IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 28, 2019, accepted May 20, 2019, date of publication May 28, 2019, date of current version June 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2919610

Continuous Prediction for Quality of Experience in Wireless Video Streaming

WENJUAN SHI[1,2], (Member, IEEE), YANJING SUN[1], (Member, IEEE), AND JINQIU PAN[3], (Student Member, IEEE)

[1]College of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China
[2]College of New Energy and Electronic Engineering, Yancheng Teachers University, Yancheng 224007, China
[3]College of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Corresponding authors: Wenjuan Shi (shiwj@yctu.edu.cn) and Yanjing Sun (yjsun@cumt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771417, Grant 51734009, and Grant 51804304, in part by the National Key Research and Development Program under Grant 2016YFC0801403, and in part by the Natural Science Foundation of Jiangsu Province of China under Grant BK20180640.

ABSTRACT Due to the rapid development of communication technologies, the requirement of mobile video streaming services is extremely increased in recent years. However, the bandwidth limitation of the wireless network often causes video impairments, such as compression artifacts and rebuffering event, when users are watching online videos. Hence, this problem often causes the reduction of quality of experience (QoE). Predicting the QoE can provide a reference to improve resource allocation strategies, accordingly providing users with a higher quality of video streaming services. In order to predict the impact of video impairment, continuous prediction for the QoE in wireless video streaming is proposed. The input of the predicted model consists of three vectors that characterize frame quality, the state of rebuffering events, and memory effect, respectively, while the output consists of continuous predicted the QoE. The predicted model uses a block-structured nonlinear Hammerstein-Wiener model. The experimental results confirm that our proposed model can effectively predict the continuous QoE for wireless video streaming.

INDEX TERMS Quality of experience (QoE), continuous prediction, wireless video streaming, bitrate drop, rebuffering event.

I. INTRODUCTION

With the rapid development of communication technologies, mobile video has gradually become mainstream business of streaming media for various communication operators and content providers. Users expect to be able to view high-quality video anytime and anywhere through mobile devices, such as phones and tablets. Due to the dynamic variation characteristics of wireless channel, the network throughput is apt to change and difficult to predict, which causes events such as dynamic changes in video bitrate and interruption of video playing, affecting the subjective experience of users. In order to improve the competitiveness of communication operators and content providers in mobile video services and to measure the performance of mobile video services, it is necessary to monitor video quality for end-users, to predict the video quality, and to provide a reference for evaluating performance of bitrate control strategies in real time.

In the course of video playing, due to the too little data in the buffer, the video will be paused and wait for new data to fill the receiving buffer, which is called refuffering event. Ghadiyaram *et al.* [1] had shown that frequent rebuffering events (RE) can cause viewers to abandon watching videos on mobile devices. In addition, compared with video playing clarity, end users are more sensitive to the fluency of videos, therefore ensuring smooth playout of videos can effectively enhance user experience.

In order to adapt to the dynamic changes of network bandwidth in real time, many state-of-the-art techniques have been developed. In Refs. [2]–[4], HTTP-based adaptive streaming (HAS) is proposed, which divides streaming videos into small video clips, encodes each small video clip into different bitrates and resolutions and expresses it in various quality levels; then select an appropriate bitrate for any video clip according to the estimated network condition or buffer capac-

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaoqing Pan.

ity, try to avoid the occurrence of rebuffering events, and ensure video playing fluently. Since HAS relies on transfer control protocol (TCP), in order to ensure reliable transmission of data packets, TCP numbers each data packet, which can ensure terminal to receive data packets orderly. The above adaptive protocol attempts to make video playing fluently by reducing frequency and times of rebuffering, and reduce the incidence of low-quality videos at the same time. But it causes frequent switch of video quality of the terminal, thereby significantly affecting quality of experience (QoE) [5]–[7]. In the applications of streaming video, QoE of end-users is the ultimate standard for measuring video quality. Accurate and real-time prediction of QoE can help network optimize resource allocation strategies to balance resource allocation and user satisfaction in unstable wireless network conditions.

In order to study the impact of low bitrate, rate change and RE on QoE of mobile terminals, this paper considers continuous prediction for QoE as a time series prediction problem, and then analyzes factors affecting continuous subjective QoE, and adopts frame quality (FQ), characteristics of RE and human memory effects (ME) which are related to subjective QoE perception to establish a predictive model for mobile terminal QoE, implementing accurate prediction of user experience and providing a reference for online evaluation of video stream control strategy performance.

The remainder of this paper is arranged as follows. In Section II, the related research of the proposed method is introduced. Section III analyzes the factors influencing on QoE. Section IV proposes QoE prediction model. We evaluate the performances of the prediction model in section V and conclude our work in Section VI.

II. RELATED RESEARCH

The goal of studying QoE is to design a model that can accurately and automatically perceive subjective experience of users, and further effectively solve the problem of resource allocation, thereby ensuring visual satisfaction of users. The existing QoE models can be divided into QoE retrospective prediction models and QoE continuous prediction models [8].

QoE retrospective models measure overall QoE of videos by only one score, and use Video Quality Assessment (VQA) method to calculate video quality. According to the dependence on the original video, VQA method is divided into full reference (FR) [9]–[11], reduced reference (RR) [12]–[14] and none reference (NR) [15]–[17] VQA model.

Even if different video content is encoded with the same bitrate, different quality of videos are still produced, and most of existing QoE retrospective models do not consider the interaction between video quality and RE. In order to solve the above problems, for video streaming media, Duanmu *et al*. [18] proposed a QoE prediction method based on HTTP adaptive stream (Streaming QoE Index, SQI) by studying subjective response of humans to video compression coding, initial buffering and video buffering during videos playing.

The above VQA methods finally measure video quality or quality degradation degree by only one score. Although it has high consistency with subjective perception, it cannot describe the situation of video quality variation affected by events such as rebuffering and bitrate changing during video playing.

In Ref. [19]–[21], the effects of bitrate changing and RE on video quality and QoE have been studied during video playing online, and a continuous QoE prediction model is proposed. Aiming at the influence of bitrate changes on QoE during video playing online, Chen *et al*. [19] proposed a dynamic system model based on Hammerstein-Wiener, which is used to predict subjective quality of bitrate adaptive video. In order to study the impact of RE on QoE, Bampis *et al.* [20] created a new video database with both RE and bitrate changes, and tested by existing objective VQA methods. Bampis *et al*. [21] proposed a continuous QoE prediction model based on Nonlinear Autoregressive Neural Network with Exogenous Variables (NARX). This continuous QoE prediction model takes an objective measure of perceptual video quality, rebufferring-aware information, and a QoE memory descriptor as three QoE-aware inputs. The dynamic neural network NARX with one input layer, one hidden layer which has 8 hidden nodes, and one output layer is used as the prediction model and is trained by the Levenberg-Marquardt algorithm to predict QoE.

