Graduate School of Engineering and Science 2020

Shibaura Institute of Technology

Master's Thesis

Title : Improving QoE Prediction Performance for Video Streaming Services

Major	Graduate School of Engineering and Science (Master's Program) Global Course of Engineering and Science
Student ID No.	MG18502
Name	Nguyen Duc Tho
Supervisor	Prof. Kamioka Eiji

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university.

Nguyen Duc Tho
April 2024

Acknowledgements

I would like to express deep and sincere gratitude to my supervisor, Prof. Kamioka Eiji, for the continuous help and indispensable guidance which led me to the completion of this thesis, for his patience and motivation.

I would also like to thank Dr. Phan Xuan Tan, for inspiring and guiding me throughout my studies. It would have been more difficult without his invaluable advice and suggestions.

I would like to thank my family for their love and hard work to make it possible for me to complete this thesis. I would like to especially thank my parent for being there for me always. To my little brother, you have been my main inspiration during all these years.

Last but not least, I have been very thankful to my friend, Tran Minh Chanh, for his support in brainstorming new ideas and encouragement to complete this thesis.

Abstract

The growing demand on video streaming services increasingly motivates the development of reliable and accurate models for the assessment of the user's Quality of Experience (QoE) in real-time to deliver high-quality streaming content to the user. However, the complexity caused by the temporal dependencies in sequential QoE data and the non-linear relationships among QoE influence factors has introduced challenges to continuous QoE prediction. This thesis proposes three novels QoE prediction models in order to improve the QoE prediction performance in terms of prediction accuracy and computational complexity.

- First, to enhance the QoE prediction accuracy, the BiLSTM-QoE model is proposed. The model utilizes a Bidirectional Long Short Term-Memory network for predicting the user's QoE.
- Second, the CNN-QoE model is introduced. The model leverages advantages of the Convolutional Neural Network to overcome the computational complexity drawbacks of Long Short Term-Memory networks while improving QoE prediction accuracy. Based on a comprehensive evaluation, the CNN-QoE model provides a high QoE prediction performance and outperforms the existing approaches.
- Third, human-related factors have a significant influence on QoE and play a crucial role in QoE modeling. However, these factors were not considered in the BiLSTM-QoE and CNN-QoE model due to the lack of data on QoE influence factors. Therefore, in order to precisely model the user's QoE, the impact of the human-related factors, namely perceptual factors, memory effect, and the degree of interest is investigated. Based on the investigation, a novel QoE model is proposed that effectively incorporates those factors to reflect the user's cumulative perception. Evaluation results indicate that the model performs excellently in predicting cumulative QoE at any moment within a streaming session.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Acronyms / Abbreviations

BiLSTM	Bidirectional Long Short-Term Memory
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
MOS	Mean Opinion Score
PCC	Pearson Pearson Correlation Coefficient
QoE	Quality of Experience
RMSE	Root Mean Squared Error
SROCC	Spearman Rank Correlation Coefficient
TCN	Temporal Convolutional Network

Chapter 1

Introduction

1.1 Motivation

In recent years, video streaming has become the most dominant contributor to global Internet traffic. The Cisco Visual Networking Index forecasts an increase in video traffic, which is expected to reach 82% by 2021, up from 73% in 2016 [2]. The rapid increase of video streaming services creates an extremely huge profit for streaming service providers. In the context of a highly competitive streaming service market, service providers such as YouTube, Netflix, or Amazon must improve and ensure a sufficient video quality to satisfy the user's expectation, resulting in high quality of experience (QoE). However, video streaming services are frequently influenced by dynamic network conditions (e.g., throughput, available bandwidth) which can lead to distorted events (e.g., bitrate switching, rebuffering events). These distorted events can negatively affect the user's satisfaction, resulting in the deterioration of the user's QoE. The capability of continuously predicting and monitoring the user's QoE in real-time can help video streaming controllers perform a QoE-based network control and management to alleviate the QoE deterioration, resulting in higher overall levels of the user's QoE [3, 4]. Therefore, there is a need for developing reliable QoE prediction models in order to quickly and accurately determine the user's QoE.

However, the continuous prediction of QoE is challenging since the user's QoE is affected by many influence factors such as video quality, video content, bitrate switching, rebuffering, etc. Moreover, in order to accurately predict the user's QoE, it needs to capture the complex temporal dependencies in sequential QoE data and the non-linear relationships among these QoE influence factors [5–8]. Additionally, QoE prediction models have to adapt to the dynamic changes in network conditions in real-time. Therefore, it is necessary to improve the prediction accuracy of QoE models that can perform consistently well across diverse scenarios of video streaming. Furthermore, it is necessary to optimize the model

computational complexity for real-time QoE monitoring. These factors form the motivation for this thesis.

The main goal of the thesis is to improve the QoE prediction performance in terms of both prediction accuracy and computational complexity of QoE prediction models for QoE-based network control and management.

1.2 Research Contributions

Based on the main goal, the work present in the thesis resulted in three novel QoE prediction models. These models are summarized as follows:

1. A QoE prediction model, namely BiLSTM-QoE, is proposed for continuously predicting the user's QoE. The BiLSTM-QoE model utilizes Bidirectional Long Short-Term Memory networks to deal with the complex temporal dependencies in sequential QoE data. The evaluation results show that the BiLSTM-QoE model achieves promising performance in terms of accuracy compared with other referenced models.

2. Despite the high accuracy of the BiLSTM-QoE model, the sequential processing characteristic in BiLSTM architecture increases the computational complexity of the model. Thus, the CNN-QoE model is presented to overcome the computational complexity drawbacks, while at the same time improve QoE prediction accuracy. Based on a comprehensive evaluation, the CNN-QoE model can provide a high QoE prediction performance that outperforms the existing approaches.

3. The BiLSTM-QoE and CNN-QoE models focused on continuously predicting the instantaneous QoE which can provide the user's instant perceived video quality at a certain moment. However, in order to correctly determine the user's QoE for performing a QoE-based network control and management, it is necessary to produce a highly accurate QoE prediction either at any moment or at the end of a streaming session. Thus, the user's cumulative QoE can be potentially utilized as a better alternative than instantaneous QoE. The cumulative QoE is cumulatively estimated from the time when the viewer starts watching a streaming video content to any moment of the streaming session. A novel cumulative QoE prediction model is proposed that precisely assesses the user's cumulative perception. Moreover, in order to accurately model the user's cumulative QoE, the model also takes into account the human-related factors (i.e., memory effects, user's interest in the video content) which is not considered in the BiLSTM-QoE and CNN-QoE models. Evaluation results

indicated that the model performs excellently in predicting cumulative QoE at any moment within a streaming session.

1.3 Thesis Organization

The rest of this thesis is organized as follows:

- Chapter 2 provides a more detailed background on QoE in video streaming, QoE assessment methodologies, and the QoE influence factors that need to be considered in QoE modeling. It also covers the literature review of the methods used for measuring and modeling the user's QoE.
- Chapters 3, 4, and 5 presents three novel QoE prediction models. The work presented in these chapters has been published in [9], [10], and [11].
- After the three models have been presented, Chapter 6 discusses the feasibility to utilize these models for QoE-based network control and management in video streaming.
- Chapter 7 concludes the thesis and shows potential future research work.

Chapter 2

Background and Literature Review

This chapter provides the necessary background on QoE, its assessment methods, and the QoE influence factors in video streaming. It also presents the literature review of the approaches used for modeling the user's QoE.

2.1 QoE Definition

The definition of QoE proposed by European Network on QoE in Multimedia Systems and Service (the EU Qualinet community) is: "QoE is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state" [12, 13]. This definition has pointed out that QoE is a user-centric metric used to measure the user's satisfaction with a particular service and their perception of the service's quality.

2.2 QoE Assessment

In order to improve and monitor the user's QoE, video streaming service providers have to develop various techniques to measure the QoE. These techniques assess the user's QoE and quantify it in measurable metrics. The user's QoE can be assessed both subjectively and objectively.

Subjective QoE assessment involves users and requires user surveys to gather subjective evaluations of a given service. The most common method to capture subjective QoE evaluations is Mean Opinion Score (MOS) which is an ITU standardized [14]. MOS is a 5-point scale ranging from 1 to 5, which correlates to bad, poor, fair, good, and excellent. Despite

the high accuracy of the subjective assessment, it does have certain disadvantages. Firstly, a large number of participants need to be gathered in order to conduct the surveys. Secondly, this method is generally expensive in terms of cost and time consumption. Finally, it cannot be used for real-time QoE measurement or monitoring for video streaming.

Alternatively, objective QoE assessment is the most frequently used technique since it can be used to estimate the user's QoE without requiring human interaction. Objective QoE models are more suited for real-time QoE measurements. However, it can be less accurate than subjective QoE evaluations.

2.3 QoE Influence Factors

The user's QoE in video streaming is affected by multiple factors. These factors can be classified into the following four categories [15, 5]:

1. **System-related Influence Factors** consider the technical aspects of video streaming quality. They include the influences of the user device (e.g., computing power, screen size), network-related (e.g., bandwidth, delay), and also the application layer (e.g., video adaptation strategy, rebuffering events, bitrate switching events).

2. **Human-related Influence Factors** are related to the user's information and characteristic. They include psychological factors of the user such as user expectations, memory and recency effects, or user's background and usage history.

3. **Context-related Influence Factors** capture the environment and the context where the user views the video content. They address the user's location and space, subscription type, time of the day, the purpose of viewing the video, etc.

4. **Content-related Influence Factors** consider the characteristics of the video content. They include the influence of encoding rate, encoding format, resolution, playback duration, type of video, etc.

