Video Quality Assessment Based on Short-Term Memory

Ying Fang[1], Weiling Chen[1], Tiesong Zhao[1], Yiwen Xu[1*] and Jing Chen[1]
[1] Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou, Fujian, China
[e-mail: fangying@fzu.edu.cn, weiling.chen@fzu.edu.cn, t.zhao@fzu.edu.cn, xu_yiwen@fzu.edu.cn, 2376719548@qq.com]
*Corresponding author: Yiwen Xu

Received December 17, 2020; revised May 10, 2021; accepted June 10, 2021; published July 31, 2021

Abstract

With the fast development of information and communication technologies, video streaming services and applications are increasing rapidly. However, the network condition is volatile. In order to provide users with better quality of service, it is necessary to develop an accurate and low-complexity model for Quality of Experience (QoE) prediction of time-varying video. Memory effects refer to the psychological influence factor of historical experience, which can be taken into account to improve the accuracy of QoE evaluation. In this paper, we design subjective experiments to explore the impact of Short-Term Memory (STM) on QoE. The experimental results show that the user's real-time QoE is influenced by the duration of previous viewing experience and the expectations generated by STM. Furthermore, we propose analytical models to determine the relationship between intrinsic video quality, expectation and real-time QoE. The proposed models have better performance for real-time QoE prediction when the video is transmitted in a fluctuate network. The models are capable of providing more accurate guidance for improving the quality of video streaming services.

Keywords: Video Quality Assessment, Short-Tterm Memory, Expectation, Quality of Experience

This paper is supported by Natural Science Foundation of Fujian Province [No. 2019J01222], National Natural Science Foundation of China [No. 61901119] and Educational Research Project for Young and Middle-aged Teachers of Fujian Provincial Department of Education [NO. JAT190016].

http://doi.org/10.3837/tiis.2021.07.012 **ISSN : 1976-7277**

1. Introduction

With the rapid development of network and video technologies, streaming media services and applications, such as YouTube, Netflix and other major online video platforms, have been increasing popularity [1]. Due to bandwidth fluctuation, video transmission rate should be automatically adapted to the changes of network states to reduce network congestion[2]. Thus the video quality is fluctuated, which severely impact user's Quality of Experience (QoE) [3]. In order to provide better services to users, service providers need to monitor and evaluate QoE in real time [4]. It is necessary to develop a reliable model of QoE for time-varying video. Many Video Quality Assessment (VQA) methods are based on image quality assessment methods [5]. They measure the perceptual quality of a given video by combining the quality of frame-level qualities, such as Peak Signal-to-Noise Ratio (PSNR) [6], Structural SIMilarity (SSIM) [7], Multi-Scale SSIM (MS-SSIM) [8], and so on. They only take spatial information of video into account without considering temporal features. In general, QoE influence factors consist of system, context and human factors [9]. Human influence factor refers to the user's gender, age, mood, memory, expectation, attention, and so on [10]. Memory effects indicate the psychological influence factor of historical experience. Because of the temporal feature of video and the video quality fluctuation, memory effects can result in a certain deviation from the user's actual experience quality. We can improve accuracy of QoE evaluation by adding the memory effects to VQA model.

When studying video quality evaluation, we must consider not only the quality of each frame of video, but also the impact of previous viewing experience on the user's quality judement. Memory system includes sensory memory, Short-Term Memory (STM) and Long-Term Memory (LTM). LTM provides the lasting retention of information and skills. It gradually becomes silent over time and is difficult to recall information. STM takes in visual and audio information and keeps a copy of it over the time-scale of seconds [11]. The information is kept for a short time but it is relatively easy to recall information. During video, STM refers to the perceptual information accumulated by previous videos. Information is constantly updated, the perceived quality of current video will be more easily affected by STM during video. For example, the memory of poor quality of video sequence causes subjects to provide lower or higher quality scores for the subsequent video sequence compared to the intrinsic quality of video. Intrinsic quality refers to the perceived video quality uneffected by memory [12]. Accordingly, without considering the memory effects, the real-time QoE prediction may has deviation. It is of great importance to study memory-based QoE.

STM plays an important role in influencing the real-time QoE. Some researches have studied the impact of memory effects, including hysteresis effect, primary effect and recent effect. However, these studies do not fully consider memory characteristic and human psychology, such as memory time and the expectations generated by STM. If the duration of previous viewing experience is longer, people are more impressed with the visual information, and the impact of the memory effects may be stronger. It is necessary to investigate the relationship between the duration of past viweing experience and the memory effects. To the best of our knowledge, human memory is difficult to be described completely by mathematical models, the effect of STM formation time on QoE is not well-understood by far. This paper analyzes the influence of STM formation time on real-time QoE and points out stable STM formation time according to subjective experiments.

Meanwhile, this paper introduces the concept of expectation to specify the impact of STM on user's QoE. Expectations based on personal experience, information transmitted by others are beliefs about something that will occur in future and play a role in affectively responding and forecasting [13]. In psychology, the previous experience will make users have certain psychological expectations for current system or condition. Expectations-confirmation theory indicates that the combination of expectations and perceived performance lead to user's satisfaction. Users evaluate perceived performance with respect to their original expectation generated by a period of initial consumption [14]. If current system or condition outperforms expectations, user will feel satisfaction. If current system or condition falls short of expectations, user is likely to be dissatisfied. We integrate confirmation theory and memory effects and extend them in the context of VQA. In consequence, user's satisfaction judgement of current system or condition is based on the level of expectations. This paper utilizes expectation to indicate the influence of previous viewing experience on instantaneous QoE. The expectation is determined by the intrinsic quality of the video watched previously, i.e., quality expectation.

In this paper, we design subjective experiments to investigate the influence of STM on the real-time QoE. The contributions of our works are summarized as follows:

1) Studied the impact of STM formation time on user's QoE and got the stable STM formation time. Different STM formation times lead to different impacts of human memory on video quality evaluation.

2) Investigated the relationship between user expectations and STM as well as the effect of quality expectations on instantaneous QoE based on the stable STM formation time. The expectation built by constant experience quality is different from what is generated by fluctuate experience quality.

3) Developed the comprehensive models to describe the quantitative relationship between user expectations and real-time QoE by integrating the experimental results. The proposed models are beneficial to selecting video quality when video data is transmitted through a fluctuate network.