Considering that Hammerstein-Wiener model can better reflect the characteristics of human memory, this paper uses the model, taking frame quality assessment model, which has high correlation with subjective QoE, characteristics of RE and human memory effect as inputs to establish a continuous QoE prediction model.

III. ANALYSIS OF FACTORS INFLUENCING ON QoE

The analysis of continuous subjective QoE helps to understand the impact of various events (e.g., the duration and frequency of stalls caused by rebuffering event) on QoE during video playing. Especially in the case of changeable network bandwidth, the analysis of continuous subjective QoE can provide a reference for the design of quality-aware video stream bitrate switching algorithm, so as to adjust the duration, number and position of stalls according to QoE prediction to maximize the experience of mobile users during video transmission. Therefore, it is necessary to analyze the influence of bitrate changes, stall events and various subjective memory effect on continuous subjective QoE.

Effect of Bitrate and Its Variation on Continuous Subjective QoE: Due to the dynamic characteristic of wireless network, bitrate will inevitably change during video transmission. Bampis *et al.* [21] showed that bitrate variation affects subjective perception of terminal video quality.

Human visual system (HVS), which is very sensitive to edge regions of images, has orientation selectivity mechanism for visual content extraction [22]. It can effectively extract image structure and conduct scene perception and understanding. At present, the orientation selectivity mechanism

has become one of the standard models for visual signal representation in primary visual cortex (PVC). Wu et al. [22] showed that when HVS perceives images, it stimulates different arrangement of excitation or inhibition of visual cortical neurons, thereby generating different orientation selectivity visual patterns (OSVP) to understand the image.

For image F, its OSVP is defined as an arrangement of the spatial correlation between its central pixel and adjacent pixels as

$$P(x|X) = A(\Lambda(x|x_1), \Lambda(x|x_2), \cdots, \Lambda(x|x_n)) \tag{1}$$

where x is the central pixel, x_i is an adjacent pixel, and $X = x_1, x_2, \ldots, x_n$. $\Lambda(x|x_i)$ is the spatial correlation between x and x_i.

The correlation based synaptic plasticity rule indicates that visual cortical neurons with similar preferred orientations have higher connection probabilities and are more likely to exhibit an excitatory response. Therefore, the interaction type between visual cortical neurons depends on their preferred orientations. In section III, we uses the difference between the center pixel x and the adjacent pixel x_i to indicate orientation similarity and defines a threshold of similarity. If the difference is less than the threshold, they have similar preferred orientation, otherwise, they have different preferred orientations.

Define the gradient orientation of the pixel $x \in F$ as its orientation

$$\theta(x) = \arctan \frac{G_v(x)}{G_h(x)} \tag{2}$$

$$f_h = \frac{1}{3}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \quad f_v = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \tag{3}$$

where $G_h = F * f_h$, $G_v = F * f_v$, f_h is the vertical orientation of Prewitt filter, f_v is the horizontal orientation of Prewitt filter, and $*$ represents the convolution operation. According to the orientation similarity between the central pixel x and the adjacent pixel x_i, $\Lambda(x|x_i)$ can be defined as

$$\Lambda(x|x_i) = \begin{cases} + & if\ |\theta(x) - \theta(x_i)| < T \\ - & others \end{cases} \tag{4}$$

where "+" indicates an excitatory interaction, "-" indicates an inhibitory interaction, and T is a similarity threshold, below which the two pixels are considered to have similar orientations. The subjective visual masking experiment reveals that if the orientations are the same, the masking effect between adjacent gratings is stronger, and the masking effect is weakened as the orientation difference increases. When the orientation difference is greater than the threshold 12°, the masking effect becomes marginal. Considering the positive and negative property of orientation difference, T is set to 6° [22].

The OSVP mode between the center pixel and the surrounding pixels is obtained by the equations (1) ∼ (4). As shown in Fig. 1, OSVP between the center pixel and surrounding pixels of 8 neighborhood is $P(x|X) = \{+ - - - - - - -+\}$.

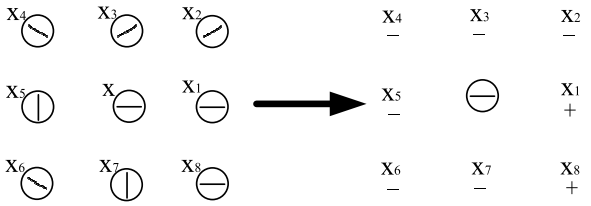


FIGURE 1. OSVP between the center pixel and surrounding pixels of 8 neighborhood.

In the 8 neighborhood, there are 8 pixels around the center pixel, so there are 2^8 possibilities for OSVP mode. Since pixels with the same number of excitation states in OSVP display the same visual information [22], for example, both $\{- + + + + + + +\}$ and $\{+ - + + + + + +\}$ have 7 excitation states which mostly exist in ordered regions (such as uniformly structured sky).Therefore, in order to reduce the amount of calculation, these OSVP types which have the same number of excitation states are combined, and the input image is mapped into an OSVP-based histogram (OSVPH), as shown in equation (5). Thus, there are 9 OSVPH states with the same number of excitation states in 8 neighborhood.

$$HIS(k) = \sum_{i=1}^{M} w(x_i)\delta(P(x_i), P^k) \tag{5}$$

$$\delta(P(x_i), P^k) = \begin{cases} 1 & if\ P(x_i) = P^k \\ 0 & else \end{cases} \tag{6}$$

where $HIS(k)$ represents the k^{th} histogram value for the k-bin, M represents the pixel number of the image, P^k represents the k^{th} OSVP arrangement vector and $w(x_i)$ is the weighting factor. Since pixels having larger luminance variation are more attractive to human, the weighting factor $w(x_i)$ of the pixel x_i is directly related to the luminance variation

$$w(x_i) = \mathrm{var}(x_i) \tag{7}$$

where $\mathrm{var}(x_i)$ is local variance of pixel x_i. According to Eq. (5), pixels having the same OSVP type are combined, and the input image is mapped to an OSVP-based histogram. Although there are 9 OSVP states with the same number of excitation states in 8 neighborhood, pixels with 7 excitation states and pixels with 8 excitation states both exist in the ordered region. Therefore, pixels with 7 excitation states are combined with pixels with 8 excitation states.