2.4 Literature Review on QoE Modeling

QoE modeling for video streaming services has received enormous attention due to its critical importance in QoE-aware applications. A number of different continuous QoE prediction models have been proposed [6, 16–19, 7, 20–22]. The authors in [6] modeled the time-varying subjective quality (TVSQ) using a Hammerstein-Wiener model. The work

in [19] proposed a model based on the augmented Nonlinear Autoregressive Network with Exogenous Inputs (NARX) for continuous QoE prediction. It should be noted that these models did not consider rebuffering events which usually happen in video streaming [23, 15]. On the other hand, the study in [18] took into account rebuffering events, perceptual video quality, and memory-related features for QoE prediction. However, the QoE prediction accuracy varied across different video streaming scenarios. The reason is that the model suffered from the difficulty in capturing the complex dependencies among QoE influence factors, leading to unreliable and unstable QoE prediction performances.

In order to address the above challenges, the authors in [1] proposed a QoE prediction model, namely, LSTM-QoE, which was based on Long Short-Term Memory networks (LSTM). The authors argued that the continuous QoE is dynamic and time-varying in response to QoE influencing events such as rebuffering [24] and bitrate adaptation [6]. To capture such dynamics, LSTM was employed and the effectiveness in modeling the complex temporal dependencies in sequential QoE data was shown. The model was evaluated on different QoE databases and outperformed the existing models in terms of QoE prediction accuracy. However, the computational complexity of the model was not fully inspected. Since the recurrent structure in LSTM can only process the task sequentially, the model failed to effectively utilize the parallel computing power of modern computers, leading to a high computational cost. Therefore, the feasibility to utilize the LSTM-QoE model for real-time video streaming applications remains an open question.

Therefore, in this thesis, we aim to improve the QoE prediction performance of QoE prediction models by enhancing the QoE prediction accuracy and optimizing the computational complexity.

Chapter 3

BiLSTM-QoE Model

3.1 Introduction

Recently, huge research efforts have been carried out to utilize the Long Short-Term Memory (LSTM) approach to model and predict several types of sequential data. The work in [1] was one of the first studies to apply the LSTM to QoE prediction in video streaming. The study proposed a continuous QoE prediction model, namely LSTM-QoE, utilizing the LSTM to capture the nonlinearities and the complex temporal dependencies in sequential QoE data. The LSTM model can leverage past events occurring during a streaming session by passing them through its chain-like gated structure. Although having shown promising results, there also have been suspicions that useful information may be filtered out or not efficiently passed through the chain-like structure of the unidirectional LSTM model since it only forwards the information within a single direction (from past to future). There is a high possibility that useful information may be missed out or not effectively forwarded [25]. Therefore, it is highly necessary to take into consideration backward direction (from future to past) as well, or, in other words, a bi-directional model.

The idea aligns well with the impact of memory on human perception: when conducting subjective QoE tests, users often recall previous events happening during the streaming session that have a significant influence on his/her satisfaction. This shows the potential of using a Bidirectional LSTM (BiLSTM) model to improve the accuracy of continuous QoE prediction.

In this chapter, the BiLSTM-QoE is presented. First, Section 3.2 introduces the architecture of the BiLSTM-QoE model. Then, the evaluation results of the model is presented in Section 3.3. Section 3.4 then discusses the advantages and disadvantages of the model. Finally, Section 3.5 summaries this chapter.

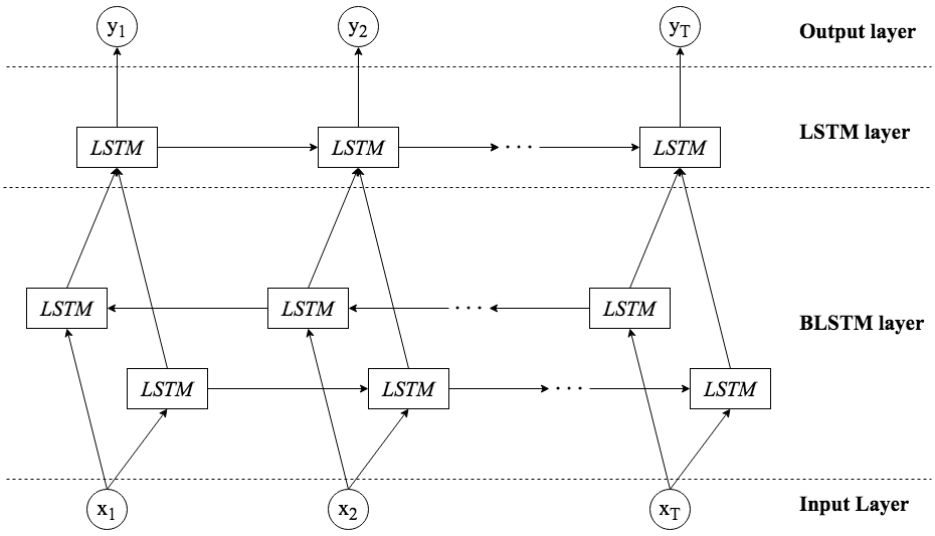


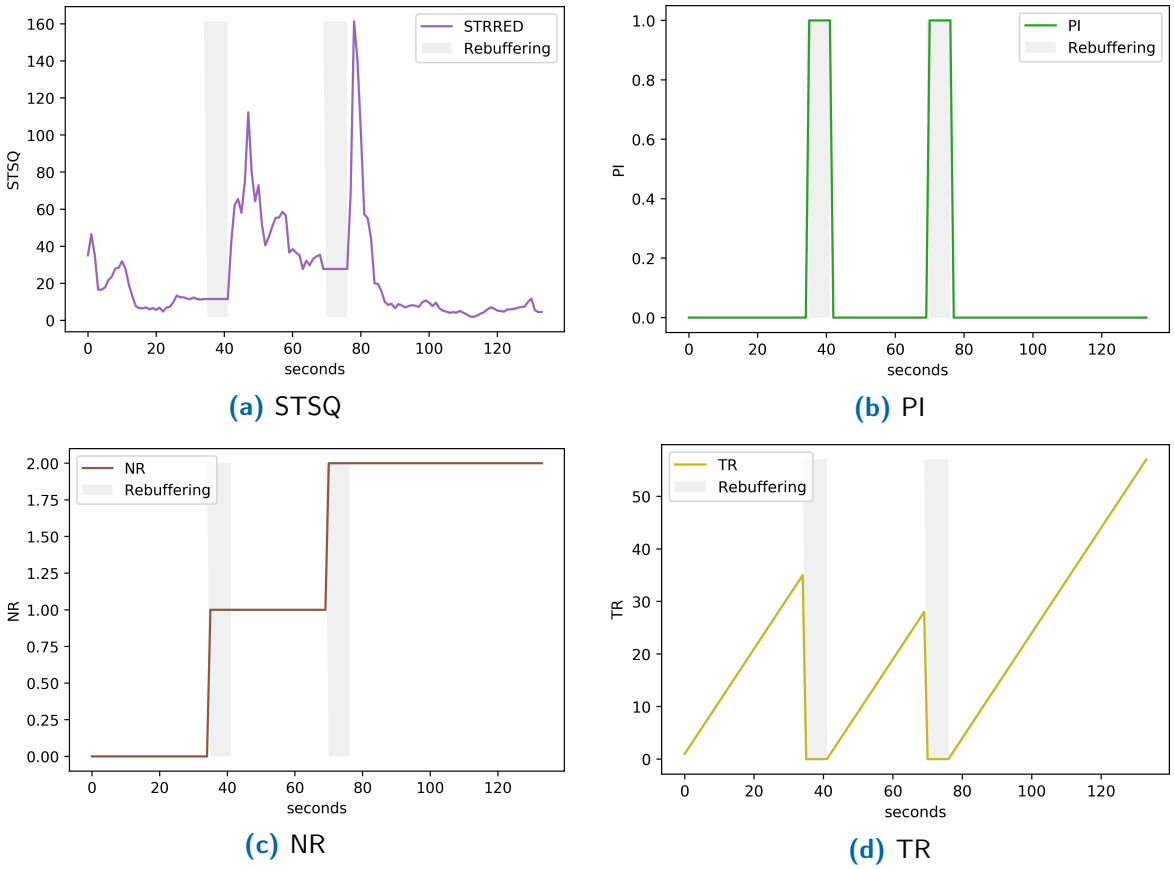
FIGURE 3.1 The proposed BiLSTM-QoE architecture.

3.2 Model Architecture

The work in [1] predicts the user's QoE using LSTM which is a special kind of Recurrent Neural Networks. With recurrent connections in the hidden layer, LSTM can capture temporal dependencies in sequential QoE data. However, it is very complex to rely on a single LSTM to capture these dependencies and the complex interactions among QoE influence factors such as video quality, bitrate switching, rebuffering events. Moreover, unidirectional LSTM structure only considers forward dependencies, thus, useful information may be filtered out. Therefore, a novel QoE prediction model based on BiLSTM is taken into account.

BiLSTM processes sequential features in both forward and backward directions, thereby it can access temporal dependencies in both directions and result higher level of representations of input features. The bidirectional networks have been proved that it has better performance compare with unidirectional ones in many fields [25–27].

The proposed architecture of the BiLSTM-QoE model, illustrated in Figure 3.1, consists of two main layers: BiLSTM layer and LSTM layer. BiLSTM layer connects two separate hidden layers to learn the sequential input features in two directions (forward and backward). The output of BiLSTM layer will be fed into the LSTM layer to predict the instantaneous QoE.



(a) STSQ (b) PI

(c) NR (d) TR

FIGURE 3.2 Example of rebuffering and bitrate-related features represented by STSQ, PI, NR, and TR

3.3 Performance Evaluation

3.3.1 Input Features for QoE Prediction

Video streaming users are sensitively affected by the video quality, known as *short time subjective quality* (STSQ) [6]. STSQ is defined as the visual quality of video being rendered to the user and can be predicted using any of the robust video quality assessment (VQA) metrics, such as Spatio-Temporal Reduced Reference Entropic Differences (STRRED) [28], Multi-Scale Structural Similarity (MS-SSIM) [29], Peak Signal to Noise Ratio (PSNR) [30], etc. Recent experiments have demonstrated that STRRED is a robust and high-performing VQA model when being tested on a very wide spectrum of video quality datasets, on multiple resolution and device types [31, 20, 1]. Therefore, STRRED is utilized to measure the STSQ.