The rest of this paper is organized as follows. In Section 2, we review the previous works related to memory effects. Section 3 describes the subjective experiment setting, including the quality assessment environment and test methodology. Section 4 qualitatively analyzes the results of the subjective experiments and illustrates the influence of STM formation time and quality expectation on QoE. Section 5 presents STM-QoE models based on memory and expectation at the certain STM formation time, and Section 6 concludes the paper with a summary.

2. Related Work

Existing researches have asserted that memory has impact on user's QoE. Thus some scholars used memory effects to improve the accuracy of predicting QoE.

To evaluate QoE, certain previous works have explored the impact of memory effects, including hysteresis effect, primary effect and recent effect. The hysteresis effect means that the memory of poor video quality elements leads to no significant reflection of the subsequent quality even if it is improved [12]. The primary effect refers to the impact of the user's initial experience on subsequent processes [15]. The recency effect determines that the QoE is evaluation heavily depended on the recent experiences [16]. Kalpana *et al.* [12] verified the impact of the hysteresis effect of human memory on quality of user experience by designing subjective experiments. Samira Tavakoli *et al.* [17] studied the QoE effects of video block

length, switching amplitude, switching frequency, and recent effect. However, conclusions drawn from hysteresis, recent and primary effects are sometimes ambiguities [18].

Several studies optimized QoE models based on hysteresis effect, recent effect, and primary effect. Deepti Ghadiyaram *et al.* [19] proposed a continuous-time video QoE predictor that employed the hysteresis of memory and multiple influencing factors to accurately predict the instantaneous QoE. Chen *et al.* [20] proposed a Hammerstein-Wiener model for predicting time-varying subjective quality of rate-adaptive video, which considered memory effects. They employed the proposed model to simulate human memory effects to find the optimal duration of memory effects. Christos G. Bampis *et al.* [21] proposed QoE continuous prediction of streaming video in dynamic networks. The QoE prediction model was driven by three QoE-aware factors: objective measurement of perceived video quality, rebuffering of perceptual information and memory effects. They used cross-validation methods to find the hysteresis coefficient of memory. Shi *et al.* [22] proposed continuous prediction of QoE in wireless video streams and predicted the impact of video impairment. The inputs of the predictive model were frame quality, the state of the rebuffered event, and the memory effects. The predictive model utilized the block structure nonlinear Hammerstein-Wiener model to simulate the memory effect. Although the impact of human memory has been considered to optimized QoE evaluation models, they only took advantage of the characteristics of hysteresis, recent or primary effects without comprehensively considering the impact of STM time and the varying video quality on memory. In our opinion, the hysteresis effect, recent effect, and primary effect do not fully describe the memory effects.

It has been identified that user expectations have high correlation with the QoE. It has been defined in [23] that QoE is "the overall acceptability of an application or service, as perceived subjectively by the end user, which may be influenced by user expectations and context." Previous viewing experience generates initial expectation of a specific service, users assess the video quality with respect to individual expectation. In this paper, we investigate the relationship between the memory of video quality and quality expectation, then analyze the impact of expectations on QoE evaluation. The VQA method based on expectation-confirmation theory is proposed in [18]. However, the discussed influence of quality adaptations can be applied to videos consisting of two segments, and the duration of video segment used for QoE evaluation was eight seconds. We increase video segments and the duration of video to further investigae the relationship of expectation, intrinsic quality of video and real-time QoE.

Summarizing, although the impact of memory effect have been already studied in previous researches, many questions have not been appropriately resolved. The existing literatures on the study of memory effect mainly concentrated on the characteristics of hysteresis, recent and primary effects. However, above mentioned factors do not fully describe the memory effect. To fully express human memory effects on VQA, memory time and different memory of video quality must be considered. As the memory consolidation time increases, the visual information is more likely to stick to user memory, which may make the memory effects be stronger. Most existing studies obtained the optimal duration of memory effects based on mathematical models. However, there is a deviation from human perception. In addition, quality expectation is built by the memory of video quality, which affects evaluation of video quality. In this paper, we designed experiments to study the influence of STM formation time and the impact of the memory of video quality on QoE, including constant video quality and varying video quality.

3. Subjective Experiment Setting

To examine the impact of STM, three separate experiments were carried out. Experiment I focused on the perceptual impact of STM formation time. The effect of constant viewing experience quality was investigated in Experiment II. Experiment III was designed to study the influence of varying viewing experience quality. All subjective tests were conducted according to ITU-R BT.500-13[24].

3.1 Subjective video database

Ten source videos provided by YouTube, namely *landscapes, hot air balloons, stone house, park fountains, square streets, human activities1, human activities2, house vistas, news interviews, football matches,* were selected in experiments, and the screenshots are shown in **Fig. 1**. Each source video is captured at 24 frames per second and a resolution of 3840×2160. In order to simulate the video streaming application, the video format is mp4, which reflects the video quality required in transmission system. The duration of each video is two minutes. The Spatial Information (SI) and Temporal Information (TI) values of these videos were calculated by (1)-(2) [25], as follows:

$$SI = \max_{\text{time}}\{\text{std}_{\text{space}}[\text{sobel}(F_n)]\}, \tag{1}$$

$$TI = \max_{\text{time}}\{\text{std}_{\text{space}}[F_n(i,j) - F_{n-1}(i,j)]\}, \tag{2}$$

where $F_n(i, j)$ is the pixel at the i-th row and the j-th column of the n-th frame, $\text{std}_{\text{space}}$ is the standard deviation and $\max_{time}$ represents the maximum over all frames.

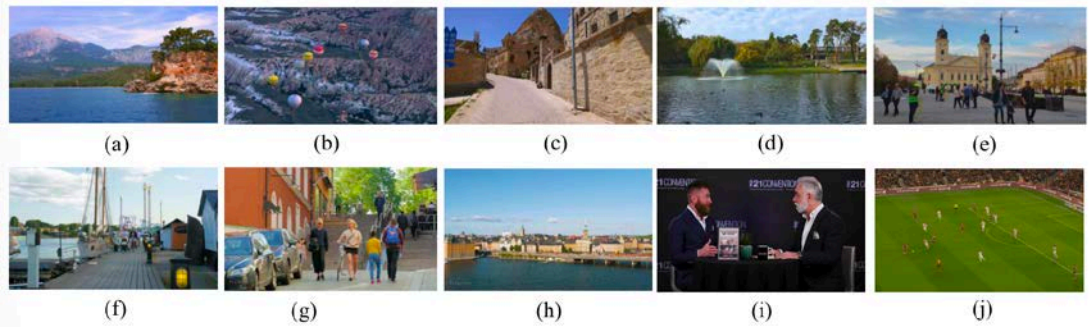


Fig. 1. Snapshots of source videos
(a) landscapes, (b) hot air balloons,(c) stone house, (d) park fountains,(e) square streets, (f) human activities1, (g)human activities2, (h) house vistas, (i) news interviews, (j) football matches