Thus, each frame of a video is mapped into an OSVPH, which has 8 bins by extracting OSVP types of each pixel which presents visual information. Fig. 2 shows OSVPH distribution of the first frame of the original video and four mobile videos with different bitrates ($R_1 < R_2 < R_3 < R_4$). Moorthy *et al.* [23] showed that, in subjective video quality experiments, higher bitrate brings better subjective QoE, and when bitrate gradually increases, larger bitrate brings better

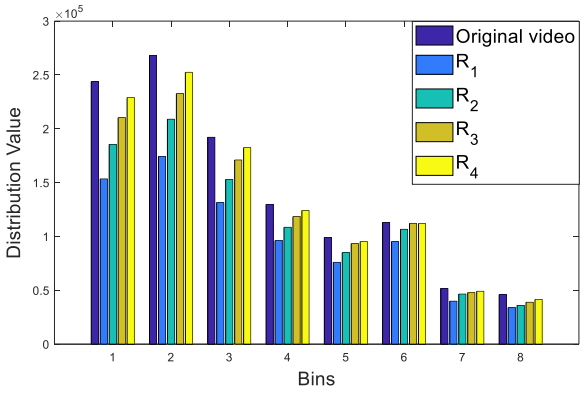


FIGURE 2. OSVPH of the first frame in original video and four videos with different bitrate.

subjective QoE. As it can be seen from Fig. 2, the reference frame has the highest OSVPH distribution value in the eight bins, and as the bitrate increases, OSVPH distribution value of each bin increases and approaches the OSVPH of the reference frame. Therefore, OSVPH can be used as the spatial domain feature of bitrate changing video to measure quality of its frames. In summary, quality of bitrate changing frames extracted by OSVPH method can be used to measure the impact of bitrate change on continuous subjective QoE. Therefore, frame quality (FQ) is extracted frame by frame using OSVPH, and the extracted FQ is taken as one of the inputs of the prediction model.

Impact of Rebuffering Event on Continuous Subjective QoE: Rebuffering event can cause stalls during the video playing, which often affects QoE [24]. In order to study the effect of RE on subjective QoE, this paper analyzes the RE effect on subjective QoE by analyzing characteristics of stalls such as number, duration and position of stalls.

A. NUMBER OF STALLS

In order to analyze the influence of number of stalls on QoE, in this paper, subjective QoE of videos with initial delay and different number of stalls in LIVE mobile stall video database II is studied, without considering duration of stalls, as shown in Fig. 3.

Consider the duration of each stall is medium (5-9 seconds). Fig. 3(a) shows a continuous subjective QoE of videos with long initial delay and few stalls (x-lfs) and videos with long initial delay and multiple stalls (x-lms); Fig. 3(b) shows a continuous subjective QoE of videos with a short initial delay and few stalls (x-sfs) and videos with a short initial delay and multiple stalls (x-sms). From Fig. 3, while the number of stalls increases, QoE tends to become smaller even if video quality returns to an acceptable level after stall occurrence. This shows that number of stalls seriously affects QoE.

B. DURATION OF STALLS

In order to analyze the influence of duration of stalls on continuous subjective QoE, assuming that initial delay

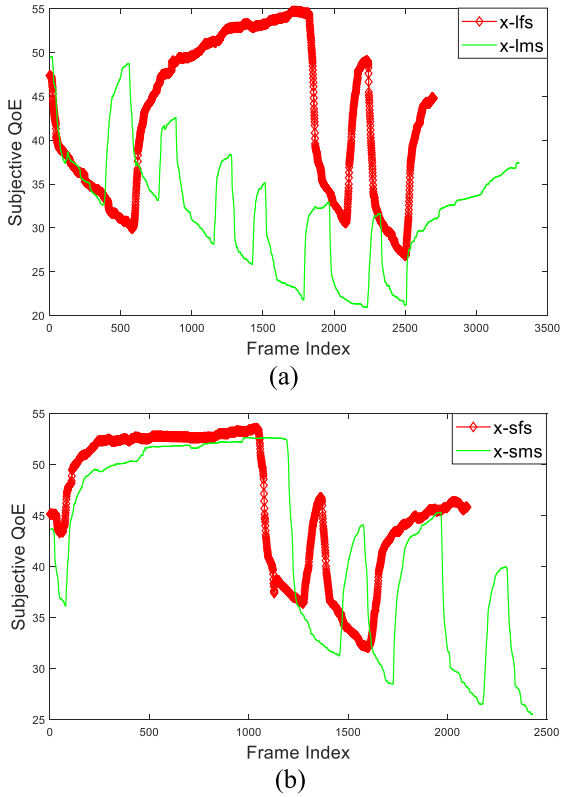


FIGURE 3. Subjective QoE of videos with initial delay and different number of stalls.

and number of stalls are constant, QoE of videos with short/medium/long stalls in the LIVE mobile stall video database II is studied, as shown in Fig. 4. Fig. 4(a) shows a continuous subjective QoE of videos with short initial delay and medium/long stalls(x-sms and x-sls); Fig. 4(b) shows a continuous subjective QoE of videos with short initial delay and short/medium stalls(x-sss and x-sms); Fig. 4(c) shows a continuous subjective QoE of videos with long initial delay and medium/long stalls(x-lms and x-lls); Fig. 4(d) shows a continuous subjective QoE of videos with long initial delay and short/medium stalls(x-lss and x-lms). From Fig. 4, QoE of videos that resumes playing after a medium/long duration of stall is lower than QoE of resumed videos after a short stall. This shows that duration of stalls also has a serious impact on QoE, and long-term stall will reduce subjective QoE after video recovery.

C. POSITION OF STALLS

The position at which stall occurs is at the beginning, middle or end of videos, as shown in Fig. 5.

In order to study the influence of stall positions on QoE, assuming that stall duration and initial delay are fixed, the continuous subjective QoE of videos with stalls occurring at different positions in the LIVE mobile stall video database II are compared, as shown in Fig. 6. Fig. 6(a)-(f) show continuous subjective QoE of videos in which stalls occur at the