Rebuffering greatly impacts the user's QoE [32]. Thus, rebuffering information such as rebuffering length, rebuffering position and the number of rebuffering events must be investigated. As a result, two rebuffering-related inputs are employed. Firstly, *playback*

indicator (PI) [18, 20, 1] is defined as a binary continuous-time variable, specifying the current playback status, i.e., 1 for rebuffering and 0 for normal playback. Secondly, as the user's annoyance increases whenever a rebuffering event occurs [32], the *number of rebuffering events* (NR) happened from the start to the current time instant of the session is considered.

Besides, the user's QoE is also affected by memory factors. For example, more recent experiences have larger impacts on the user's perceived video quality, known as the recency effect [33, 8, 34]. To capture the relation between the recency effect and the user's QoE, *time elapsed since the last video impairment* (i.e., bitrate switch or rebuffering occurrence) [18, 20, 1], denoted as TR, is utilized.

All the considered QoE influence factors are fed to the BiLSTM-QoE model to predict the instantaneous QoE. The examples of four factors (including STSQ, PI, NR, and TR) are illustrated in Figure 3.2.

3.3.2 LIVE Netflix Video QoE Database

The BiLSTM-QoE model is evaluated on the LIVE Netflix Video QoE Database [8]. The database consists of 112 distorted videos generated from 14 video contents by 8 different playout patterns including bitrate changing, rebuffering events and mixtures of both [8]. The training and testing strategy are defined in [1]. For each video j in the database, one train-test set is created, the model is trained on the set of videos that do not have the same content and the playout pattern as the video j in the test set. Therefore, there are 112 train-test sets, each contains 1 testing videos and 91 training videos (excludes 14 videos with the same content and 7 with the same playout pattern). With this strategy, content and pattern dependencies are eliminated.

TABLE 3.1 QoE prediction performance of the BiLSTM-QoE over the LIVE Netflix Video QoE Database.

	PCC	SROCC	RMSE
BiLSTM-QoE	**0.894**	**0.830**	**7.43**
LSTM-QoE [1]	0.802	0.714	7.78
NLSS-QoE [20]	0.655	0.483	16.09
NARX [18]	0.621	0.557	8.52

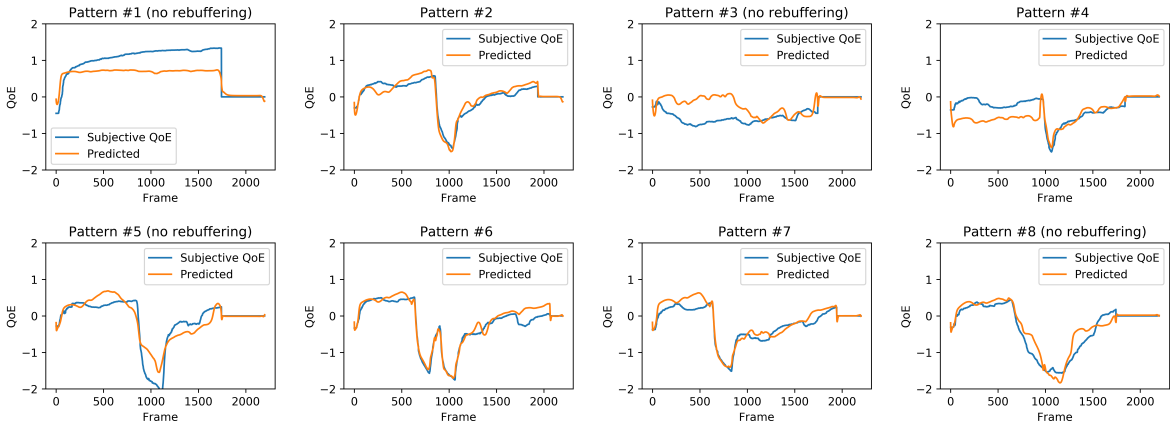


FIGURE 3.3 QoE prediction performance of the BiLSTM-QoE over the LIVE Netflix Video QoE Database.

3.3.3 Evaluation Results

The BiLSTM-QoE model was trained and tested as described above. The QoE prediction on the database using 4 features (i.e. STSQ, PI, TR and NR) are illustrated in Figure 6. The mean QoE prediction performance results are tabulated in Table 3.1, also compared with other QoE models in the same database. There are three considered evaluation criteria: Pearson Correlation Coefficient (PCC), Spearman Rank Correlation Coefficient (SROCC), and Root Mean Squared Error (RMSE). The proposed model outperforms LSTM-QoE [1], NLSS-QoE [20] and NARX [18] in terms of QoE prediction accuracy. This is because the BiLSTM networks are very suitable to learn more useful information from time-varying features, thereby allows this model to achieve the best performance among all the referenced models.

3.4 Discussion

The results demonstrate that the proposed model using BiLSTM, measuring both forward and backward directions, is capable of learning more useful information from QoE influence factors such as video quality, bitrate switching, and rebuffering events. For example, in pattern #1 (which is ideal circumstances where no rebuffering or bitrate changes occurred), #2 (one rebuffering occurred), #6 (two rebufferings occurred close together), the QoE prediction performance is very good and the trends of the predicted QoE are pretty similar with the subjective QoE. It has proved that the model can adapt well to different scenarios occurring during a streaming session. However, the accuracy of the model may fluctuate across different playout patterns in the database. For example, the accuracy on patterns without rebuffering

(#1, #3, #5, #8) are relatively worse, due to the fact that among four features, only the STSQ values are nonzero, thereby, the prediction model depends only on STSQ inputs that may hurt the accuracy. Thus, it is necessary to consider the other general valuable features whose values are nonzero in any scenarios, to achieve higher prediction accuracy.

3.5 Summary

In this chapter, the BiLSTM-QoE model was presented to improve the QoE prediction accuracy. Unlike unidirectional LSTM only measuring forward direction, BiLSTM considers both forward and backward dependencies, thereby it has the ability to avoid filtering out the useful information. Despite the various QoE influence factors and the complex interactions among them, the model can provide an accurate prediction user's QoE.

In the next chapter, the CNN-QoE model is introduced to further improve the QoE prediction accuracy. Furthermore, the model also focus on minimizing the computational complexity, resulting in a QoE prediction model that can be utilized in real-time QoE modeling and continuous monitoring.

Chapter 4

CNN-QoE Model

4.1 Introduction

The BiLSTM-QoE model achieved high accuracy since it is capable of capturing temporal dependencies in sequential QoE data. However, the chain structure in LSTM layers requires a high computational cost for practically predicting the user's QoE due to the use of sequential processing over time. It means that the subsequent processing steps must wait for the output from the previous ones. This leads to an open question about the performance of the model on real-time QoE-based monitoring and management.

Recently, Temporal Convolutional Network (TCN) [35], a variation of Convolutional Neural Network (CNN), has emerged as a promisingly alternative solution for the sequence modeling tasks. TCN adopts dilated causal convolutions [36–38] to provide a powerful way of extracting the temporal dependencies in the sequential data. Different from LSTM, the computations in TCN can be performed in parallel, providing computational and modeling advantages. In practical deployments, TCN convincingly outperforms canonical recurrent architectures including LSTMs and BiLSTMs across a broad range of sequence modeling tasks [35]. Enlightened by the great ability of TCN, an improved TCN-based model, namely QoE-CNN, is introduced for improving the QoE prediction accuracy and optimizing the computational complexity.

The remainder of this chapter is organized as follows: Section 4.2 discusses the TCN architecture in detail. The proposed model is presented in Section 4.3. Section 4.4 and 4.5 provide evaluation results of the proposed model and their discussion, respectively. Finally, this chapter is concluded in Section 4.6.

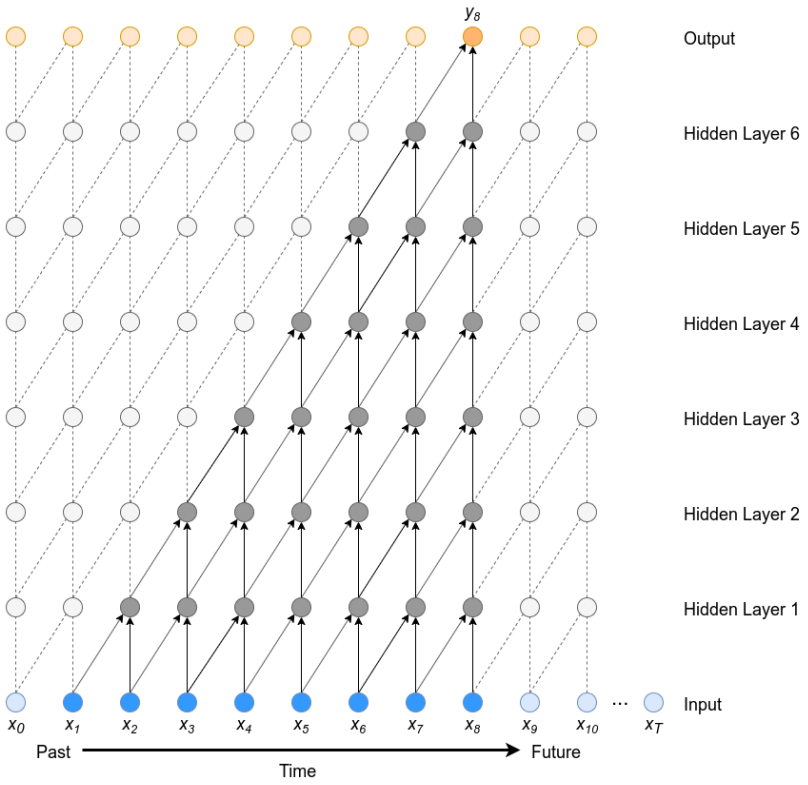


FIGURE 4.1 An illustration of a stack of causal convolution layers with the convolution filter size of 1×2.