SI and TI of all videos are given in **Fig. 2**. Higher SI indicates more spatially complex scenes and larger TI indicates higher motion video sequence [25]. As shown in **Fig. 2**, the videos are of different spatial and temporal complexity, the source videos selected in this work are representative. Re-buffering is always more unpleasant to subjects than bitrate changes. Therefore, most existing online video playback platforms usually adjust the video spatial resolution according to the network condition. For example, YouTube supports multiple formats and resolutions of the same video [26] and automatically adjusts the format or resolution of the video during video playback. To simulate the typical video streaming application, we selected long video sequences and divided each of them into two segments

with different spatial resolutions, including previous experience video (P-video) and current video (C-video). P-video was designed to test the memory of video quality and made users generate subsequent video quality expectations. C-video was an anchor video. Each source video was encoded with six common spatial resolutions using FFmepg, including 2560 ×1440, 1920×1080, 1280×720, 720×480, 640×360, and 400×300.

We constructed the anchor video and the test video. Each test video sequence consists of P-video and C-video, and video content is continuous. The duration of anchor video was set as 30s. By comparing the Mean Opinion Score (MOS) of anchor video with that of C-video, we investigated the influence of STM.

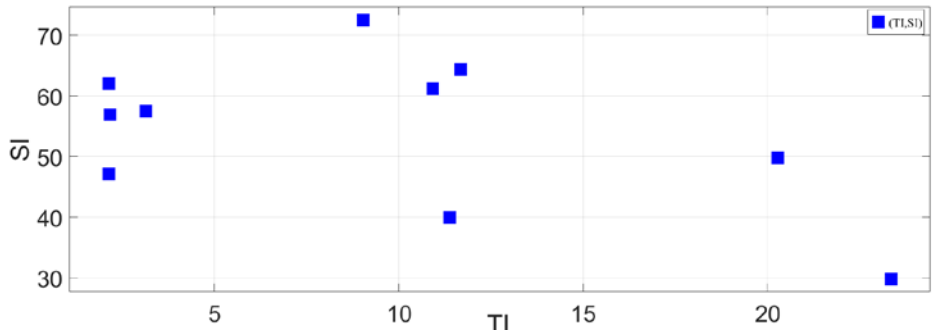


Fig. 2. SI and TI of source videos

Experiment I investigated the effect of STM formation time. There were ten anchor videos with resolution of 1280×720. The duration of P-video was set to five different durations: 15s, 30s, 45s, 60s, and 90s, respectively. The duration of C-video was set as 30s. There were five kinds of test video duration: 45s, 60s, 75s, 90s, and 120s. The test video sequence settings are shown in **Fig. 3**. When the spatial resolution of video switches from 640×360 to 1280×720, the change of video quality is easily to be detected by users. We selected this resolution switching mode to study the impact of STM formation time. The spatial resolution of each P-video was set to 640 ×360. The spatial resolution of each C-video was 1280×720. There were 50 test videos.

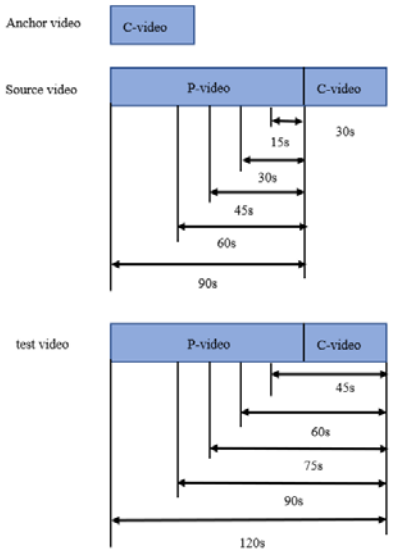


Fig. 3. Videos in Experiment I

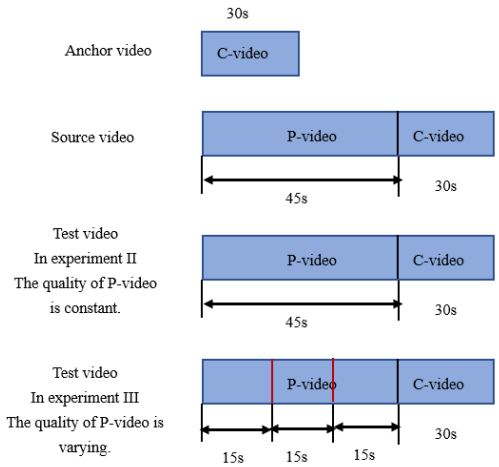


Fig. 4. Videos in Experiment II and III

To further analyze the relationship of expectations and instantaneous QoE, different previous experience qualities were set in Experiment II. The quality of P-video is constant. The test video and anchor video were constructed as **Fig. 4** shows. Thirty anchor videos have the resolutions of either 720×480, 1280×720 or 1920×1080. When the previous experience quality is constant, the expectation genetrated by STM can be represented by the perceptual quality of P-video [14]. P-video and C-video were set to different spatial resolution in order to detect the impact of quality expectation. The spatial resolution settings of P-video and C-video are shown in **Table 1**. The change of video quality included low-to-high resolution switch and high-to-low resolution switch. The duration of P-video was set as 45s. According to the results of experiment I, STM effects tend to be stable when the duration of previous experience is 45s (see **Sec. 4.1**).The duration of C-video was set as 30s.

