies have not considered the combined effects of compression artifacts and rebuffering events. However, in [27]–[29], these scenarios via subjective testing were studied, while others [30], [31] used measurements of video bitrate, resolution, and frame rate, along with rebuffering event information, to objectively predict QoE. However, none of these efforts incorporated perceptual VQA models to supply visual quality predictions to their objective systems, although distortions are an important aspect of video QoE. Recently, more general QoE-aware models that use perceptual VQA for retrospective (noncontinuous) QoE prediction were proposed in [32] and [33], but this does not supply a tool that could be used for real-time bitrate decisions.

Continuous time QoE prediction is a more challenging problem that requires accounting for the instantaneous temporal effects of subjective QoE. In this direction, Chen *et al.* [4] developed a Hammerstein–Wiener (HW) model of the temporal subjective quality of HTTP video streams. Ghadiyaram *et al.* [5] have also used HW model of the effects of rebuffering on continuous time subjective QoE. However, neither of these approaches considered a combination of rebuffering and video compression artifacts.

Here, we develop a continuous time QoE prediction engine that relies on simple, but highly descriptive "QoE-aware" inputs: objective VQA, playback status information, and QoE memory descriptors. These inputs are continuously measured on videos and continuously fed into a nonlinear prediction engine expressed as a single hidden layer neural network.

II. Dataset

The recently designed LIVE-NFLX Video QoE Database [34] consists of approximately 5000 retrospective human QoE

Manuscript received February 8, 2017; revised April 12, 2017; accepted May 5, 2017. Date of publication May 18, 2017; date of current version June 8, 2017. This work was supported by Netflix, Inc. A software implementation can be found at http://live.ece.utexas.edu/research/NARX_QoE/narx_qoe_index.html. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Joao M. Ascenso. *(Corresponding author: Christos G. Bampis.)*

C. G. Bampis and A. C. Bovik are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: bampis@utexas.edu; bovik@ece.utexas.edu).

Z. Li is with Netflix, Inc., Los Gatos, CA 95032 USA (e-mail: zli@netflix.com).

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

Digital Object Identifier 10.1109/LSP.2017.2705423

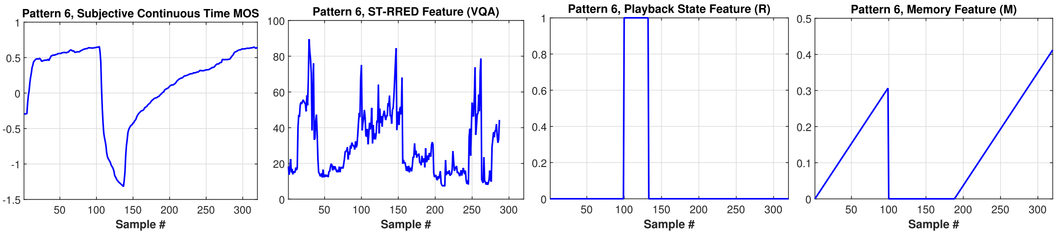


Fig. 1. Input and external continuous time QoE variables. All inputs are downsampled to match the prediction rate of NARX.

opinion scores and the same number of continuous time subjective QoE traces collected from 56 subjects that viewed video content on a mobile device (using a properly designed interface). The new database was designed to simulate a set of realistic playout/rebuffering patterns based on a simple available bandwidth model. It contains 14 diverse contents and 8 patterns (per content). We briefly describe the three types of playout patterns: constant encoding at 500 kb/s (#0) and 250 kb/s (#2), adaptive rate drops at 66 kb/s (#4) and 100 kb/s (#7), and mixtures of rebuffering events and compressed bitrate patterns such as constant encoding interrupted by rebuffering once (#1, #3) or twice (#5) and adaptive rate drops with rebuffering (#6). These playout patterns were designed using a bandwidth usage equalization model to reflect tradeoffs in practical adaptive streaming scenarios. We refer the interested reader to [34] for more details.