4.2 Temporal Convolutional Network

In this section, TCN architecture is briefly discussed to summarize its advantages and disadvantages in sequence modeling tasks. Thereby, the conclusions of this section will be the crucial foundation for the subsequent improvements proposed in CNN-QoE, which are stated in Section 4.3.

4.2.1 1D Convolutions

CNN was traditionally designed to operate on two dimensions (2D) data such as images. An input image is passed through a series of 2D convolution layers. Each 2D convolution applies and slides a number of 2D filters through the image. To adapt CNN for time-series data, TCN utilizes 1D convolution where the filters exhibit only one dimension (time) instead of two dimensions (width and height). Concretely, a time-series input is convolved with a filter size of $1 \times k$.

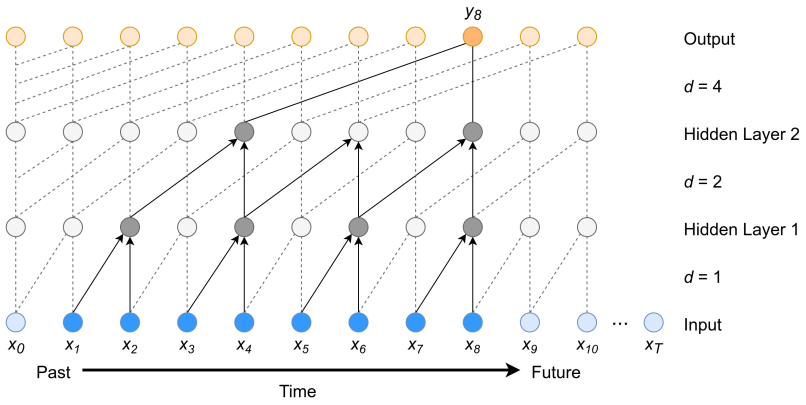


FIGURE 4.2 An illustration of a stack of dilated causal convolution layers with the convolution filter size of 1×2.

Furthermore, 1D convolutions are well-suited for real-time tasks due to their low computational requirements. 1D convolutions require simple array operations rather than matrix operations, hence, the computational complexity is significantly reduced in comparison with 2D convolutions. In addition, the convolution operations allow fully parallel processing, resulting in a significant improvement of computational speed.

4.2.2 Causal Convolutions

A *causal convolution* is a convolution layer to ensure there is no information "leakage" from future into past. In other words, given an time-series input $x_0, ..., x_T$, the predicted output $\widehat{y}_t$ at a time instant t depends only on the inputs at time t and earlier $\mathbf{x}_t, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-r+1}$. For instance, as illustrated in Figure 4.1, the predicted $\widehat{y}_8$ is computed by a combination of the inputs $x_1, ..., x_8$. It can be observed that, in order to achieve a long effective history size or a large receptive field size, an extremely deep network or very large filters are needed, which significantly increases the model computational complexity. Thus, TCN architecture utilizes dilated causal convolutions rather than causal convolutions. The advantages and disadvantages of dilated causal convolutions are discussed below.

4.2.3 Dilated Causal Convolutions

TCN adopts a dilated causal convolution comprising of the causal and the dilated convolutions. The causal convolution has already been described in the previous subsection. Meanwhile, dilated convolution [36–38] is a convolution where the convolution filter is applied to a larger area than its length by skipping input values with several steps. Therefore, the dilated causal convolution can effectively allow the network to operate on a larger scale than the one with a

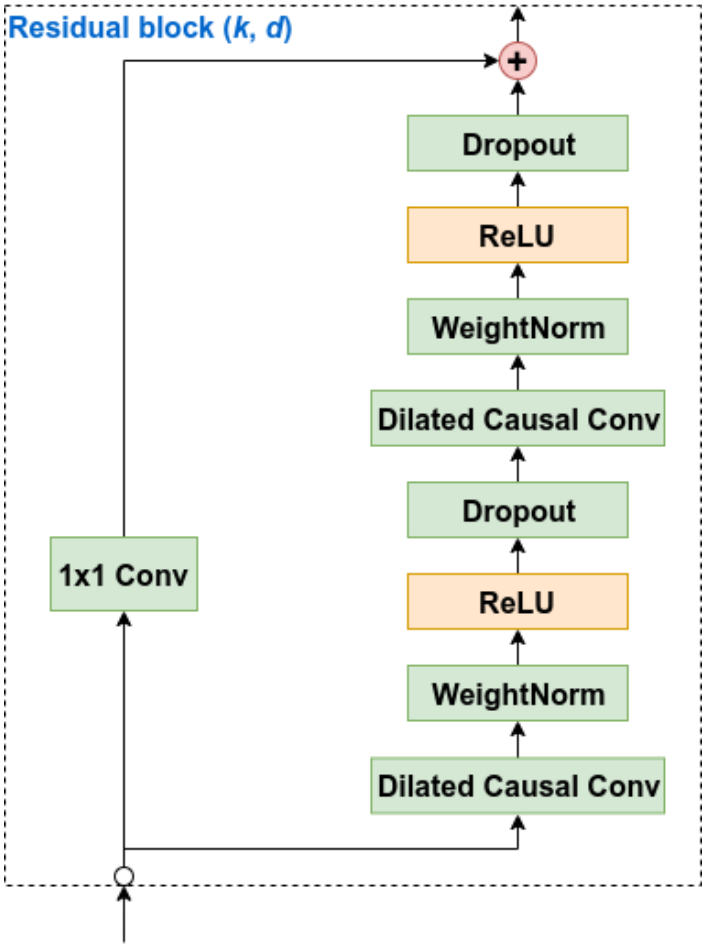


FIGURE 4.3 The residual block in TCN architecture.

normal convolution while ensuring that there is no leakage of information from the future to the past. The dilated causal convolution is defined as:

$$D(t) = \sum_{i=0}^{k-1} f(i) \cdot \mathbf{x}_{t-d \cdot i} \tag{4.1}$$

where, d is the dilation factor, f is a filter size of $1 \times k$. d exponentially increases with the depth of the network (i.e., $d = 2^l$ at layer l of the network). For instance, given the network with L layers of dilated causal convolutions $l = 1, ..., L$, the dilation factors exponentially increase by a factor of 2 for every layer:

$$d \in [2^0, 2^1, ..., 2^{L-1}] \tag{4.2}$$

Figure 4.2 depicts a network with three dilated causal convolutions for dilations 1, 2, and 4. Using the dilated causal convolutions, the model is able to efficiently learn the connections between far-away time-steps in the time series data. Moreover, as opposed to causal convolutions in Figure 4.1, the dilated causal convolutions require fewer layers even though the receptive field size is the same. A stack of dilated causal convolutions enables the network to have a very large receptive field with just a few layers, while preserving the computational efficiency. Therefore, dilated causal convolutions reduce the total number of learnable parameters, resulting in more efficient training and light-weight model.

However, the dilated causal convolutions have problem with local feature extraction. As shown in Figure 4.2, it can be seen that the filter applied to the time-series input is not overlapped due to the skipping steps of the dilation factor. As long as the dilation factor increases, the feature is extracted from only far-apart time-steps, but not from adjacent time-steps. Therefore, the local connection among adjacent time-steps is not fully extracted at higher layers.

4.2.4 Residual Block

The depth of the model is important for learning robust representations, but also comes with a challenge of vanishing gradients. The residual block has been found to be an effective way to address this issue and build very deep networks [39]. A residual block contains a series of transformation functions F, whose outputs are added to the input $\mathbf{x}$ of the block:

$$o = Activation(x + F(x)) \tag{4.3}$$

The residual block is used between each layer in TCN to speed up convergence and enable the training of much deeper models. The residual block for TCN is shown in Figure 4.3. It consists of dilated causal convolution, ReLU activation function [40], weight normalization [41], and spatial dropout [42] for regularization. Having two layers of dilated causal convolution in the TCN's residual block is suitable for complex challenges such as speech enhancement [35]. Compared with speech signal data, sequential QoE data is much simpler. That is to say, the two layers of dilated causal convolution are redundant and are not optimal for the QoE prediction problem.

In TCN architecture, equations (4.1) and (4.2) suggest that the TCN model heavily depends on the network depth L and the filter size k.

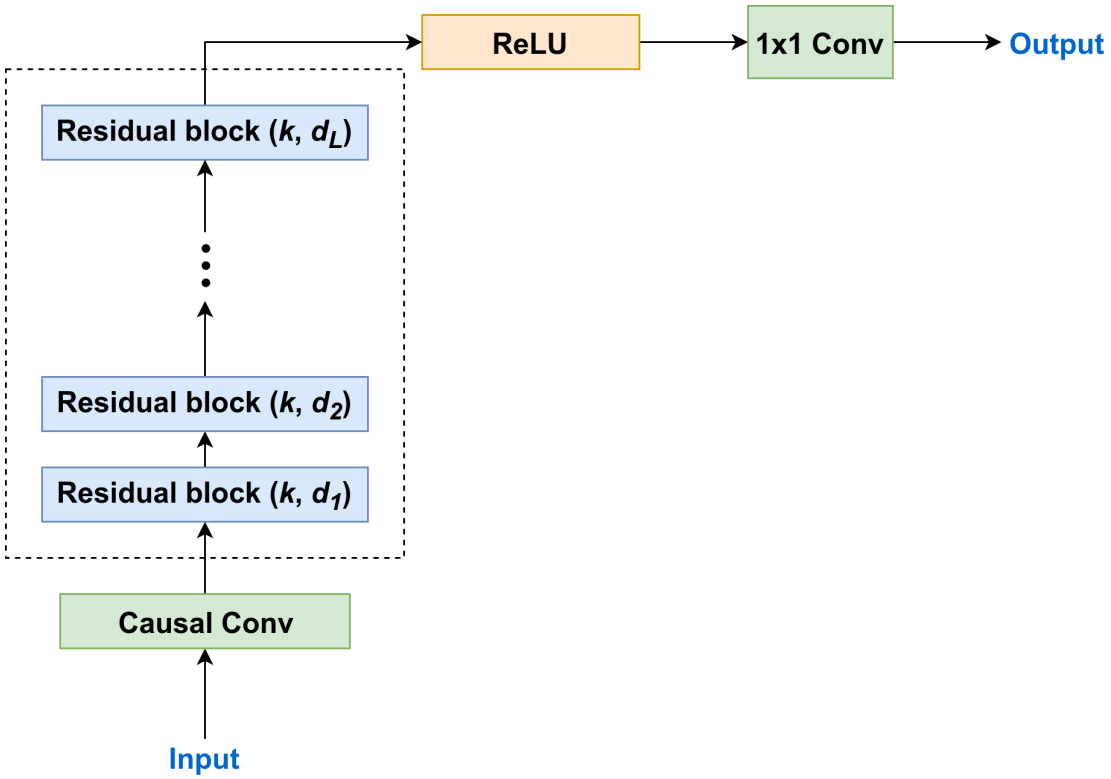


FIGURE 4.4 The proposed CNN-QoE architecture.