Table 1. P-video and C-video spatial resolution in experiment II

Resolution switching	Resolution of P-video	Resolution of C-video
Low-to-High	400×300	720×480
	640×360	720×480
	400×300	1280×720
	640×360	1280×720
	720×480	1280×720
	640×360	1920×1080
	720×480	1920×1080
	1280×720	1920×1080
High-to-Low	1280×720	720×480
	1920×1080	720×480
	2560×1440	720×480
	1920×1080	1280×720
	2560×1440	1280×720

Table 2. P-video and C-video spatial resolution in experiment III

Resolution switching	Resolution of P-video	Resolution of C-video	Name
Low-to-High	400×300、640×360、640×360	720×480	S4801
	640×360、400×300、640×360	720×480	S4802
	640×360、640×360、400×300	720×480	S4803
	400×300、720×480、720×480	1280×720	S7201
	720×480、400×300、720×480	1280×720	S7202
	720×480、720×480、400×300	1280×720	S7203
	400×300、720×480、720×480	1920×1080	S10801
	720×480、400×300、720×480	1920×1080	S10802
	720×480、720×480、400×300	1920×1080	S10803
High-to-Low	2560×1440、1280×720、1280×720	720×480	X4801
	1280×720、2560×1440、1280×720	720×480	X4802
	1280×720、1280×720、2560×1440	720×480	X4803

Experiment III investigated the relationship between expectations and fluctuating viewing experience quality as well as the impact on real-time QoE. Test video and anchor video were constructed as shown in **Fig. 4**. There were 30 anchor videos and 120 test videos. The

resolutions of P-video and C-video are shown in **Table 2**. The duration of P-video is 45s. The quality of P-video was varying, which was different from Experiment II. P-video included three video segments of different qualities but continuous content. The duration of each video segment was set to 15s. Two of the three video segments kept the same quality, the quality of another segment was different from them. Subjects watched and evaluated the perceived quality of each video segment. The MOS of each video segment referred to the segment quality. The change of video quality included low-to-high resolution switch and high-to-low resolution switch.The duration of C-video was set as 30s.

3.2 Experimental method

The video sequences were presented on a 27-inch LCD panel with a resolution of 5120×2880 and 32-bit true-color. Since the source video has the resolution of 3840×2160, the screen resolution was adjusted to 3840×2160. The video was played using the playback software potplayer. A total of 25 subjects, including 12 males and 13 females aged between 20 and 25, participated in the subjective experiments. All of subjects are non-experts in video processing, normal visual acuity (or corrected-to-normal acuity), and color vision. The Single Stimulus (SS) methodology is employed in the experiments. After watching the video sequence, subjects had to score it based on their experience with the video quality. The quality rating is based on the ITU-R eleven-grade (ACR-11) quality scale [24], where 0 represents very bad, and 10, perfect, quality. The subjects were asked to take a 10-minute break after 30 minutes to avoid visual fatigue.

Firstly, in order to remove LTM effect, a training session was performed before each experiment to familiarize subjects with the original 4K video and the worst quality video as the reference group. Then, all the anchor videos were displayed in a random order. Subjects watched and evaluated the perceived quality of various anchor videos. Finally, all test videos were displayed in a random order, and subjects watched the video sequence. Separate subjective opinions were collected for the P-video and C-video. In order not to interfere subjects' viewing experience, when subjects watched the video, the recorder asked them the current experience quality and recorded it during the viewing process.

3.3 Experimental data screening

Following the subjective data collection, subject rejection strategy was applied to identify potential outliers in the rating process. The abnormal data was removed by comparing the correlation between the subjective data and the MOS value of each video sequence. We used two evaluation criteria, Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank order Correlation Coefficient (SRCC). If PLCC or SRCC is less than 0.7, the subjective data is removed [25]. In all experiments, each correlation coefficient was in the range of 0.80-0.92, all of the subjective data can be used to analyze the results.

4. Experimental Results and Analysis

To analyze the impact of STM, we focus on how the duration of past viewing experience affect the instantaneous QoE according to experiment I and obtain the stable STM formation time. Then we study the effect of previous experience quality, including constant and varying quality, on instantaneous QoE.

4.1 Influence of STM formation time

In order to observe the MOSs deviation of C-video and anchor video, ΔMOS between C-video and anchor video was calculated. We plot ΔMOS versus different STM formation time in **Fig. 5**. The results show that when the spatial resolution of P-video varies from 640×360 to 1280×720, ΔMOS is positive. The MOS of C-video is higher than that of anchor video. One reason may be that the memory of poor quality of video sequence generates low expectation, thus the quality improvement makes users more pleasant and surprised. When STM formation time is less than 45s, ΔMOS gets larger as the durations of viewing experience increases, which means that the impact of STM effect on QoE becomes stronger with the increase of time. However, when the duration of P-video is longer than 45s, ΔMOS shows no obvious fading and it converges to around 0.75, the impact of STM on QoE tends to be stable.

The results of experiment I show that STM formation time is one of the potential influence factors on QoE. The longer the STM formation time is, the more sensitive the user is to the change of video quality. When the STM formation time is 45s, STM effect tends to be stable. The conclusion can be used to guide video quality switching based on the duration of viewing experience during video playback. When the duration of video watched previously is longer than 45s, the quality of the subsequent video should not change too drastically from the previous one.

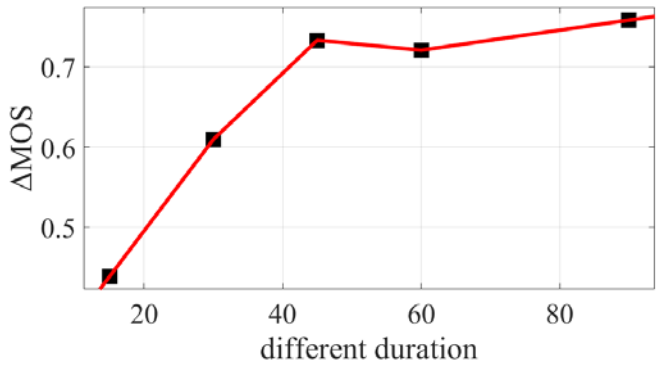


Fig. 5. Relationship between ΔMOS and STM formation time

4.2 Influence of quality expectation

Base on the confirmation of STM formation time, we preliminarily study the impact of quality expectation on QoE. Generally, after watching some videos, users have an expectation for the quality of subsequent video and evaluate the instantaneous QoE by comparing the video quality with their expectations. The user's expectations come from the memory of previous visual information. Experiment II and Experiment III were carried out to analyze the impact of quality expectation built by STM.

4.2.1 Influence of stable STM

In Experiment II, we set P-video quality without fluctuation and quality expectation is determined by P-video. The resolution of C-video was set to either 720 ×480, 1280×720, or 1920×1080. As shown in **Fig. 6**, the change of video quality results in the variation of QoE. The MOS value of anchor video denotes intrinsic quality of C-video. The MOS value of C-video is the perceptual quality affected by STM.