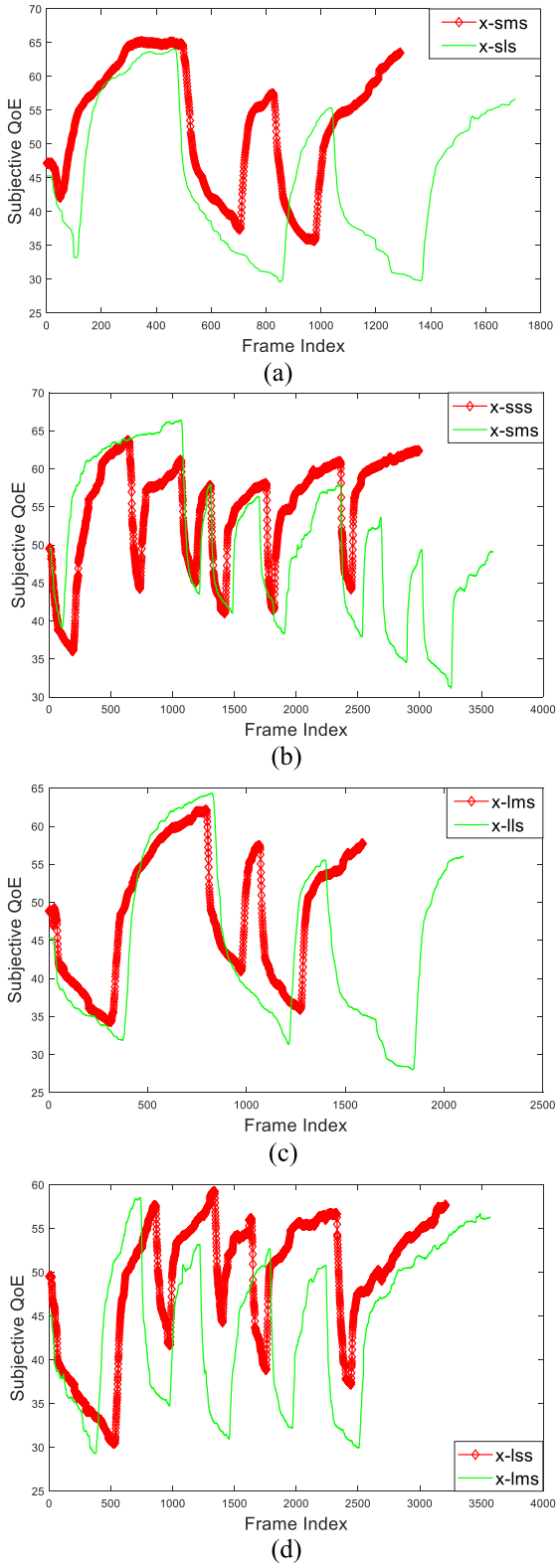


FIGURE 4. Subjective QoE distribution of videos with different duration of stalls.

beginning, middle, end, beginning and middle, middle and end, beginning and middle and end positions, respectively.

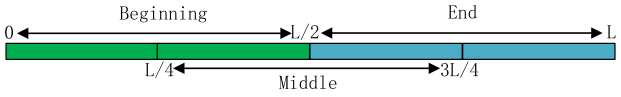


FIGURE 5. Positions at which stalls occur.

In order to observe the influence of stall positions on continuous subjective QoE better, the Dynamic Time Warping (DTW) [25] method is used to normalize the continuous subjective QoE of videos with different stall positions and the corresponding reference video in Fig. 6. The continuous subjective QoE distribution before and after the DTW is shown in Fig. 7, and DTW distance from the reference video is shown in Table 1. The DTW method is a similarity measurement between two time series. The smaller the DTW is, the greater the likelihood that the two time series are similar is. It can be seen from Fig. 6, Fig. 7 and Table 1 that whether the stalls occur at the beginning, the middle or the end, it will cause a decrease in QoE, especially when multiple consecutive stalls occur, QoE will sharply decrease no matter where it occurs. What's more, stalls have a greater impact on QoE when it occurs in the middle or the end.

In summary, video stalls caused by RE have a great influence on continuous subjective QoE. In particular, frequent or long-duration RE have a severe adverse effect on subjective QoE. Therefore, in this paper, we use state of RE as one of the inputs of the prediction model.

D. EFFECT OF MEMORY EFFECT ON CONTINUOUS SUBJECTIVE QOE

Since people are ultimate observers of videos, while people are watching videos, they are inevitably affected by memory effect (ME) such as recency effect, primacy effect and hysteresis effect.

1) RECENCY EFFECT

In social cognitive psychology, when people remember a series of things, ME on the end part is better than other parts, which is called recency effect [26]. Taking stalling event as an example, from Fig. 7 and Table 1, the continuous subjective QoE of videos with stalls at the end is lower than which has stalls at other positions, and it has a larger DTW distance from the continuous subjective QoE of reference videos. This shows that due to the influence of recency effect, stalls occurring at the end of videos impacts the continuous subjective QoE more seriously than that occurring at the other positions.

2) PRIMACY EFFECT

Primacy effect refers to the influence of "first impression" on subsequent cognition of objects in the process of one's social cognition [27]. Taking stalls as an example in Table 1, DTW distance between the continuous QoE of videos with stalls at the beginning and that of the reference video is larger than videos with stalls at the middle. Therefore, stalls that occur