4.3 Model Architecture

In this section, the model CNN-QoE is introduced to leverage the advantages and handles the problems of the TCN architecture [35] in QoE prediction for video streaming services. The architecture employed for the CNN-QoE model is discussed in detail in the following subsections. The model architecture hyperparameters are then analyzed to find the optimal values which can improve the QoE prediction accuracy, while minimizing the computational complexity.

4.3.1 Proposed Model Architecture

Figure 4.4 illustrates the overview of the CNN-QoE's architecture. The CNN-QoE leverages the advantages of 1D convolutions, dilated causal convolutions and residual block in TCN architecture. To adapt the TCN to QoE prediction tasks, a number of improvements are made as follows:

- An initial causal convolution layer is added to the input and then connects to residual block which includes a dilated causal convolution layer.

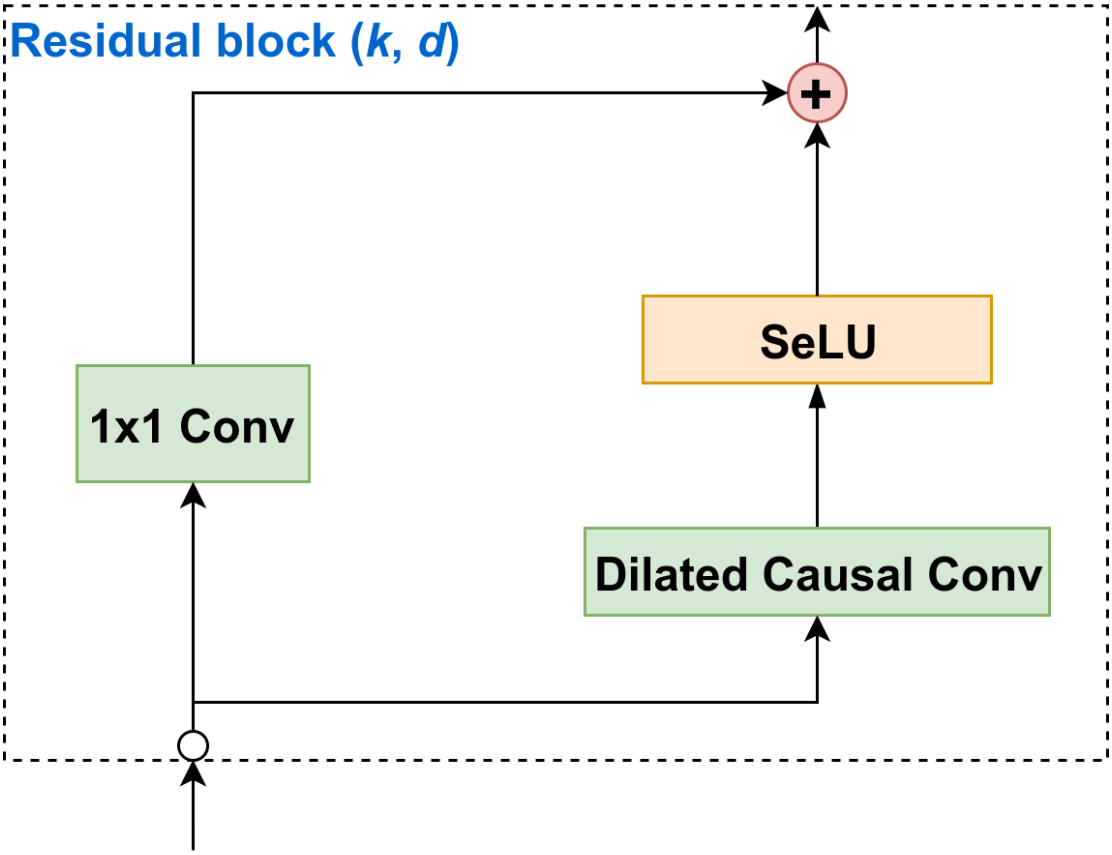


FIGURE 4.5 The proposed residual block used in the proposed architecture.

- The residual block is simplified by leveraging the advantages of Scaled Exponential Linear Units (SeLU) activation function [43].

These distinguishing characteristics are discussed as below.

4.3.1.1 Causal convolution to extract local features

The architecture of the proposed model comprises of one causal convolution layer and a stack of dilated causal convolutions, while the TCN consists of only a number of dilated causal convolutions. A causal convolution layer is added between the input time-series and the first residual block as shown in Figure 4.4. This causal convolution layer can extract the local features of the adjacent time-steps in the sequential QoE data. Afterward, the following dilated causal convolution layers are leveraged to extract the global features between far-apart time steps. These layers help the model to learn the most informative features in the time series input, resulting in higher accuracy.

TABLE 4.1 Hyperparameters for the best performance model.

Architecture Hyperparameters	Description	Derived Value
r	Receptive field size	8
k	Filter size	2
L	Number of dilated causal convolution layers	3
n	Number of filters on each convolution layer	32

4.3.1.2 SeLU activation function

Activation function plays an important role in allowing the model to learn non-linear representations of the input features. When training a deep learning model, the vanishing and exploding gradient are the most challenging problems that prevent the network from learning the optimal solution. The TCN model gets rid of these problems by integrating ReLU activation function [40], weight normalization [41] and dropout [42] layer as shown in Figure 4.3. In the proposed CNN-QoE, those layers are replaced with the SeLU to leverage its advantages and simplify the residual block as shown in Figure 4.5. SeLU is a self-normalizing activation function. It converges to zero mean and unit variance when propagated through multiple layers during network training, thereby making it unaffected by vanishing and exploding gradient problems. Moreover, SeLU also solves the "dying ReLU" problem where the ReLU function always outputs the same value of 0 for any input, so the gradient descent is not able to alter the learnable parameters. At the same time, SeLU also reduces the training time and learns robust features more efficiently than other networks with normalization techniques, such as weight normalization [43]. SeLU activation function described as follow [43]:

$$SeLU(x) = \lambda \begin{pmatrix} x, & if\, x > 0 \\ \alpha exp(x) - \alpha, & if\, x \leq 0 \end{pmatrix} \tag{4.4}$$

where $\alpha = 1.67733$ and $\lambda = 1.0507$. These are the same values as the ones proposed in [43].

4.3.2 Architecture Hyperparameters Selection

When training the model, an adequate set of architecture hyperparameters must be selected to achieve the best performance. The proposed model consists of L residual block layers, each layer contains a dilated causal convolution with a filter size of $1 \times k$, as shown in Figure 4.4 and 4.5. Each dilated convolution layer has a dilation factor d doubled at each layer up, as shown in (4.2). The proposed model depends on the network depth L and the filter

size k. These hyperparameters control the trade-off between QoE prediction accuracy and computational complexity of the model. To effectively optimize the hyperparameters, it is important to set a boundary for the space of possible hyperparameter values.

The user's QoE is mostly affected by the recent experiences, also known as the recency effect [33, 8, 34]. The recency effect gradually decreases within 15 to 20 seconds [8, 34] after distorted events (e.g., bitrate fluctuations, rebuffering events). Therefore, the effective history or the receptive field size r of the model cannot be larger than 20 time-steps

$$r \leq 20 \tag{4.5}$$

Moreover, the receptive field depends on the number of dilated causal convolution layers L and the filter size k. For example, with $l \in [1, L]$, the receptive field r can be determined by (4.6) [44, 35]

$$r = 2^L, \text{if} \quad k = 2 \tag{4.6}$$

or (4.7) [45]

$$r = 2^{L+1} - 1, \text{if} \quad k = 3 \tag{4.7}$$

Figure 4.2 shows an example of a three-layer ($L = 3$) dilated convolutional network. In this figure, given the filter size of 1×2 ($k = 2$), the receptive field is computed by $r = 2^3 = 8$. From (4.5), (4.6), and (4.7), the range of L values can easily be defined $L \in [2, 3, 4]$.

In a 1D convolution, the number of filters n is also important to effectively extract the information from the inputs. To minimize the computation complexity of the model, the range of n is set to $n \in \{16, 32, 64\}$. We conduct a simple grid-search of the model architecture hyperparameters with $k \in [2, 3]$, $L \in [2, 3, 4]$, and $n \in \{16, 32, 64\}$. Table 4.1 shows the values of r, k, L, and n that achieves the best performance.

4.4 Performance Evaluation

In this section, we evaluate the performance of the CNN-QoE in terms of QoE prediction accuracy and computational complexity. The evaluation is performed by comparing the proposed model with numerous baseline models across multiple databases. In the following subsections, firstly, a brief explanation of baseline models is showed. Then, the evaluation results on accuracy and computational complexity are presented. Finally, the overall performance of the proposed model is discussed to illustrate its capability for real-time QoE prediction.