Based on the resolution switching setting in the Experiment II, we classify the resolution changes in low-to-high switching and high-to-low switching. As illustrated in **Fig. 6**, the same C-video has different MOS values when the quality of P-video is different. When the spatial resolution of the test video switches from high to low, the MOS of C-video is lower than that of anchor video though C-video and anchor video are the same video. When the spatial resolution switching is low to high, the MOS of C-video is higher than that of anchor video. The change of P-video quality affects the subjects' opinions on C-video. Different viewing experiences have substantially different impacts on the perceptual quality of subsequent video.

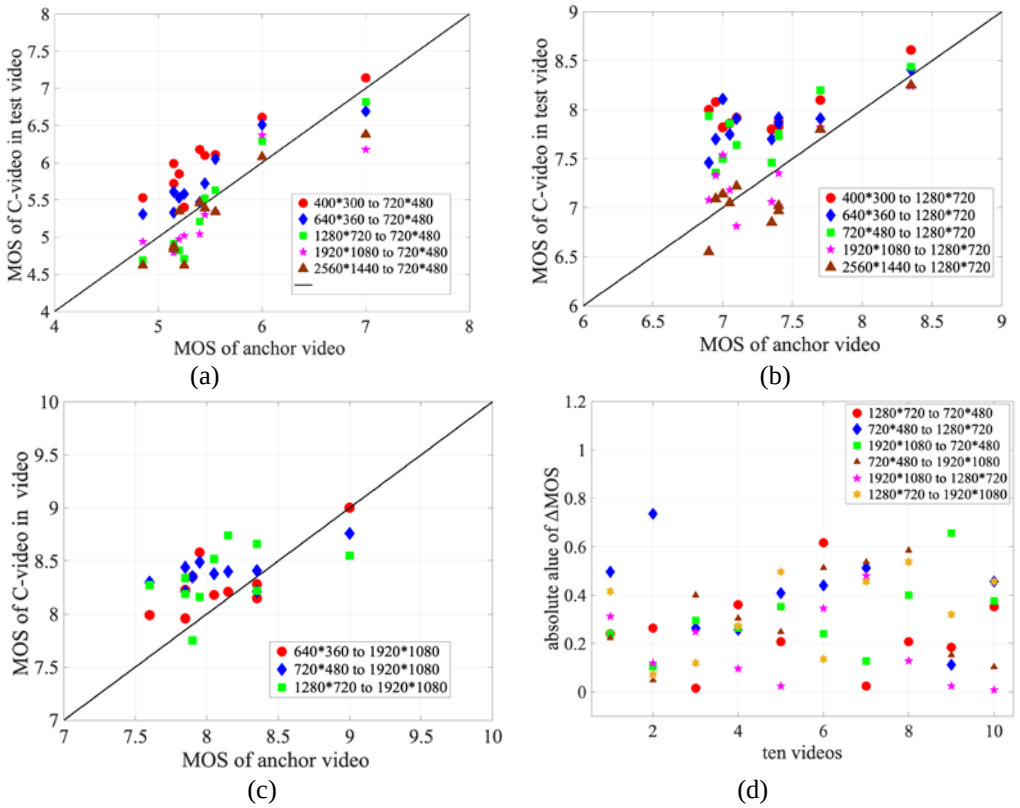


Fig. 6. Comparison of MOS value of anchor video and C-video in Experiment II (a) the resolution of anchor video is 720×480, (b) the resolution of anchor video is 1280×720, (c) the resolution of anchor video is1920×1080 , (d) absolute value of the MOS difference between anchor videos and C-video

Different STM has a distinct effect on user's QoE. When users are previously exposed to low quality of videos, the viewing experience builds low expectations for the subsequent video quality, thus the quality improvement makes users feel more pleasant. Users tend to give reward to quality improvement. Conversely, when the spatial resolution switching is high to low, users tend to give penalty to quality degradation because of high expectations. If the change of video quality is in a positive direction, the perceived quality of the subsequent video is generally higher than its intrinsic quality and vice versa, which is consistent with the conclusion in [18].

Furthermore, from the results shown in **Fig. 6(d)**, the impact of quality increase on the QoE is stronger than the impact of quality decline. The amplitude of variation in video resolution is

the same, but the absolute value of MOS difference between anchor video and C-video caused by low-to-high resolution switch is greater than that caused by high-to-low resolution switch. Subjects are even more delighted by the increasing quality of the video, so that a reward for video quality improvement is higher than the penalty for quality decline.

As final note, different quality switching amplitudes have different impacts on the judgement of the subsequent video quality. As shown in **Fig. 6**, as the quality switching amplitude increases, the effect of previous viewing experience on QoE gets stronger. Previous low-quality viewing experience results in lower expectations. Users will give more rewards to video services if the quality of service is beyond their expectation. Similarly, higher previous experience quality generates higher expectations. Once the video quality declines, the quality of service is contrary to users' expectation, users will inevitably give corresponding penalties.

4.2.2 Influence of fluctuating STM

According to the results of Experiment III, we compared the MOS of C-video with that of anchor video and analyzed the relationship of quality expectations and previous fluctuating viewing experience as well as the impact of expectations on VQA.

P-video includes three video segments. Two of them kept the same quality, the quality of another one is different from them. The experimental results are shown in **Fig. 7**, where the horizontal axis shows MOSs of anchor videos (MOS1), and the vertical axis indicates MOSs of C-videos (MOS2).