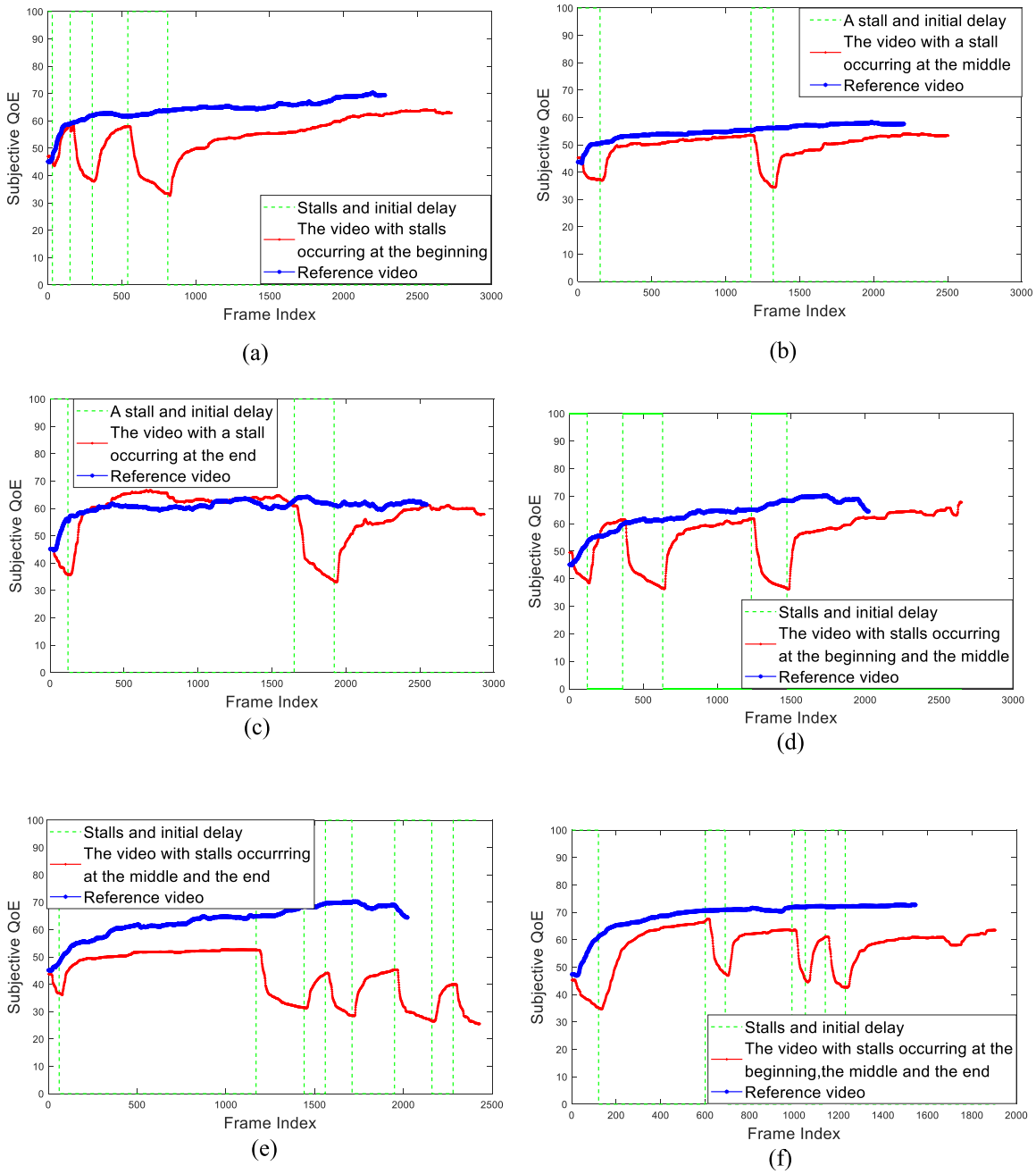


FIGURE 6. Subjective QoE of reference video and videos with different stall positions.

at the beginning has a greater influence on the continuous QoE than that occurs at the middle. It can be seen that due to primacy effect, the impact of stalls occurring at the beginning of videos on continuous QoE cannot be ignored.

3) HYSTERESIS EFFECT

Seshadrinathan and Bovik [28] showed that, in the process of observing videos, there is a hysteresis effect on continuous subjective QoE, which means that due to the occurrence of events such as rebuffering or bitrate reduction, observers respond sharply to the degradation of video quality, and give a lower score for this part of videos and do not have obvious reaction to the improved video quality after these events. Dramatic degradation of video quality during video playing gives observers a bad impression, and even that video quality returns to an acceptable level for observers after these events, the bad impression still retains in their memory, which leads to a lower score for videos with rebuffering or bitrate reduction event.

Therefore, from the above analysis that the continuous subjective QoE will be affected by the unpleasant memory such as recency effect, primacy effect, and hysteresis effect.

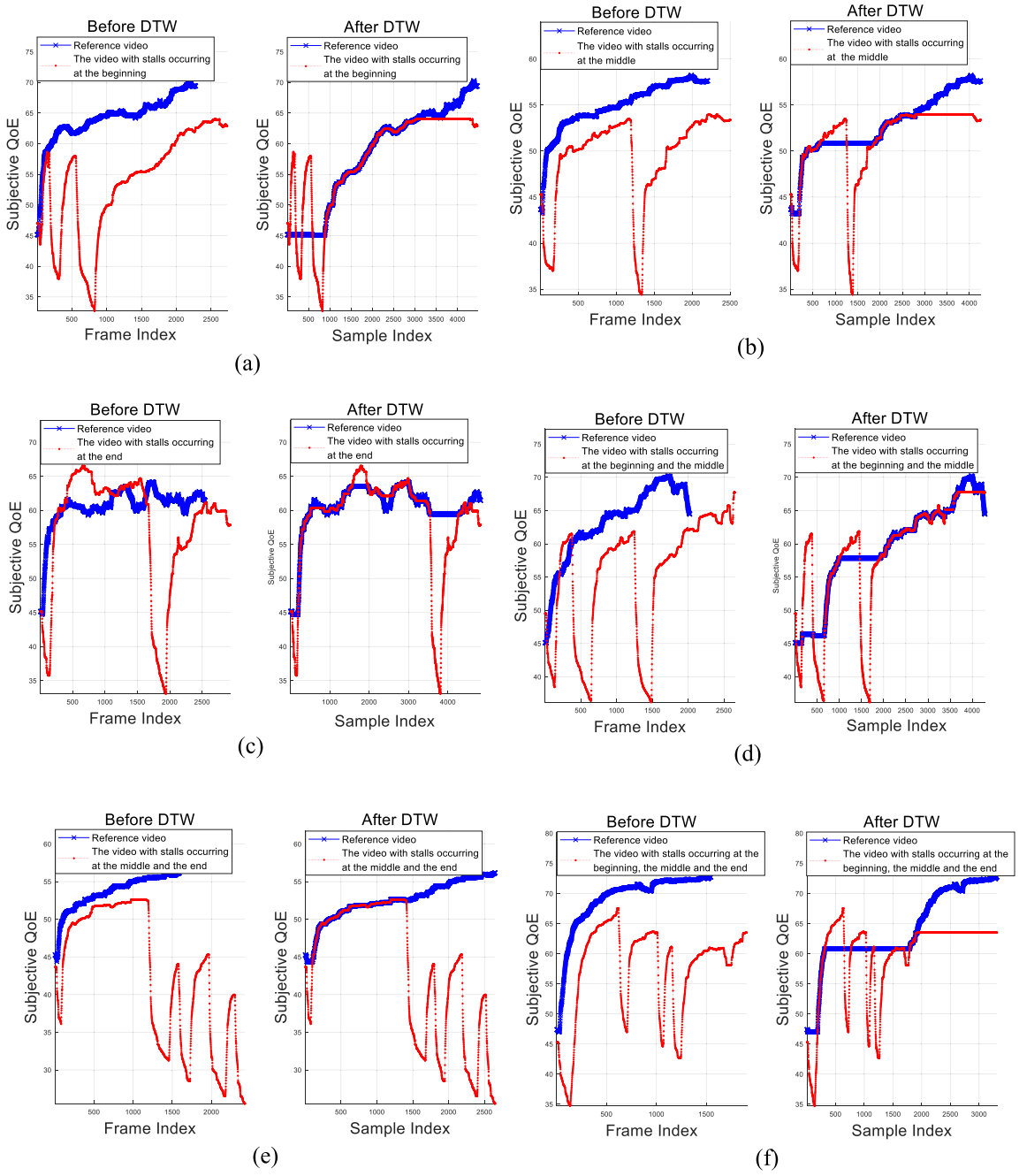


FIGURE 7. Subjective QoE of reference video and videos with different stall positions before/after DTW.

TABLE 1. The DTW of videos with stalls occuring at different position.

Video Type	DTW Distance
Videos with stalls occurring at the beginning	3.6154
Videos with stalls occurring at the middle	3.4256
Videos with stalls occurring at the end	3.8518
Videos with stalls occurring at the beginning and the middle	4.3839
Videos with continuous stalls occurring at the middle and the end	10.0695
Videos with continuous stalls occurring at the beginning, the middle and the end	9.0808

The influence of these effects on continuous subjective QoE should be considered when designing continuous QoE prediction model.

IV. QoE PREDICTION MODEL

Considering Hammerstein-Wiener (HM) model can simulate memory effect such as hysteresis effect, a block-structured nonlinear HW model is used as the prediction model and the input parameters, output parameters, and prediction model structure are as follows.