TABLE 4.2 An overview of the three public QoE databases used in the proposed model evaluation.

Database	Device Type	Rebuffering Events	Bitrate Fluctuations	Duration	QoE Range
LFOVIA Video QoE Database	TV	yes	yes	120 secs	[0, 100]
LIVE Mobile Stall Video Database II	Mobile	yes	no	29-134 secs	[0, 100]
LIVE Netflix Video QoE Database	Mobile	yes	yes	at least 1 minute	[-2.28, 1.53]

4.4.1 Baseline Models

To evaluate the QoE prediction accuracy of the proposed model, the comparison with the state-of-the-art QoE models comprising of LSTM-QoE [1], NLSS-QoE [20], SVR-QoE [34], and NARX [18] will be performed. It is worth noting that we also make a comparison with the original TCN model, or TCN-QoE for short, in the QoE prediction task. The TCN-QoE model uses the same network hyperparameters and input features with ones described in Section 4.3.2 and 3.3.1.

To evaluate the computational complexity of the proposed model, we focus on the comparison with deep learning-based QoE prediction models since they achieve exceptionally higher accuracy. Particularly, LSTM-QoE [1] and TCN-QoE are utilized in the comparison. It is important to note that the LSTM-QoE [1] model hyperparameters are employed as reported in its respective works in order to ensure a fair comparison.

4.4.2 Accuracy

4.4.2.1 Databases

There are three public QoE databases used for the evaluation of QoE prediction accuracy, including LFOVIA Video QoE Database [34], LIVE Netflix Video QoE Database [8], and LIVE Mobile Stall Video Database II [32]. The descriptions of these databases are summarized in Table 4.2.

To evaluate the QoE prediction accuracy, the evaluation procedures performed on each database are described as follows:

- LFOVIA Video QoE Database [34] consists of 36 distorted video sequences of 120 seconds duration. The training and testing procedures are performed on this database in the same way as the one described in [1]. The databases are divided into different train-test sets. In each train-test sets, there is only one video in the testing set, whereas

the training set includes the videos that do not have the same content and playout pattern as the test video. Thus, there are 36 train-test sets, and 25 of 36 videos are chosen for training the model for each test video.

- LIVE Netflix Video QoE Database [8]: The same evaluation procedure as described for LFOVIA Video QoE Database is employed. There are 112 train-test sets corresponding to each of the videos in this database. In each train-test set, the training set consists of 91 videos out of a total of 112 videos in the database (excludes 14 with the same playout pattern and 7 with the same content).
- LIVE Mobile Stall Video Database II [32]: The evaluation procedure is slightly different from the one applied to the above databases. Firstly, 174 test sets corresponding to each of 174 videos in the database are created. For each test set, since the distortion patterns are randomly distributed across the videos, randomly 80% videos from the remaining 173 videos are then chosen for training the model and perform evaluation over the test video.

4.4.2.2 Evaluation Settings

To evaluate the accuracy of the proposed model, the model hyperparameter sets and input features are used as described in Section 4.3.2 and 3.3.1, respectively. The QoE prediction performance of the proposed model is compared with baseline models described in Section 4.4.1.

4.4.2.3 Evaluation Criteria

Three evaluation metrics, namely, Pearson Correlation Coefficient (PCC), Spearman Rank Order Correlation Coefficient (SROCC) and Root Mean Squared Error (RMSE) are considered for QoE prediction accuracy assessment. The SROCC measures the monotonic relationship, while PCC measures the degree of linearity between the subjective and the predicted QoE For PCC and SROCC, a higher value illustrates a better result, while for the RMSE, the lower value is better.

4.4.2.4 Results

Figures 4.6, 4.7 and 4.8 illustrate the QoE prediction performance over the three databases using the proposed CNN-QoE model. In general, the proposed model produces superior and consistent QoE prediction performance in different situations with and without rebuffering events. Patterns #1-#3 in Figure 4.6, 4.7, and 4.8 show that the proposed model can effectively

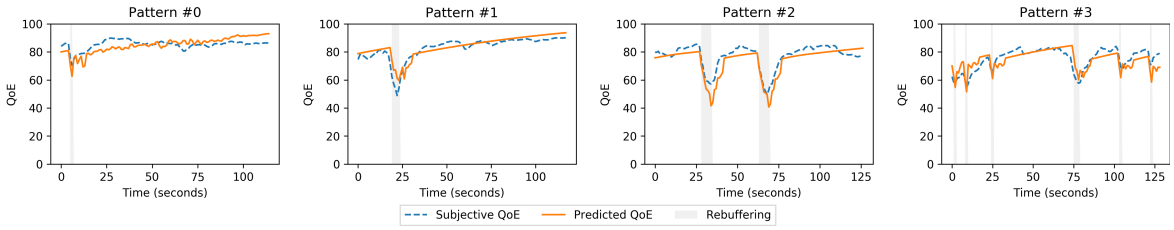


FIGURE 4.6 QoE prediction performance of the CNN-QoE over the LFOVIA Video QoE Database.

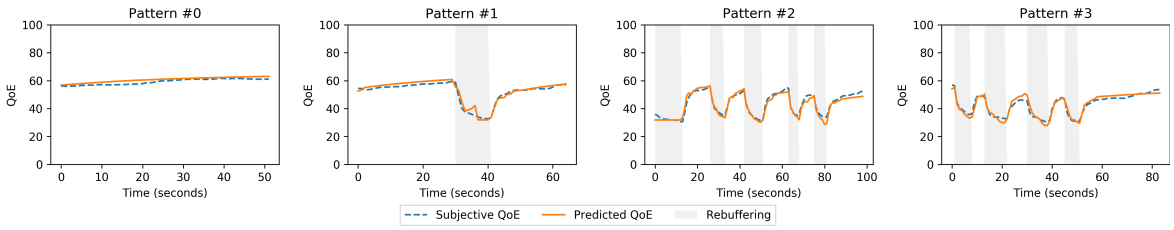


FIGURE 4.7 QoE prediction performance of the CNN-QoE over the LIVE Mobile Stall II Video Database.

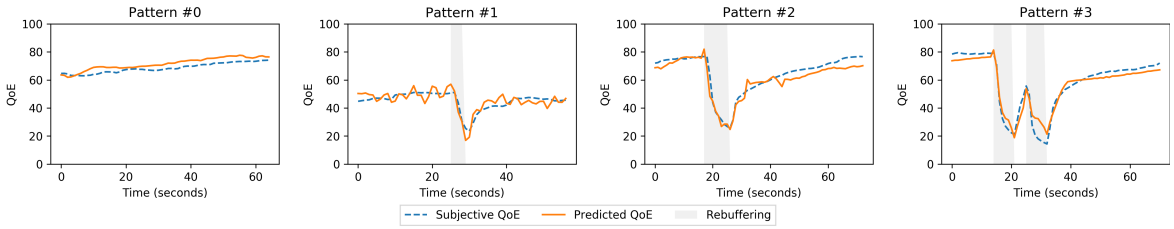


FIGURE 4.8 QoE prediction performance of the CNN-QoE over the LIVE Netflix Video QoE Database.

capture the effect of rebuffering events on the user's subjective QoE. Especially, even the rebuffering event repeatedly occurs as illustrated in pattern #3 in Figure 4.6 and patterns #2, #3 in Figure 4.7, the QoE predictions still correlate well with the subjective QoE. Meanwhile, pattern #0 in Figure 4.6 and pattern #1 in Figure 4.8 show some fluctuations in the predicted QoE. However, the amplitudes of these fluctuations are small and the varying trends in the subjective QoE are still adequately captured by the proposed model. Additionally, a linear trend is subsequently introduced in the predicted QoE after each rebuffering event as shown in patterns #1 - #3, Figure 7. It means that the model is not overfitting and can be trained on the LFOVIA Video QoE Database with larger epochs to further increase its nonlinearity and QoE prediction accuracy.

The QoE prediction performance results over each database in comparison with existing models are shown in the Tables 4.3, 4.4 and 4.5. It is important to note that the Hammerstein-Wiener model in [22] was employed as reported in their work in order to ensure a fair

TABLE 4.3 QoE prediction performance of the CNN-QoE over the LFOVIA Video QoE Database.

	PCC	SROCC	RMSE (%)
CNN-QoE	**0.820**	**0.759**	**4.81**
TCN-QoE	0.670	0.732	5.47
LSTM-QoE [1]	0.800	0.730	9.56
NLSS-QoE [20]	0.767	0.685	7.59
SVR-QoE [34]	0.686	0.648	10.44

TABLE 4.4 QoE prediction performance of the CNN-QoE over the LIVE Mobile Stall Video Database II. Boldface indicates the best result.

	PCC	SROCC	RMSE (%)
CNN-QoE	**0.892**	**0.885**	**5.36**
TCN-QoE	0.667	0.603	9.71
LSTM-QoE [1]	0.878	0.862	7.08
NLSS-QoE [20]	0.680	0.590	9.52

comparison. From these tables, it is revealed that the CNN-QoE outperforms the existing QoE models within all the criteria, especially in terms of RMSE. Moreover, the accuracy produced by CNN-QoE is consistent across the databases, thus marking it as an efficient comprehensive model. The results illustrate that the CNN-QoE architecture is capable of capturing the complex inter-dependencies and non-linear relationships among QoE influence factors. Interestingly, there is a significant improvement in QoE prediction accuracy when comparing with TCN-QoE. It means that the enhancements in the proposed architecture have made the model more suitable for QoE prediction.

4.4.3 Computational Complexity

In this subsection, the computational complexity of the CNN-QoE model is investigated to show its effectiveness in comparison with baseline methods including TCN-QoE and LSTM-QoE [1]. These models are trained and tested on the LFOVIA Video QoE Database with a training:test ratio of 80:20.