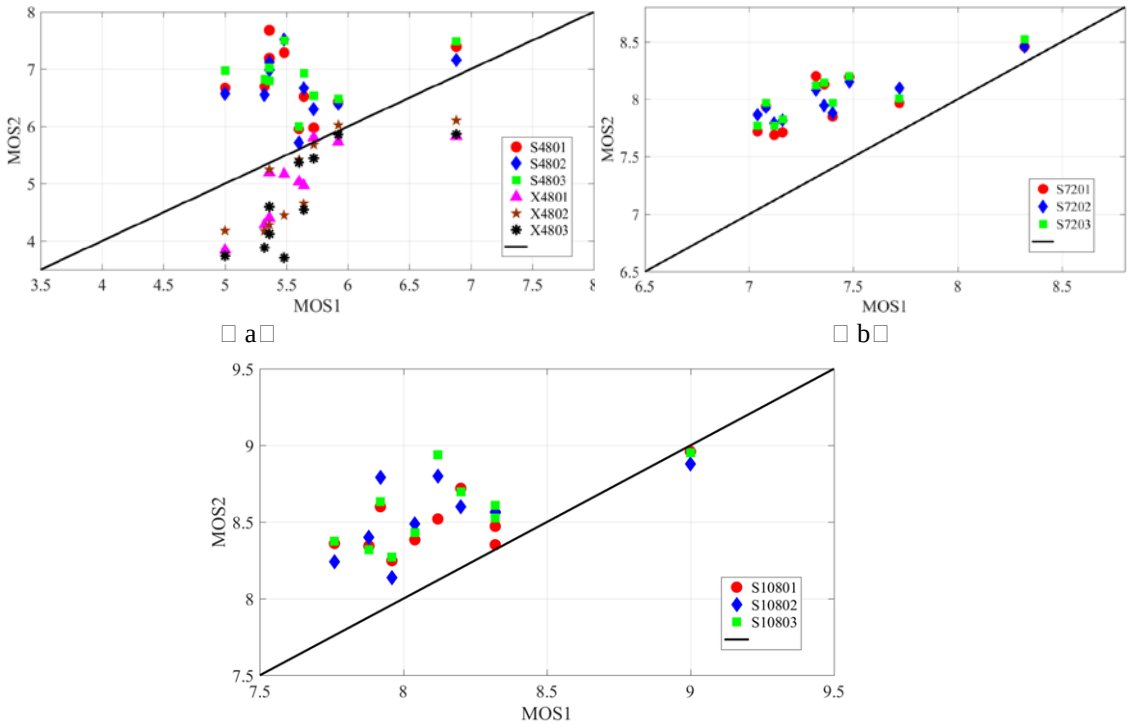


Fig. 7. Comparison of MOS value of anchor video and C-video in Experiment III

When the overall quality of P-video is different from the quality of C-video, the impact of expectation is the same as described in Sec. 4.2.1. When the video resolution is converted from low to high, the MOS of C-video is greater than that of anchor video, vice versa. However, the different variations in the quality of three video segments lead to different judgement of the perception quality of C-video. As shown in **Fig. 7**, when the quality of first two P-video segments is the same but the quality of the third P-video segment changes, the difference between MOS2 and MOS1 of the C-video is the greatest. The impact of the quality of the third video segment on real-time QoE is strongest, which means that the impact of the recent experiences is the greatest.

5. STM-QoE Model

In this section, on the basis of empirical findings from our subjective experiments, we propose new objective QoE assessment model based on STM (STM-QoE). More specifically, we investigate the influence of constant quality expectation and construct Stable STM-based QoE assessment model (SSTM-QoE) and Fluctuation STM-based QoE assessment model (FSTM-QoE).

In the application of streaming media transmission platform, we should design a model that has simple structure and high computational efficiency and can accurately predict the perceived QoE. We employ a simple linear weighted model that can be adpoted to directly and explicitly model the combined effects of expectation and instinc qualtiy as follows:

$$Q_t = \alpha \cdot E_t + \beta \cdot q_t + C, \tag{3}$$

where Q_t is the instantaneous QoE at moment t, E_t is the quality expectation generated by video watched previously and it is determined by the intrinsic quality of previous video, q_t is the intrinsic quality of video at moment t, α and β denote the weights of expectation and intrinsic quality, respectively, and C is adjustment coefficient.

Expectation is a subjective factor and may not be directly obtained in many streaming applications. In order to improve the usability of the model, we transform expectation into objective quality. In generally, PSNR and SSIM are used to evaluate the quality of video. In this work, we employ SSIM as the objective video quality index. We plot the MOS values of P-videos and anchor videos versus their SSIM values in **Fig. 8**. There is a non-linear correlation between SSIM of the video and the MOS value. In order to improve the model prediction accuracy, we study the relationship between SSIM and MOS value (Q_{mos}). By examining the general trend of MOS variation with SSIM, we construct the model as follows:

$$Q_{\text{mos}} = e^{2.441 \cdot \text{SSIM}} - 2.694 \tag{4}$$

Root mean squared error (RMSE) and R-square are employed for performance evaluation by comparing MOS values for each video sequence and the model predictions. The RMSE and R-square are 0.896 and 0.68, respectively. The results indicate that the model is reliable. In this work, we map the SSIM of the video to Q_{mos} by (4) and employ Q_{mos} as the objective video quality index.

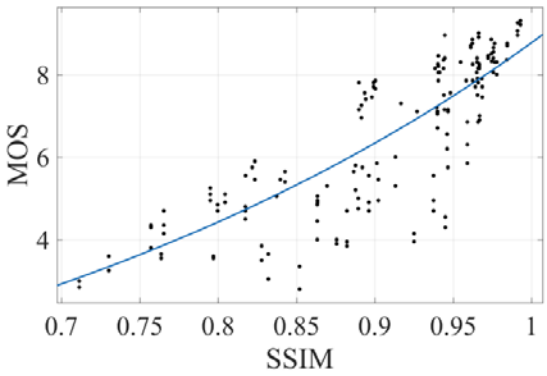


Fig. 8. Relationship between MOS and SSIM of anchor video and P-video

5.1 SSTM-QoE model

With a goal to model the effect of constant previous experience quality on QoE, the experimental data collected by experiment II were randomly divide into training and test sets with an 80/20 split and no content overlapped. The training set is used to obtain the model parameters and the testing set is used to evaluate the prediction performance of the model. To mitigate any bias due to the division of data, the process of randomly dataset splitting was repeated 45 times. To average the results of 45 trainings of (3), the SSTM-QoE prediction model is conducted as follows:

$$Q_{\mathrm{SSTM}t} = -0.465E_{\mathrm{t}} + 1.005q_t + 3.312\,, \tag{5}$$

where E_t is represented by the Q_{mos} of the video watched previously, q_t and $Q_{\mathrm{SSTM}t}$ are the Q_{mos} of the video and real-time QoE at moment t respectively.

Four evaluation criteria, including PLCC, SRCC, Kendall Rank-order Correlation Coefficient (KRCC) and RMSE, are utilized to compare the performance of the VQA models. The comparison in QoE prediction performance between our model and four state-of-the-art methods (SSIM, BSRIQA[27], HOSA[28] and VIIDEO[29]) is shown in **Table 3**. Accordingly, we observe that the proposed model provides a better performance in terms of PLCC, SRCC, KRCC and RMSE against the four state-of-the-art methods. It indicates that the model combined with STM can effectively improve the accuracy of QoE evaluation.