A. INPUT PARAMETER

Considering the influence of bitrate and its variation, RE and ME on continuous subjective QoE, in this paper, we use FQ vector, RE state vector and ME vector as the inputs of the prediction model.

1) FQ: Frame quality is calculated frame by frame by using OSVPH to form a FQ vector;

2) RE: This paper defines a Boolean continuous time variable RE_1 that describes video playing state at time t, taking $RE_1 = 1$ during rebuffering events occur and $RE_1 = 0$ at other time. This input captures the information related to RE.

3) ME: the ratio of the duration from the impairment event (rebuffering or bitrate drop) occuring to the end of the video to the total duration of the video.

B. OUTPUT PARAMETER

The output parameter is the predicted value of continuous subjective QoE, but in model training phase, in order to get better model parameters, subjective QoE is used as the output parameter.

C. PREDICTION MODEL

The HW model consists of two static nonlinear modules and one dynamic linear module. The dynamic linear module can be represented by a transfer function with n_p poles and n_z zeros. Two static nonlinear modules describe the nonlinear relationship between the inputs and continuous subjective QoE. The structure of the Multiple Input Single Output (MISO) HW model is shown in Fig. 8.

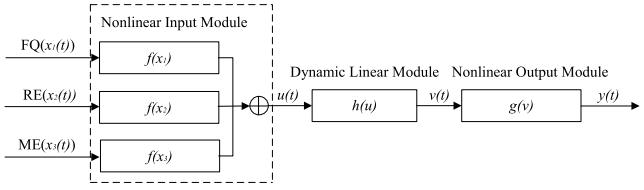


FIGURE 8. The diagram of the MISO HW model.

In Fig. 8, $f(t)$ represents the nonlinear input module function, $g(v)$ represents a nonlinear output module function and $h(u)$ represents the dynamic linear module function. In this paper, a sigmoid function is used to describe the above two nonlinear module functions, and an Infinite Impulse Response (IIR) filter that characterizes long-term ME is used to describe the linear module functions. $u(t)$ is the output of the nonlinear input module, $v(t)$ is the output of the dynamic linear module, $y(t)$ is the output of the nonlinear output module, and z-transformation of the dynamic linear module $h(u)$ is $H(z)$. $u(t)$, $H(z)$, $y(t)$ and $f(t)$ are shown as

$$u(t) = f(x_1(t)) + f(x_2(t)) + f(x_3(t)) \tag{8}$$

$$H(z) = \frac{b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}}{1 - a_1 z^{-1} - \cdots - a_n z^{-n}} \tag{9}$$

$$y(t) = g(v(t)) = \gamma_3 + \gamma_4 \frac{1}{1+\exp(-\gamma_1 v(t) + \gamma_2))} \tag{10}$$

$$f(t) = \beta_3 + \beta_4 \frac{1}{1 + \exp(-\beta_1 x_i(t) + \beta_2)} \tag{11}$$

where $x_1(t)$ represents FQ vector, $x_2(t)$ represents RE state vector, $x_3(t)$ represents ME vector, t represents frame number, $\mathbf{a} = [a_1, \ldots, a_n]^T$ and $\mathbf{b} = [b_1, \ldots, b_m]^T$ represent parameter vector and $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]$ represent the parameters of the sigmoid function.

V. THE PERFORMANCE EVALUATION OF THE PREDICTION MODEL

A. INTRODUCTION TO THE TEST VIDEO DATABASE

To test performance of the prediction model, LIVE-Netflix mobile VQA database [21] built by Image and video engineering laboratory in the university of Texas at Austin is used as test videos. The database consists of 14 original reference videos and 112 distorted videos at 1080p (1920×1080) resolution. The video library provides 8 playout patterns for each original reference video as follows. In pattern 0, the bitrate is fixed at 500 kbps and there is no RE; in pattern 1, the bitrate is 250 kbps, and there is a RE with a duration of 8s at the position of the 28th second, after it, the bitrate is restored to 250 kbps; in pattern 2, the bitrate is 160 kbps and there is no RE; in pattern 3, the bitrate is 195 kbps, and there is a RE with a duration of 4s at the position of the 30th second, after it, the bitrate is restored to 195 kbps; in pattern 4, the initial bitrate is 250 kbps, then reduced to 66 kbps at the position of the 30th second, and finally restored to 250 kbps at the position of the 40th second; in pattern 5, the initial bitrate is 250 kbps, two RE of 6.66 seconds respectively occur at the position of the 20th and the 30th second, the bitrate is restored to 250 kbps after each impairment event; in pattern 6, the initial bitrate is 250 kbps, and a RE of 8.33 seconds occurs at the position of the 20th second, then the bitrate is reduced to 160 kbps, and restored to 250 kbps after about 10 seconds; in pattern 7, the initial bitrate is 250 kbps, then reduced to 100 kbps at the position of the 20th second and restored to 250 kbps after 20 seconds. The events that cause video distortion in each mode are shown in Table 2.

B. CROSS-VALIDATION

In order to obtain the best parameters, during the process of training and testing, videos are firstly divided into con-

TABLE 2. Type of impairment event in each video mode.

pattern	Impairment Event
0	none
1	rebuffering
2	low bitrate
3	rebuffering + low bitrate
4	bitrate drop
5	rebuffering
6	bitrate change + rebuffering
7	bitrate drop

tent independent training subsets and test subsets, wherein training subset accounts for 80% and test subset accounts for 20%. The training data is then further divided into validation subsets. The independence of video contents ensures that subjective prejudice on different video contents is eliminated during training and testing.

The cross-validation strategy used in this paper is as follows. First, numbering all the videos, let $i = 1, 2, \ldots, N$, forming a database containing N videos. Second, the i-th video is randomly selected as the test time series. In order to avoid the influence of same content and playout pattern, we exclude all other videos having either the same content or the same playout pattern as the test video in the training set. Then, the training set is further divided into a training subset and a validation set. This step is repeated r times to ensure adequate data splitting and covering. Subsequently, we evaluate the parameter configuration of the model based on root mean square error (RMSE) and select the model parameters that produce the minimum RMSE as the model parameters which will be used in the test phase. In this experiment, $n_p = n_z = 6$.

C. PERFORMANCE COMPARISON

This paper uses RMSE, Outage Rate (OR) [21] and DTW [20] to evaluate the performance of the prediction model. RMSE captures the overall signal fidelity; OR measures the frequency of times when the predicted value falls outside the 95% confidence interval, and the larger the value is, the lower the prediction accuracy is [21]; DTW reflects the similarity between two time series, the smaller the value is, the greater the similarity possibility of two time series is [20].

We compare the performance of proposed model in three types where the input parameter of the first type is only FQ, the input parameters of the second type are FQ and RE, and the input parameters of the third type are FQ, RE, and ME. Among them, we adopt OSVPH as FQ assessment method. Comparison results of OR, DTW, and RMSE are shown in Table 3. From Table 3, the performance of the prediction model with three inputs consisted of FQ, RE and ME is significantly better than other models.