III. Nonlinear Autoregressive (NAR) Model

Our goal is to design a predictive engine that is able to efficiently process a nonlinear aggregate of subjective QoE measurements as inputs, including video quality during intervals of normal playback, rebuffering traces, and memory of prior events affecting QoE (recency). The nonlinear autoregressive with exogenous variables (NARX) model [35], [36] is an excellent choice for this task: It nonlinearly combines each of the inputs in an autoregressive fashion as

$$\mathbf{y}_t = f(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_{t-d_y}, \mathbf{u}_t, \mathbf{u}_{t-1}, \mathbf{u}_{t-2}, ..., \mathbf{u}_{t-d_u}) \quad (1)$$

where $f(.)$ is a nonlinear function of previous inputs $\{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_{t-d_y}\}$, previous (and current) external variables $\{\mathbf{u}_t, \mathbf{u}_{t-1}, \mathbf{u}_{t-2}, ..., \mathbf{u}_{t-d_u}\}$, d_y is the number of lags in the input, and d_u is the number of lags in the external variables.

Here, we take $\mathbf{y}_t = y_t$ to be the subjective QoE prediction at time t, and $\mathbf{u}_t = [u_{1t}\ u_{2t}\ u_{3t}]^\top$ to be a column vector containing the values of three external variables at time t:

1) *VQA*: the value u_{1t} of an objective video quality prediction at time t. Any high-performance VQA method may be used;
2) *R*: the playback status u_{2t} of the client at time t: 1 for rebuffering and 0 for normal playback;
3) *M*: the time u_{3t} that has elapsed since the last video impairment (rebuffering or bitrate drop) occurred at time t. We normalize u_{3t} by the video duration.

Examples of the external variables are shown in Fig. 1. Note that for y_t we use the z-scored continuous time mean opinion scores (MOS) (which compensate for different uses of the scale by each subject) after rejection of outlying subjects.

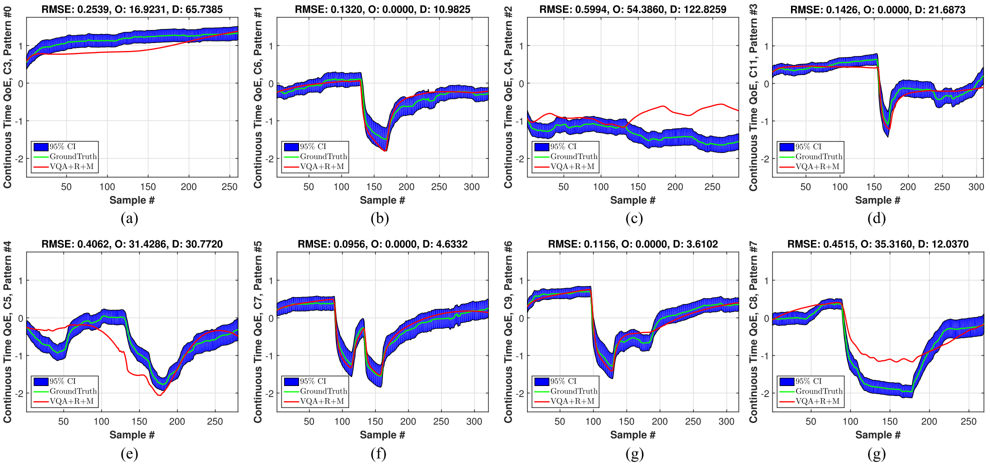
The autoregressive memory of NARX allows it to account for recency: the current QoE score depends on recent past measurements. The external variables $\mathbf{u}_t$ allow NARX to reflect present (and past) video quality, the effects of rebuffering on perceived QoE, and longer term memory effects. If no external variables are used, NARX degenerates to a NAR model $\mathbf{y}_t = f(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ..., \mathbf{y}_{t-d_y})$, where the current input is a nonlinear function of inputs within a finite window in the past (d_y). An exponential regression approach was used in [37] to model the memory of web QoE events, while we instead adopt an autoregressive neural network approach. One important challenge that might be encountered when using autoregressive models for real-time QoE prediction is that prediction errors may be propagated or amplified when the predicted outputs are fed back to the prediction engine.