Table 3. Performance comparison of SSTM-QoE and the state-of-the-art methods

Model	PLCC	KRCC	SRCC	RMSE
SSTM_QoE	**0.918**	**0.729**	**0.894**	**0.446**
SSIM	0.779	0.628	0.775	0.744
BSRIQA	0.792	0.679	0.811	0.634
HOSA	0.848	0.575	0.663	0.563
VIIDEO	0.734	0.341	0.494	0.569

5.2 FSTM-QoE model

The quality of each video segment watched previously affect the users' opinion on the subsequent video. To well understand the relationship between expectation and the segment-level perceptual video quality, we analyze the overall QoE of P-video and the MOSs

of three video segments collected in Experiments III and explore a linear weighted model as follows:

$$E = \sum_{i=1}^{3} w_i \cdot q_i \ , \tag{6}$$

where E is the expectations built by fluctuating STM, q_i is the Q_{mos} of the previous i-th video, w_i is the weight of the quality of i-th video.

To obtain the appropriate parameters w_i in the model, the experimental data collected by experiment III was partitioned into training and testing data (80/20 split) with non-overlapping content. The random split was repeated 45 times and we average the results to obtain **W** = [0.156, 0.404, 0.440]. The average PLCC, SRCC, KRCC, RMSE between the predicted expectation and the overall QoE of P-video over these 45 iterations are calculated. Meanwhile, we evaluate the performance of our proposed model and the average strategy of **W** = [1/3, 1/3, 1/3]. After 15s of short viewing experience, the memory may be disturbed, the quality of the third segment of P-video has greatest effect on the subsequent video. Our proposed model accounts for recency effect and provides a better performance against the average pooling strategy, as shown in **Table 4**. We can predict each video segment quality and then calculate expectation by (6).

Table 4. Performance comparison of weight strategy and average strategy

Model	PLCC	KRCC	SRCC	RMSE
linear weighted	0.775	0.555	0.689	0.947
Average	0.763	0.528	0.663	0.973

To obtain the appropriate parameters for FSTM- QoE prediction mode, the dataset is split into disjoint 80% training and 20% test sets. The random split was repeated 45 times and the median PLCC, SRCC, SRCC, and RMSE results are reported in **Table 5**. The values of the weights α and β in (3) are determined by averaging the results of 45 trainings. The model as follows:

$$Q_{\text{FSTM}t} = -0.846E_t + 1.071q_t + 4.964 \, , \tag{7}$$

where E_t denotes the expectation calculated by (6), q_t and $Q_{\text{FSTM}t}$ are the Q_{mos} of the video and real-time QoE at moment t respectively.

The proposed model is able to account for the effect of fluctuating viewing experience and expectation. As shown in **Table 5**, the proposed model performs better than the state-of-the-art methods.

Table 5. Performance comparison of FSTM-QoE and the state-of-the-art methods

Model	PLCC	KRCC	SRCC	RMSE
FSTM_QoE	**0.928**	**0.731**	**0.869**	**0.549**
SSIM	0.724	0.521	0.651	1.061
BSRIQA	0.763	0.528	0.635	0.839
HOSA	0.611	0.593	0.695	1.032
VIIDEO	0.809	0.524	0.651	0.821

The proposed models are useful in better understanding the impact of STM in evaluating time-varying video quality and providing more accurate guidance for video quality switching. For time-varying video, the instantaneous QoE can be predicted at the server side by STM-QoE models. Specifically, according to the perceived quality of video watched previously, content provider can select the quality of subsequent video before video transmission in dynamic network and provide users with better quality of service.

6. Conclusion

In conclusion, STM formation time is one of the potential influence factors on QoE. The impact of positive video quality switching is greater than that of negative video quality switching. The degree of STM effect on user's QoE is correlated with the video quality variation amplitude. When the STM is fluctuating, each stage of memory has an impact on the video quality evaluation, but the recent memory has the greatest impact. We demonstrated that the influence of different quality expectations built by STM results in different impacts on user's QoE. We proposed STM-QoE models for QoE prediction of time-varying video, including SSTM-QoE model and FSTM-QoE model. The models perform better against SSIM. The models can be adopted in QoE monitoring when video data is transmitted through a fluctuated network.

For future work, we will desire more video quality of varying patterns to better understand the STM effect. Considering long-term dependencies of viewing experience, we will investigate the influence of LTM. Additionally, we plan to study the combined impact of LTM and STM on real-time QoE. Furthermore, due to channel error or packet loss in the video transmission system, the received video sequences might be different. We will investigate the impact of these factors on QoE to optimize the models. Moreover, the impact of temporal masking effects on the current model is another challenging problem that desires further investigations. We wish our study in memory effects will inspire further research on developing an accurate and low complexity model for QoE prediction.

References

[1] P. Juluri, V. Tamarapalli and D. Medhi, "Measurement of Quality of Experience of video-on-demand services: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 401–418, February 2015. Article (CrossRef Link).

[2] A. Biernacki, "Improving Video quality by diversification of adaptive streaming strategies," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 1, pp. 374-395, 2017. Article (CrossRef Link).

[3] L. Gao, Y. Xie, J. Qi and Z. Li, "A multifunctional video quality assessment system," in *Proc. of 5th International Congress on Image and Signal Processing*, 16-18 Oct. 2012. Article (CrossRef Link).

[4] X. Liu, M. Chen, T. Wan and C. Yu, "Hybrid No-Reference Video Quality Assessment Focusing on Codec Effects," *KSII Transactions on Internet and Information Systems*, vol. 5, no. 3, pp. 592-606, 2011. Article (CrossRef Link).

[5] A. B. Watson and J. Malo, "Video quality measures based on the standard spatial observer," in *Proc. of International Conference on Image Processing*, 22-25 Sept. 2002. Article (CrossRef Link).

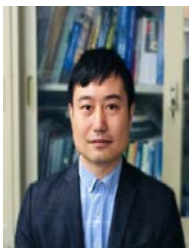
[6] M. Zink, J. Schmitt, and R. Steinmetz, "Layer-encoded video in scalable adaptive streaming," *IEEE Transactions on Multimedia*, vol.7, no.1, pp.75–84, Feb.2005.