In this paper, FR image quality assessment (IQA) methods PSNR, SSIM [29] and MS-SSIM [30], RR IQA methods STRRED [13] and OSVPH, NR IQA method NIQE [31] are used as assessment method for FQ. In this paper, taking FQ, RE and ME as inputs, RMSE, OR, and DTW are used to compare the median performance of the prediction model by different FQ assessment methods. The results of the comparison are shown in Table 4, where the best performance is marked with red.

TABLE 3. Median Performance comparison of three different input parameter models in LIVE-NFLX database.

FQ			FQ+RE			FQ+RE+ME		
RMSE	OR	DTW	RMSE	OR	DTW	RMSE	OR	DTW
0.2366	21.8605	81.5358	0.1074	18.2156	31.8224	**0.0566**	**2.0111**	**27.7486**

In order to compare the performance of prediction models using different FQ assessment methods, we use F-test method [32] to estimate the statistical significance of the proposed model in comparison to the state-of-the-art FQ assessment methods conducted on LIVE-Netflix mobile VQA database. F_{test} is performed by computing the squared ratio of RMSE values of a metric A and a metric B, which is defined as equation (12).

$$F_{test} = \frac{(RMSE_A)^2}{(RMSE_B)^2} \tag{12}$$

Table 5 shows the performance of prediction model using different FQ assessment methods. ''1'' indicates that performance of the prediction model with FQ assessment method on column vector is superior to that with FQ assessment method on row vector in statistical results (95% certainty); ''0'' indicates that performance of the prediction model with FQ assessment method on column vector is equivalent to that with FQ assessment method on row vector in statistical results (95% certainty); ''−1'' indicates that performance of the prediction model with FQ assessment method on row vector as input is superior to that with FQ evaluation method on column vector in statistical results (95% certainty).

From Table 4 and Table 5, the performance of the model using OSVPH as FQ assessment method is relatively good, which is better than the model with PSNR, SSIM and MS_SSIM as FQ assessment method, and is comparative to the performance of the model with NIQE and STRRED

TABLE 4. Comparison of median performance of prediction model using different video FQ assessment methods.

FQ assessment method	RMSE	OR	DTW
PSNR	0.0825	16.8950	39.1115
SSIM	0.0739	16.3569	27.7018
MS-SSIM	0.0753	14.6119	26.1974
NIQE	0.0630	42.9224	35.1000
STRRED	0.0593	8.3019	33.9887
OSVPH	**0.0566**	**2.0111**	**27.7486**

TABLE 5. Statistical performance comparison of RMSE of prediction model using different fq assessment methods.

FQ assessment model	PSNR	SSIM	MS-SSIM	NIQE	STRRED	OSVPH
PSNR	-	0	0	-1	-1	-1
SSIM	0	-	0	-1	-1	-1
MS-SSIM	0	0	-	-1	-1	-1
NIQE	1	1	1	-	0	0
STRRED	1	1	1	0	-	0
OSVPH	**1**	**1**	**1**	**0**	**0**	-

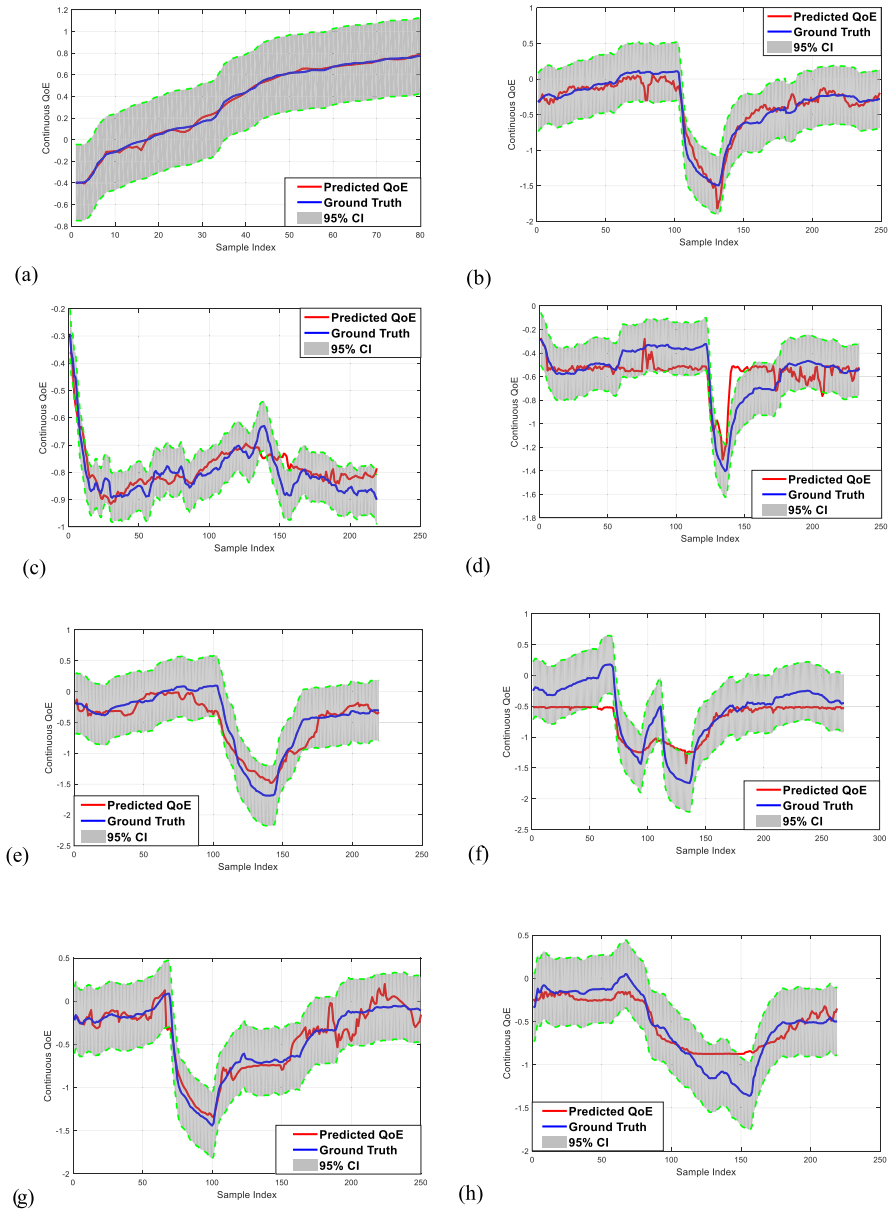


FIGURE 9. The comparison between the predicted and the ground truth QoE in different playout pattern. (a) pattern 0 (RMSE = 0.0004; OR = 0; DTW = 0.7554). (b) pattern 1(RMSE = 0.0113; OR = 0; DTW = 9.9168). (c) pattern 2 (RMSE = 0.0026; OR = 0.2755; DTW = 4.6064). (d) pattern 3 (RMSE = 0.0244; OR = 2.5641; DTW = 9.7173). (e) pattern 4 (RMSE = 0.0349; OR = 0; DTW = 12.2094). (f) pattern 5 (RMSE = 0.0870; OR = 8.9219; DTW = 54.8420). (g) pattern 6 (RMSE = 0.0139; OR = 0.8000; DTW = 11.9976). (h) pattern 7 (RMSE = 0.0291; OR = 8.6758; DTW = 21.0805).