IV. Application Experiments

We learned and evaluated the NARX QoE prediction engine on the new LIVE-NFLX QoE database. To reduce content and pattern dependencies, we divided the 14 contents into two disjoint sets: a training and a testing set containing nonoverlapping contents. To also eliminate pattern dependencies as much as possible, we applied the following strategy. Let j index the videos in the database, i.e., $j \in [1, ..., 112]$. For each j, we excluded all other videos having either the same content or the same pattern as j, defining those videos to be the jth training set, while the jth test set contains only the jth video. Therefore, we created 112 train and test sets; one for each video. Since the LIVE-NFLX QoE database contains 14 contents and 8 playout patterns per content, this implies $(14-1)(8-1) = 91$ training videos for each of the 112 testing videos. Of course, when deploying the NARX QoE model in the more general setting, it would be necessary to train it on the entire database.

For each train–test combination, we applied the NARX prediction engine to predict continuous time subjective QoE. However, evaluating the performance of each model on a test video is not trivial: Measuring the similarity between the two time series associated with each test video (predicted and ground truth QoE) depends on the type of performance measure that is used. For retrospective QoE evaluation, we can simply use the linear correlation coefficient or the Spearman rank order correlation coefficient, which measures the degree of monotonicity between two sets of measurements. However, computing the correlation between two QoE-related time series is a more difficult proposition, since we are interested in capturing the correct range of the subjective scores and achieving temporal alignment between the two time series. Therefore, following [4], we used the outage rate, the dynamic time warping (DTW) distance [38], and root-mean-square error (RMSE) for our comparisons.

In NARX, the nonlinear function $f(.)$ is approximated by a multilayer perceptron. Since the choice of NARX architecture (number of hidden layers, nodes per layer, d_y and d_u) can lead to variable results, we applied the following simple and effective design: we used a single hidden layer network with 5, 8, or 10 hidden nodes. We empirically fixed $d_y = d_u = 15$ and



(a) (b) (c) (d)

(e) (f) (g) (g)

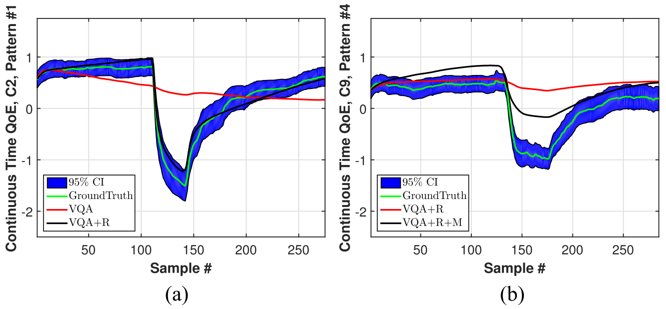
Fig. 2. Predictions on all eight patterns when ST-RRED is used as the video quality measure. C is the content index, O is the outage percentage, D is the DTW distance, and CI is the confidence interval.

divided the training set into two subsets: one for training and one for validation, to determine the best network architecture (number of nodes in the hidden layer) in terms of RMSE. Then, we trained on the whole training set and tested it on each test video sequence. Different parameters and test sets may yield different results; in practice, this cross-validation approach is sufficient to train such a predictor. Since the initialization of the network can also affect results, we repeated the training process five times and averaged the computed error metrics. To speed up the training/testing process, we designed the NARX model to predict one value every 0.25 s. We used the Levenberg–Marquardt algorithm for training.

A. Qualitative Experiments

We begin by visualizing the outputs of the proposed dynamic network. First, consider 8 of the 14 contents from the LIVE-NFLX Video QoE Dataset and the (closed loop) predictions for these 8 patterns, as shown in Fig. 2. These contents cover diverse spatiotemporal complexities, e.g., C5: a scene that contains a water scene and that is harder to encode, and C3: a slow moving scene of a human dialogue. Clearly, the prediction quality varies with the playout pattern. For example, in playout patterns #1, #3, #5, and #6 [see Fig. 2(b), (d), (f), and (g)] where there is at least one rebuffering event, the proposed model was able to closely capture the effects of rebuffering on subjective QoE. Since rebuffering is always unpleasant and obvious to subjects, the external continuous variable *R* capturing the playback status effectively describes the occurrence of rebuffering.