4.4.3.1 Evaluation Settings

For running these deep learning-based QoE prediction models, a personal computer running 18.04 Ubuntu LTS with an Intel i7-8750H @ 2.20GHz and 16GB RAM system is used. It should be noted that the GPU computation power is not utilized on the personal computer.

TABLE 4.5 QoE prediction performance of the CNN-QoE over the LIVE Netflix Video QoE Database. Boldface indicates the best result.

	PCC	SROCC	RMSE (%)
CNN-QoE	**0.848**	**0.733**	**6.97**
TCN-QoE	0.753	0.720	7.62
LSTM-QoE [1]	0.802	0.714	7.78
NLSS-QoE [20]	0.655	0.483	16.09
NARX [18]	0.621	0.557	8.52
[22]	0.611	0.515	18.96

TABLE 4.6 Computational complexity of the CNN-QoE on the personal computer.

	Inference Time (ms)	Model Size (kB)	FLOPs	Number of Parameters
CNN-QoE	**0.673**	82.08	175,498	9,605
TCN-QoE	0.856	145.86	253,076	13,766
LSTM-QoE [1]	1.996	**41.18**	**30,953**	**6,364**

4.4.3.2 Evaluation Criteria

The following four evaluation metrics are considered:

- Inference time: the time taken to predict the user QoE at any given time instant t.
- Model size: the storage size of the trained model on the hard drive.
- FLOPs: the number of operations performed.
- Number of Parameters: number of learnable parameters in the model.

4.4.3.3 Results

Table 4.6 show the computational complexity results of the proposed CNN-QoE compared to the TCN-QoE and LSTM-QoE. In general, the CNN-QoE requires a higher number of parameters and FLOPs in comparison with LSTM-QoE to achieve higher accuracy. Although the FLOPs of the proposed model are larger, the inference time is 3 times faster than the LSTM-QoE model. This indicates that the proposed model can efficiently leverage the power of parallel computation to boost up the computing speed. It can be seen from Table 4.6 that the architecture complexity of TCN-QoE is extremely higher than our proposed CNN-QoE model in terms of number of parameters and FLOPs. However, the accuracy of TCN-QoE is not quite comparable with the CNN-QoE as shown in Tables 4.3, 4.4, and 4.5 . It proves

that the proposed improvement adapted on the original TCN architecture allow CNN-QoE to effectively capture the complex temporal dependencies in a sequential QoE data.

4.4.4 Overall Performance

Accurate and efficient QoE prediction models provide important benefits to the deployment and operation of video streaming services. As shown in subsection 4.4.2 and 4.4.3, the proposed model CNN-QoE can achieve not only the state-of-the-art QoE prediction accuracy but also the reduction on computational complexity. Therefore, the CNN-QoE can be an excellent choice for future QoE prediction systems or QoE-aware video streaming applications.

4.5 Discussion

According to the above-mentioned evaluation results, it can be seen that the proposed model completely outperforms TCN-QoE where the original TCN architecture is adopted in the QoE prediction task. Thereby, it generally demonstrates the efficiency of the proposed improvements upon the original TCN architecture in QoE prediction for video streaming services. In the following subsections, the effects of the improvements including the interactions between causal convolutions and dilated causal convolutions are discussed in detail.

4.5.1 Effects of comprising causal convolutions and dilated causal convolutions

Different from TCN [35] architecture, the proposed architecture has an initial causal convolution instead of a dilated causal convolution, as shown in Figure 4.4. Unlike dilated causal convolution, a causal convolution with denser filters is more effective in extracting the local dependencies among adjacent time-steps. However, a stack of causal convolutions dramatically increases the model complexity. Therefore, we combine causal convolutions with dilated causal convolutions to achieve desirable prediction accuracy, while eliminating the complexity possibly caused by only utilizing causal convolutions in the architecture. As a result, the proposed model can effectively capture the temporal dependencies among adjacent and far-apart time-steps in the sequential QoE data, providing a better QoE prediction accuracy, especially in terms of RMSE.

Moreover, it can be seen from Tables 4.6 that the FLOPs of the proposed model are larger than those of LSTM-QoE. The reason is that the convolution layers require more operations

for performing convolution between a number of filters and the input time series. However, the proposed model runs faster than the baseline models, which indicates that the convolution operations are fully parallelized, leading to real-time QoE prediction advantages.

4.5.2 Effects of simplifying the residual block and using SeLU

To simplify the residual block, only one dilated causal convolution is adopted in the residual block instead of two as in the original TCN architecture (as illustrated in Figure 4.3 and Figure 4.5). The reason behind this is the fact that the sequential QoE data is much simpler than the preferred data of TCN [35] (i.e., speech signal data). Therefore, two dilated causal convolution layers can make the model easily suffers from overfitting and reduces the QoE prediction accuracy. Reducing the number of dilated causal convolutions in the residual block helps the proposed model to be easily trained and reduce overfitting. Furthermore, SeLU [43] activation function also enables the model to learn faster and converge better to the optimal values, subsequently improving the QoE prediction accuracy.

In terms of computational complexity, observing from Tables 4.6, it is obvious that these improvements in the residual block tremendously reduced the number of parameters compared to the one in the original TCN architecture TCN-QoE. Thereby, the CNN-QoE can produce smaller model size and FLOPs, faster training and inference times.

In summary, the improvements in the proposed architecture help provide a more stable, accurate and light-weight QoE prediction model.

4.6 Summary

In this chapter, the CNN-QoE model is presented for continuous QoE prediction. The proposed model introduces multiple improvements to the original TCN model to leverage its strengths and eliminate its drawbacks in the QoE prediction task for video streaming services. The comprehensive evaluation across different QoE databases demonstrates that CNN-QoE model produces superior performance in terms of QoE prediction accuracy and computational complexity. Accordingly, CNN-QoE provides a highly competitive prediction performance. These results validate the robustness of the model in real-time QoE prediction.

Chapter 5

Cumulative QoE Prediction Model

5.1 Introduction

Most of existing works only focused on modeling the overall or the instantaneous QoE, which have shown insufficient characteristics. The overall QoE [46–48], which demonstrates the final subjective judgment for a streaming session, can only be assessed when the viewer finishes watching. Therefore, the overall QoE cannot be applied for real-time QoE monitor and also does not give sufficient information about events occurring during the session. Although the instantaneous QoE [6, 16], on the other hand, can provide the instant perceived video quality at a certain moment, it only reflects locally the quality assessment within a specific time range, without considering the cumulative effects of prior events. Hence, it is highly sensitive to video impairments due to hysteresis effect [49, 32] and does not precisely express the user's perceived video quality. In contrast, modeling and predicting the user's cumulative perception to a streaming video content are able to provide lots of advantages for QoE monitor and control systems since it not only reflects the user's overall satisfaction but also reveals the impact of distorted events happening during the streaming session.

In addition, according to [12], QoE is defined as the results from the fulfillment of the user's expectation to the enjoyment of the application or service based on his or her *personality* and *current state*. Here, "personality" defines "the characteristics of a person that account for consistent patterns of feeling, thinking and behaving", whereas, "current state" stands for "situational or temporal changes in the feeling, thinking or behavior of a person". Therefore, human-related influence factors (e.g., perceptual factors, memory effect, user's interest in video content) play a crucial role in accurately modeling the user's QoE.

Some studies investigate and quantify the impact of perceptual factors [46–48, 6, 16]. However, the authors usually abandon the temporal dynamics and historical experience of the user's satisfaction, which are referred to as the memory effects [50]. Some other studies

attempt to clarify the role of primacy and recency effects [17, 51, 18, 34, 7, 20, 1], resulting in the high accurate QoE prediction. Typically, the primacy and recency effects [52] determine the memory influence of impairments occurring at the beginning and the end of streaming session [8], respectively. Besides, the effect of unpleasant events which take place in the middle of the session also leaves a considerable impact on the perceived video quality [8, 32]. Theoretically, such impacts can be represented by an exponential deterioration of memory retention in time (defined by *forgetting curve*) [53–55] for infrequent events or by *repetition* [32, 55] for the repeated impairments. However, the influence of forgetting behavior and repetition has not been carefully investigated in existing QoE models. Therefore, to fully express human memory effects on QoE assessment, in addition to the primacy and recency effects, the forgetting curve and repetition should be involved in the discussion.

Apart from that, the factors that relate to video content also have a noticeable effect on perceived QoE. Those factors might be type of video, video complexity [5], etc. Additionally, some studies (e.g., [56–58]) have found that the user's interest in video content possibly generates the bias in his/her QoE evaluation. More concretely, the user tends to provide higher QoE scores for more attractive video contents. Such a behavior is influenced by the so-called degree of interest (DoI) which clarifies the interestingness of different video content, or the ability of the video content to attract the user and keep the user's interest [59]. However, existing studies often neglect this factor due to the fact that these numerical values might vary upon different users based on their personal interests.

For those reasons, in this chapter, we present a cumulative QoE model that extremely well quantifies multiple effects of human-related factors, that is to say, perceptual influence factors, memory effect and degree of interest (DoI). In order to assess the accuracy in predicting cumulative QoE, the cumulative QoE model is evaluated over LFOVIA database [34] and through the subjective evaluation. Evaluation results demonstrate that the cumulative QoE at different moments within a streaming session is precisely predicted by the model. It shows the potential of cumulative QoE that can be utilized as a better alternative than either overall QoE or instantaneous QoE in QoE monitoring and management.

The rest of this chapter is organized as follows. Section 5.2 investigates the influence of human-related factors and presents the cumulative QoE prediction model. Section 5.3 evaluates the performance of the proposal and discusses the advantages and disadvantages. Section 5.4 summarizes this chapter.

5.2 Cumulative QoE Model

In this section, we first investigate the impact of those factors on human perception in QoE evaluation and then formulate the proposed cumulative QoE model.