TABLE 6. Performance comparison of different prediction models.

Prediction Model	RMSE	OR	DTW
NARX	0.3322	60.9302	97.3374
GH	0.1131	47.2727	35.0410
Proposed Model	**0.0566**	**2.0111**	**27.7486**

TABLE 7. Statistical results of RMSE performance for different prediction models.

Prediction Model	NARX	GH	Proposed Model
NARX	-	-1	-1
GH	1	-	-1
Proposed Model	**1**	**1**	**-**

as FQ assessment method.In this paper, we compare the performance of the proposed model with the NARX model [20] and GH model [18], the results of RMSE, OR and DTW are shown in Table 6. The best performance is marked in red.

The statistical performance of different prediction models are shown in Table 7. From Table 6 and Table 7, the performance of the proposed model is better than other models.

In this paper, we use the proposed model to test different pattern of videos and show the predicted QoE in 95% confidence interval (CI) and the ground truth in Fig. 9. Fig. 9 (b)-(g) show that the proposed model can accurately predict the effect of rebuffering event on subjective QoE. Fig. 9(f) shows that the proposed model can effectively predict where there are two rebuffering events, and Fig. 9 (e)-(h) show that the model can capture bitrate drop accurately.

VI. CONCLUSION

In order to predict the impact of video impairment events such as bitrate drop and rebuffering on QoE, we have proposed a continuous QoE prediction model. The inputs of the prediction model consist of frame quality, rebuffering event state, and the vector characterizing memory effect, the output of the proposed model is predicted QoE. The proposed model uses a block-structured nonlinear Hammerstein-Wiener model. Experimental results tested on the LIVE-Netflix mobile video database show that the predicted results of the proposed model is well consistent with subjective QoE, and can accurately predict subjective QoE, which can provide a reference for the performance evaluation of wirelss video streaming control strategy.

REFERENCES

[1] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A subjective and objective study of stalling events in mobile streaming videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 183–197, Jan. 2019.

[2] Z. Pan, Y. Zhang, J. Lei, L. Xu, and X. Sun, "Early DIRECT mode decision based on all-zero block and rate distortion cost for multiview video coding," *IET Image Process.*, vol. 10, no. 1, pp. 9–15, Jan. 2016.

[3] Y. Zhang, Z. Pan, N. Li, X. Wang, G. Jiang, and S. Kwong, "Effective data driven coding unit size decision approaches for HEVC INTRA coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3208–3222, Nov. 2018.

[4] Z. Pan, J. Lei, Y. Zhang, and F. L. Wang, "Adaptive fractional-pixel motion estimation skipped algorithm for efficient HEVC motion estimation," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1, Jan. 2018, Art. no. 12.

[5] *Dynamics Adaptive Streaming Over HTTP (DASH)*, Standard ISO/IEC 23009-1, 2014.

[6] M. N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnström, and A. Raake, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Proc. 6th Int. Workshop Qual. Multimedia Exper.*, Germany, Sep. 2014, pp. 141–146.

[7] 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, Apr. 2016.

[8] C. G. Bampis, Z. Li, I. Katsavounidis, and A. C. Bovik, "Recurrent and dynamic models for predicting streaming video quality of experience," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3316–3331, Jul. 2018.

[9] S. Li, L. Ma, and K. N. Ngan, "Full-reference video quality assessment by decoupling detail losses and additive impairments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 7, pp. 1100–1112, Jul. 2012.

[10] J. You, T. Ebrahimi, and A. Perkis, "Attention driven foveated video quality assessment," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 200–213, Jan. 2014.

[11] 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.

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

[13] 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.

[14] C. G. Bampis, P. Gupta, R. Soundararajan, and A. C. Bovik, "SpEED-QA: Spatial efficient entropic differencing for image and video quality," *IEEE Signal Process. Lett.*, vol. 24, no. 9, pp. 1333–1337, Sep. 2017.

[15] F. Yang, S. Wan, Q. Xie, and H. R. Wu, "No-reference quality assessment for networked video via primary analysis of bit stream," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 11, pp. 1544–1554, Nov. 2010.

[16] 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.

[17] 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.

[18] Z. Duanmu, K. Zeng, 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. 2017.

[19] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath, Jr., 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.

[20] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5217–5231, Nov. 2017.

[21] C. G. Bampis, Z. Li, and A. C. Bovik, "Continuous prediction of streaming video QoE using dynamic networks," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 1083–1087, Jul. 2017.

[22] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Inf. Sci.*, vol. 351, pp. 18–29, Jul. 2016.

[23] 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.

[24] 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. 959911-1–959911-8, Sep. 2015.

[25] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *Proc. AAAI-Workshop Knowl. Discovery Databases*, New York, NY, USA, Jan. 1994, pp. 359–370.

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

[27] A. J. Greene, C. Prepscius, and W. B. Levy, "Primacy versus recency in a quantitative model: Activity is the critical distinction," *Learn. Memory*, vol. 7, no. 1, pp. 48–57, Jan. 2000.

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

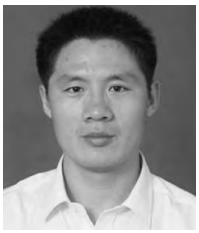
[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[31] 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.

[32] L. Li, D. Wu, J. Wu, H. Li, W. Lin, and A. C. Kot, "Image sharpness assessment by sparse representation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1085–1097, Jun. 2016.

WENJUAN SHI received the M.S. degree in mechatronics from Soochow University, Suzhou, China, in 2006, and the Dr. Eng. degree in information and control engineering from the China University of Mining and Technology, in 2018. Since 2017, she has been an Associate Professor with Yancheng Teachers University. She is currently engaged in the research of deep learning, video quality assessment, and image processing.



YANJING SUN received the Ph.D. degree in information and communication engineering from the China University of Mining and Technology, in 2008. He has been a Professor with the School of Information and Control Engineering, China University of Mining and Technology, since 2012. His current research interests include wireless communication, IBFD communication, embedded real-time systems, wireless sensor networks, and cyber physical systems.

JINQIU PAN is currently pursuing the master's degree in communication engineering with the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interest is dictionary learning, deep learning, and convex optimization.