By contrast, when applied on videos without rebuffering, the prediction quality may vary depending on the content and the type of playout pattern. For the ElFuente Lake sequence (also denoted by C5), there are two segments where the encoding scheme was observed to greatly affect subjective QoE as shown in Fig. 2(e): at the beginning of the video (the fountain scene of high spatial complexity), and the adaptive bitrate drop in the middle of the video sequence. While the proposed predictor was able to follow the drop in subjective QoE and the overall trend, it did not capture the first drop in the QoE. This might also be partly explained by the challenging content. Similar to #4, the example of #7 [see Fig. 2(h)] shows that while the predicted QoE follows the correct trend, the subjective drop was underestimated. Since the learned QoE predictor uses the temporal VQA scores internally, this shows that even high performance VQA models are not always able to capture the perceptual effects of realistic bitrate drops.



(a) (b)

Fig. 3. Effect of (a) rebuffering *R* and (b) memory *M* external variables. C is the content index.

Since the proposed method uses external variables to enhance prediction performance, it is important to understand the contribution of each continuous variable. Fig. 3 plots the contributions of the *R* and *M* external variables. In Fig. 3(a), there is rebuffering; hence, using only the *VQA* external variable cannot account for the effects of rebuffering on perceived QoE. Using the external variable *R* greatly improves the prediction result. In Fig. 3(b), the memory input helps to capture the dynamic rate drop in the middle of the sequence.

B. Quantitative Experiments

We now move to quantitative analysis of the experiment outcomes. First, we examined whether the combination of the *VQA*, *R,* and *M* external variables led to improved prediction performance. We selected various VQA models, including PSNR, SSIM [7], MS-SSIM [10], NIQE [39], VMAF [40], and ST-RRED [16] and show the results of applying them on the LIVE-NFLX Database in Table I. We also experimented with VIIDEO [21], but were not satisfied with the results. While powerful and well-proven full-reference (FR) perceptual VQA models can be used on the server side if it implements VQA calculation, no-reference (NR) VQA models would be required at the client side or other flexible application context. Of course, NR VQA models are not yet as developed or successful as FR VQA models. Clearly, ST-RRED was the best performing VQA

TABLE I
MEDIAN PERFORMANCE METRICS FOR VARIOUS VQA MODELS AND FEATURE SETS ON ALL 112 TEST SEQUENCES

External Variables	*VQA*			*VQA+R*			*VQA+R+M*		
Model/Metric	RMSE	outage %	DTW	RMSE	outage %	DTW	RMSE	outage %	DTW
PSNR	0.6048	44.9480	67.1964	0.3700	26.7594	47.2307	0.3149	19.2988	25.9719
SSIM [7]	0.4850	35.1055	53.9253	0.3256	20.6029	**30.9463**	*0.2575*	*14.2081*	*22.9150*
MS-SSIM [10]	0.5093	38.6606	53.9536	**0.3189**	20.8699	34.6884	0.3326	22.4458	24.3942
NIQE [39]	0.6335	50.9098	61.3503	0.3972	35.7967	45.6234	0.3557	23.4994	31.7299
VMAF [40]	0.5455	45.2282	54.5198	0.3450	24.3133	34.3212	0.3041	16.3525	24.3822
STRRED [16]	**0.4467**	**29.3912**	**40.3229**	0.3245	**17.7055**	32.7867	0.2685	14.3412	23.3444

The best result per feature set is in boldface; the best result overall is in boldface and italic font.

TABLE II
COMPARISON WITH THE HW MODEL ON ONLY VIDEOS SUFFERING FROM BITRATE-RELATED IMPAIRMENTS

Model/Metric	RMSE	outage %	DTW
HW [4]	0.4179	31.7281	41.4905
NARX (*VQA*)	0.3745	29.9980	44.5552

TABLE III
MEDIAN PERFORMANCE ACROSS PLAYOUT PATTERNS WHEN USING EXTERNAL VARIABLES *VQA+R+M* AND ST-RRED

Pattern #	RMSE	outage %	DTW
0	0.7608	53.8162	81.7470
1	0.1850	3.0023	9.0227
2	0.2057	14.1298	69.0217
3	0.2518	12.1916	20.7408
4	0.4201	25.8960	36.8139
5	0.2079	1.9285	11.6585
6	0.2116	5.3340	9.3749
7	0.3571	25.3791	21.7034

model when only considering the *VQA* external variable. NIQE was the worst performer, but it is a frame-based NR model. The use of the *R* external variable greatly improved the prediction results, while the combination of *VQA+R+M* performed the best for each quality model. SSIM and ST-RRED demonstrated the best prediction results.