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004. Article (CrossRef Link).
[8] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. of The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 9-12 Nov. 2003. Article (CrossRef Link).
[9] T. Zhao, Q. Liu and C. W. Chen, "QoE in video transmission: a user experience-driven strategy," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp.285–302, October 2016. Article (CrossRef Link).
[10] U. Reiter and K. Brunnström, K. D. Moor, M-C. Larabi, M. Pereira, A. Pinheiro, J. You and A. Zgank, "Factors influencing Quality of Experience," *Quality of Experience*, pp. 55–72, 2014. Article (CrossRef Link).
[11] N. Cowan, "What are the differences between long-term, short-term, and working memory?," *Progress in Brain Research*, 169, 323-338, 2008. Article (CrossRef Link).
[12] K. Seshadrinathan, A. C. Bovik, "Temporal hysteresis model of time varying subjective video quality," in *Proc. of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing* (*ICASSP*), 22-27 May 2011. Article (CrossRef Link).
[13] C. G. Bampis and Z. Li, et al., "study of temporal effects on subjective video Quality of Experience," *IEEE Transactions on Image Processing*, vol.26, no.11, pp. 5217 - 5231, Nov. 2017. Article (CrossRef Link).
[14] L. O. Richard, "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions," *Journal of Marketing Research*, vol.17, no.4, pp. 460-469, 1980. Article (CrossRef Link).
[15] J. G. Anthony, P. Colin and B. L. William, "Primacy versus recency in a quantitative model: Activity is the critical distinction," *Learn. Memory*, vol.7, no.1, pp.48–57, 2000. Article (CrossRef Link).
[16] K. Seshadrinathan, A. C. Bovik: "Recency and duration neglect in subjective assessment of television picture quality," *Applied Cognitive Psychology*, vol.15, no.6, pp.639–657, 2001. Article (CrossRef Link).
[17] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, N. García, "Perceptual 1uality of HTTP adaptive streaming strategies: cross-experimental analysis of multi-Laboratory and crowdsourced subjective studies," *IEEE Journal on Selected Areas in Communications*, vol.34, no.8, pp.2141–2153, June 2016. Article (CrossRef Link).
[18] Z. Duanmu, K. Ma, Z. Wang, "Quality-of-Experience for adaptive streaming videos: An expectation confirmation theory motivated approach," *IEEE Transactions on Image Processing*, vol.27, no.12, pp. 6135-6146, 2018. Article (CrossRef Link)
[19] D. Ghadiyaram, J. Pan, and A. C. Bovik, "Learning a continuous-time streaming video QoE model," *IEEE Transactions on Image Processing*, vol.27, no.5, pp.2257–2271, 2018.
[20] C. Chao, L. K. Choi, G. Vecianna, C. Caramanis, R.W. Heath and A. C. Bovik, "Modeling the time-varying subjective quality of HTTP video streams with rate adaptations," *IEEE Transactions on Image Processing*, vol.23, no.5, pp.2206–2221, 2014.
[21] C. Bampis, Z. Li and A. C. Bovik, "Continuous prediction of streaming video QoE using dynamic networks," *IEEE Signal Processing Letters*, vol.24, no.7, pp.1083–1087, May 2017. Article (CrossRef Link)
[22] W. Shi, Y. Sun and J. Pan, "Continuous prediction for Quality of Experience in wireless video streaming," *IEEE Access*, vol.7, pp. 70343 - 70354, May 2019. Article (CrossRef Link).
[23] "Vocabulary and effects of transmission parameters on customer opinion of transmission quality," amendment 2, ITU-T Recommendation P.10/G.100, Tech. Rep., 2017.
[24] Methodology for the subjective assessment of the quality of television pictures, ITU-R Recommendation BT.500-13, 2012.
[25] Methodology for the subjective assessment of video quality in multimedia applications, ITU-R Recommendation BT.1788, 2007.

[26] A. Finamore, M. Mellia, M. M. Munafò, R. Torres and S. G. Rao, "YouTube everywhere: Impact of device and infrastructure synergies on user experience," in *Proc. of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp.345–360, 2011. Article (CrossRef Link).
[27] Y. Fang, C. Zhang, W. Yang, J. Liu and Z. Guo, "Blind visual quality assessment for image super-resolution by convolutional neural network," *Multimedia Tools and Applications*(*MTAP*), vol. 77, pp. 29829–29846, 2018. Article (CrossRef Link)
[28] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu and D. Doermann, "Blind image quality assessment based on high order statistics aggregation," *IEEE Transactions on Image Processing*, vol.25 no.9, pp. 4444-4457, 2016. Article (CrossRef Link).
[29] A. Mittal, M. A. Saad and A. C. Bovik, "A completely blind video integrity oracle," *IEEE Transactions on Image Processing*, vol.25, no.1, pp.289–300, 2016. Article (CrossRef Link).

Ying Fang received the M.S. degree in communication and information system from Fuzhou University, Fujian, China, in 2007. She is currently pursuing the Ph.D degree in communication and information system in Fuzhou University, Fujian, China. Her research interests include video signal processing and visual quality assessment.

Weiling Chen (Member, IEEE) received the B.S. and Ph.D. degrees in communication engineering from Xiamen University, Xiamen, China, in 2013 and 2018, respectively. She is currently a Lecturer with the College of Physics and Information Engineering, Fuzhou University, China. From Sep. 2016 to Dec. 2016, she was visiting at the School of Computer Science and Engineering, Nanyang Technological University, Singapore. Her current research interests include image quality assessment, image compression, and underwater acoustic communication.



Tiesong Zhao (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Science and Technology of China, Hefei, China, in 2006, and the Ph.D. degree in computer science from the City University of Hong Kong, Hong Kong, in 2011. He served as a Research Associate with the Department of Computer Science, City University of Hong Kong (2011-2012), a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo (2012-2013) and a Research Scientist with the Ubiquitous Multimedia Laboratory, The State University of New York at Buffalo (2014-2015). He is currently a Minjiang Distinguished Professor in the College of Physics and Information Engineering, Fuzhou University, China. His research interests include multimedia signal processing, coding, quality assessment and transmission. Dr. Zhao has around 50 publications in these fields. Due to his contributions in video coding and transmission, he received the Fujian Science and Technology Award for Young Scholars in 2017. He has also been serving as an Associate Editor of IET Electronics Letters since 2019.

Yiwen Xu received the Ph.D degree in the department of electronic engineering from Xiamen University, Xiamen, China, in 2012. He has been an Associate Professor with the College of Physics and Information Engineering, Fuzhou University, Fujian, China, since 2013. His research interests lie in multimedia information processing, video codec and transmission, and video quality assessment.

Jing Chen received the M.S. degree in electronics and communication engineering from Fuzhou University, Fujian, China, in 2020. Her research interests include image and video signal processing and visual quality assessment.