We also compared our approach with the dynamic system proposed in [4] (HW). The HW implementation is not suitable for videos of different durations, so we relied on the System Identification Toolbox in MATLAB, which allows for versatility in the number and duration of the inputs. We applied input and output nonlinearities using a sigmoidal network with ten neurons. The model parameters of the linear block were selected using the same validation scheme as in the NARX case. A drawback of the HW method is that it is not applicable to videos suffering from rebuffering, hence we trained NARX using the VQA input only and report results only on videos that are impaired by bitrate-related impairments. As shown in Table II, the NARX architecture yielded better performance than the HW model in terms of both RMSE and outage rate. The DTW distance was larger for NARX; but its purpose is to capture temporal trends rather than precision, and is used as a complement to the other metrics.

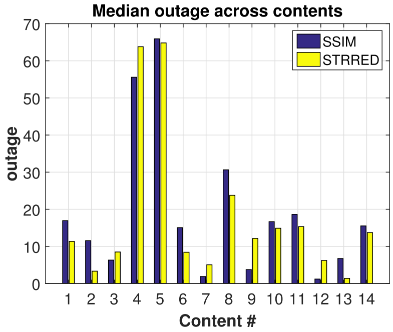


Fig. 4. Median outage % across all 14 contents in the database.

As mentioned earlier, the performance of the proposed model may vary across different playout patterns. To investigate this claim in a different light, Table III shows the median results per playout pattern. Observe that for pattern #0, the performance dropped considerably, which may be due to the fact that of the three external variables, only the VQA values were nonzero. Notably, VQA models are not designed for long-term quality prediction; hence, predictors relying only on VQA inputs may suffer in performance. Also, the performance on patterns without rebuffering (#0, #2, #4, and #7) was relatively worse, which again highlights the benefit of deploying a high performance video quality algorithm. This could also be due to error propagation when the NARX predictor is applied. We also investigated the "per content" performance of our proposed model. Naturally, the differing spatio-temporal complexities of the video contents could lead to variations in prediction performance. Fig. 4 shows the per content behavior of the continuous time QoE prediction model. For some contents, such as #4, #5, and #8, the outage rate was considerably higher, while for others such as #3 and #12 it was much lower. Going forward, it will be of great interest to account for, and ameliorate content-dependence.

V. FUTURE WORK

We envision building larger and more accurate continuous time models, e.g., by investigating the effects of the underlying parameters, extending this work to include other QoE-relevant features or by using other potentially powerful learning systems, such as recurrent neural networks. Also, it would be interesting to incorporate a moving average component to deal with measurement noise (i.e., a NARMAX model [41]), which could help address the effects of error propagation. To reduce any potential risks of overfitting and parameter sensitivity, aggregating multiple prediction models [42] or training/testing on multiple databases could also prove beneficial.

ACKNOWLEDGMENT

The authors would like to thank A. K. Moorthy, I. Katsavounidis, and A. Aaron for their help in designing the LIVE-NFLX Video QoE Database.

REFERENCES

[1] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.

[2] D. S. Hands and S. Avons, "Recency and duration neglect in subjective assessment of television picture quality," *Appl. Cogn. Psychol.*, vol. 15, no. 6, pp. 639–657, 2001.

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

[4] 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, no. 5, pp. 2206–2221, May 2014.

[5] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A time-varying subjective quality model for mobile streaming videos with stalling events," in *Proc. SPIE Opt. Eng.+ Appl.*, 2015, pp. 959 911–959 918.

[6] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual quality of http adaptive streaming strategies: Cross-experimental analysis of multi-laboratory and crowdsourced subjective studies," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2141–2153, Aug. 2016.

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

[8] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process., Image Commun.*, vol. 19, no. 2, pp. 121–132, 2004.

[9] K. Seshadrinathan and A. C. Bovik, "A structural similarity metric for video based on motion models," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, USA, Jun. 2007, pp. 869–872.

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

[11] M. H. Pinson, L. K. Choi, and A. C. Bovik, "Temporal video quality model accounting for variable frame delay distortions," *IEEE Trans. Broadcast.*, vol. 60, no. 4, pp. 637–649, Dec. 2014.

[12] 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.

[13] 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. Image Process.*, 2011, pp. 2505–2508.

[14] J. Y. Lin, T.-J. Liu, E. C.-H. Wu, and C.-C. J. Kuo, "A fusion-based video quality assessment (FVQA) index," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2014, pp. 1–5.

[15] K. Manasa and S. S. Channappayya, "An optical flow-based full reference video quality assessment algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2480–2492, Jun. 2016.

[16] R. Soundararajan and A. C. Bovik, "Video quality assessment by reduced reference spatio-temporal entropic differencing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684–694, Apr. 2013.

[17] K.-C. Yang, C. C. Guest, K. El-Maleh, and P. K. Das, "Perceptual temporal quality metric for compressed video," *IEEE Trans. Multimedia*, vol. 9, no. 7, pp. 1528–1535, Nov. 2007.

[18] F. Yang, S. Wan, Y. Chang, and H. R. Wu, "A novel objective no-reference metric for digital video quality assessment," *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 685–688, Oct. 2005.

[19] Y. Kawayoke and Y. Horita, "NR objective continuous video quality assessment model based on frame quality measure," in *Proc. IEEE Int. Conf. Image Process.*, 2008, pp. 385–388.

[20] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind prediction of natural video quality," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1352–1365, Mar. 2014.

[21] A. Mittal, M. A. Saad, and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 289–300, Jan. 2016.

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

[23] 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. Signal Inf. Process.*, 2014, pp. 989–993.

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

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

[26] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the quality of experience of HTTP video streaming," in *Proc. 12th IFIP/IEEE Int. Symp. Integr. Netw. Manage. Workshops*, 2011, pp. 485–492.

[27] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hobfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, Jan.–Mar. 2015.

[28] M.-N. Garcia *et al.*, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. 2014 6th Int. Workshop Quality Multimedia Experience*, 2014, pp. 141–146.

[29] W. Robitza, M. N. Garcia, and A. Raake, "At home in the lab: Assessing audiovisual quality of HTTP-based adaptive streaming with an immersive test paradigm," in *Proc. 2015 7th Int. Workshop Quality Multimedia Experience*, 2015, pp. 1–6.

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

[31] I.-T. P.1201, "Parametric non-intrusive assessment of audiovisual media streaming quality. Amendment 2: New Appendix III – Use of ITU-T P.1201 for non-adaptive, progressive download type media streaming," Dec. 2013.

[32] Z. Duanmu, Z. Kai, K. Ma, A. Rehman, and Z. Wang, "A quality-of-experience index for streaming video," *IEEE J. Sel. Topics. Signal Process.*, vol. 11, no. 1, pp. 154–166, Feb. 2016.

[33] C. G. Bampis and A. C. Bovik, "Learning to predict streaming video QoE: Distortions, rebuffering and memory," *Trans. Image Process.*, submitted for publication.

[34] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Temporal effects on subjective video quality of experience," *Trans. Image Process.*, to be published.

[35] T. Lin, B. G. Horne, P. Tino, and C. L. Giles, "Learning long-term dependencies in NARX recurrent neural networks," *IEEE Trans. Neural Netw.*, vol. 7, no. 6, pp. 1329–1338, 1996.

[36] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent NARX neural networks," *IEEE Trans. Syst., Man, Cybern.*, vol. 27, no. 2, pp. 208–215, Apr. 1997.

[37] T. Hoßfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, and M. Fiedler, "The memory effect and its implications on web QoE modelling," in *Proc. 2011 23rd Int. Teletraffic Congr.*, 2011, pp. 103–110.

[38] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. 3rd Int. Conf. Knowl. Discovery Data Min. Workshop*, Seattle, WA, USA, 1994, pp. 359–370.

[39] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[40] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric." 2016. [Online]. Available: http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html

[41] S. Chen and S. Billings, "Representations of non-linear systems: The narmax model," *Int. J. Control*, vol. 49, no. 3, pp. 1013–1032, 1989.

[42] C. G. Bampis, Z. Li, I. Katsavounidis, and A. C. Bovik, "Recurrent and dynamic networks that predict streaming video quality of experience," *IEEE Trans. Image Process.*, under review.