5.2.1 Perceptual factors

In video QoE assessment, perceptual factors [23] including video quality, rebuffering frequency, rebuffering duration are directly perceived by the user. Sub-section 3.3.1 has introduced four input features (i.e. STSQ, PI, NR, and TR) that are related to perceptual factors. These input features are fed into an LSTM-QoE model [1] to predict the instantaneous QoE as follows [1]:

$$q(t) = LSTM^0(\boldsymbol{x}(t), \boldsymbol{c}(t-1)) \tag{5.1}$$

where, $q(t)$ represents the predicted instantaneous QoE at the time instant t, $\boldsymbol{x}(t)$ is the input features, $\boldsymbol{c}(t)$ is the memory cells which encode the knowledge of the inputs that have been observed up to the time t. $LSTM$ provides two functionalities: $LSTM^0$ for output QoE prediction and $LSTM^c$ for memory cells update which is given by [1]:

$$\boldsymbol{c}(t) = LSTM^c(\boldsymbol{c}(0:t-1), q(0:t-1)), \forall t \geq 1 \tag{5.2}$$

where, $\boldsymbol{c}(0:t-1)$ and $q(0:t-1)$ respectively refer to the past memory cells and the past predicted QoE.

The architecture of LSTM-QoE model are illustrated in Figure 5.1.

5.2.2 Memory Effects

Memory effects refer to the influence of historical/past experiences on the perceived video quality. Primacy and recency are two common effects which were investigated in numerous studies [1, 7, 18]. In addition to these factors, the effect of forgetting curve characteristic and repetition are also considered in our proposed model. The next parts of this subsection will discuss the role and mathematical function of these factors. Based on that, a memory weight is proposed for the cumulative QoE model.

5.2.2.1 Primacy Effect

The primacy effect [60, 52] describes the human behavior to recall (bitrate or rebuffering) initial events occurred at the beginning of the streaming session when providing the overall

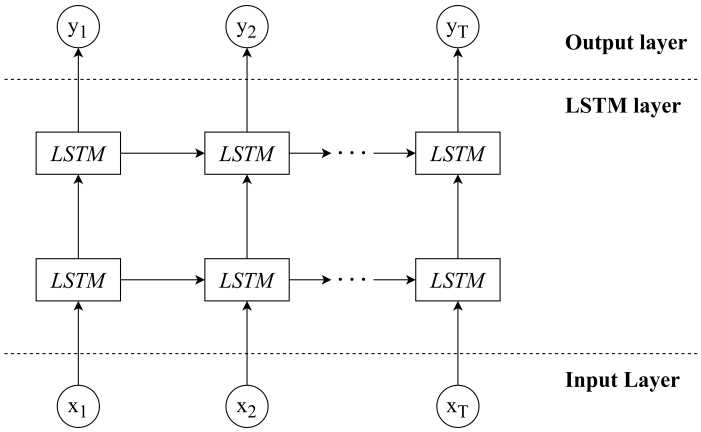


FIGURE 5.1 LSTM network [1] for the user's instantaneous QoE prediction. The network is composed of two LSTM (Long Short-term Memory) layers. The inputs to the layers are four features including STSQ, PI, NR, and TR. The outputs combine the LSTM layers' hidden states, representing the predicted instantaneous QoE values.

evaluation [61]. In fact, the primacy effect always exponentially decreases by time [60]. Therefore, its characteristics can be expressed by an exponential curve as follows:

$$f_P(t) = exp(-\alpha_P * t), \quad 0 \leq t \leq L \tag{5.3}$$

where α_P determines the *intensity* of primacy effect (how fast the primacy effect diminishes over time) and t denotes a time instant within a session of L seconds.

5.2.2.2 Recency Effect

The recency effect [60, 52] refers to the ability of the human memory to recall the most recent events [61], hence, the evaluated QoE heavily depends on the recent experiences. The recency effect also can be described by an exponential curve represented by the following equation:

$$f_R(t) = exp(-\alpha_R * (L-t)), \quad 0 \leq t \leq L \tag{5.4}$$

where α_R determines the *intensity* of recency effect.

The primacy effect and the recency effect can be combined as the U-shaped form [52], quantifying the influenced weight of the events occurring from the beginning to the end of a video session. As shown in Figure 5.2, it can be observed that both Eq. 5.3 and Eq. 5.4 reflect the primacy and recency effect extremely well.

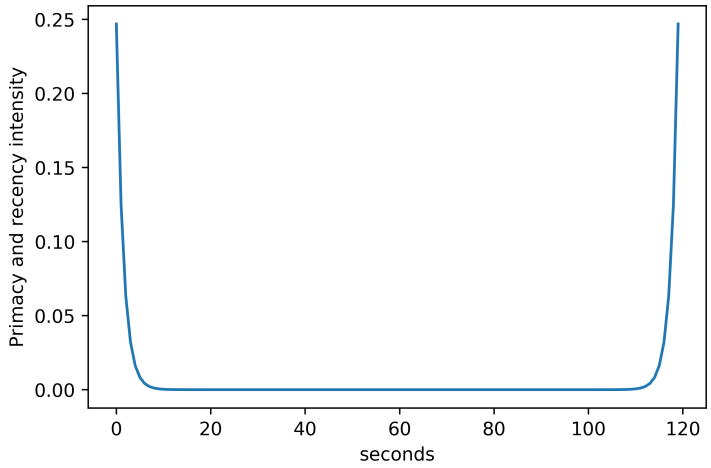
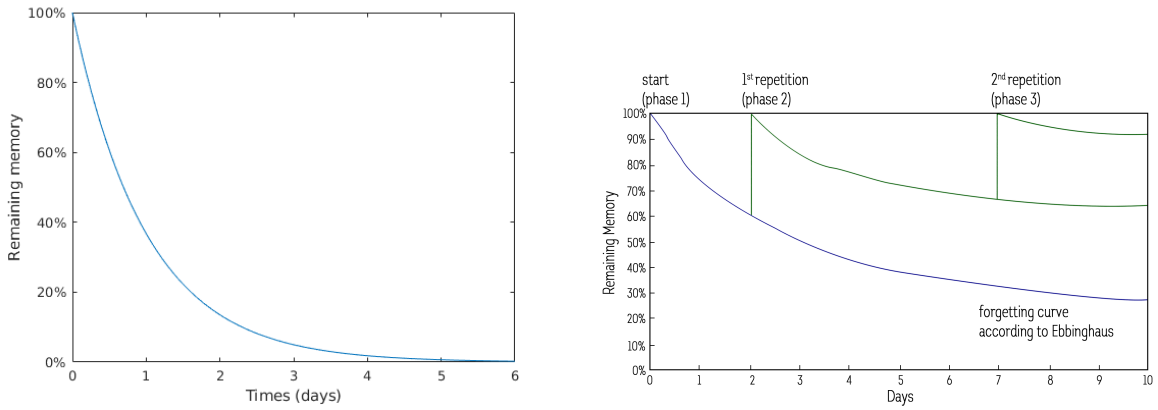


FIGURE 5.2 A typical U-shaped curve combined primacy and recency effects.



(a) An example of forgetting curve

(b) Forgetting curve and repetition

FIGURE 5.3 Examples of forgetting curve and repetition.

5.2.2.3 Forgetting Curve and Repetition

Due to the significant impact of the negative experience caused by distorted events, the primacy and recency effect can be neglected under repeated bitrate switches or rebuffering [5]. In such situations, forgetting behavior and repetition should be taken into account. The forgetting behavior, in other words, forgetting curve characteristic [55] is a natural process, describing the exponential loss of memory over time. As shown in Figure 5.3a, when information is learned, its memory retention declines at an exponential rate. Accordingly, any occurred events can be exponentially forgotten by time if there is no attempt to retain it. The level of remaining memory about such events at a specific time point depends on:

- The strength of memory (memory intensity): The durability that memory traces in the brain. The more annoyance the event is, the stronger the user memorizes it and the longer it lasts.
- The time has elapsed since the occurrences of events: As shown in Figure 5.3a, the user will forget an average of 60% of what they experience within the first period of time [54, 55].
- Repetition: The more frequently an event occurs, the more likely it sticks to the user memory (shown in Figure 5.3b)

In a typical streaming session, an interruption (bitrate switching or rebuffering) can happen regularly. When an event, especially rebuffering repeatedly occurs, the strength of memory of those events will trendily increases [55], negatively influencing the perceived video quality. Consequently, as the number of negative events increases, QoE will recover at a slower pace after the occurrence of each event. Such memory characteristics can be formulated as the following equation [62]:

$$f_{RP}(t) = exp(-\frac{\alpha_{RP}}{NR(t)} * TR(t)), \quad 0 \leq t \leq L \tag{5.5}$$

where $NR(t)$ is the number of rebuffering events occurring until the time t, $TR(t)$ is the time elapsed since the last video impairment, and α_{RP} is the intensity of memory related to a rebuffering event. The ratio $\frac{\alpha_{RP}}{NR(t)}$ determines the retention of the user's memory after the $NR(t)$-th rebuffering. Accordingly, the lower $\frac{\alpha_{RP}}{NR(t)}$ is, the higher retention rate, making f_{RP} declines at a lower rate.

5.2.2.4 Proposed Memory Weight

As discussed in the previous sub-subsections, the effects of primacy, recency, forgetting behavior and repetition are significantly crucial for the evaluation of the cumulative QoE. Therefore, in the proposed cumulative QoE model, we introduce a novel *memory weight* incorporating the effects of those factors to accurately assess the cumulative human perception during a streaming session. The proposed memory weight is represented by Eq. 5.6. An example of time-varying memory weight is illustrated in Figure 5.4. In fact, Eq. 5.6 is a linear combination effect of the above-mentioned memory factors obtained from Eq. 5.3, Eq. 5.4 and Eq. 5.5.

$$w_t = \beta_1 f_P(t) + \beta_2 f_R(t) + \beta_3 f_{RP}(t) \tag{5.6}$$

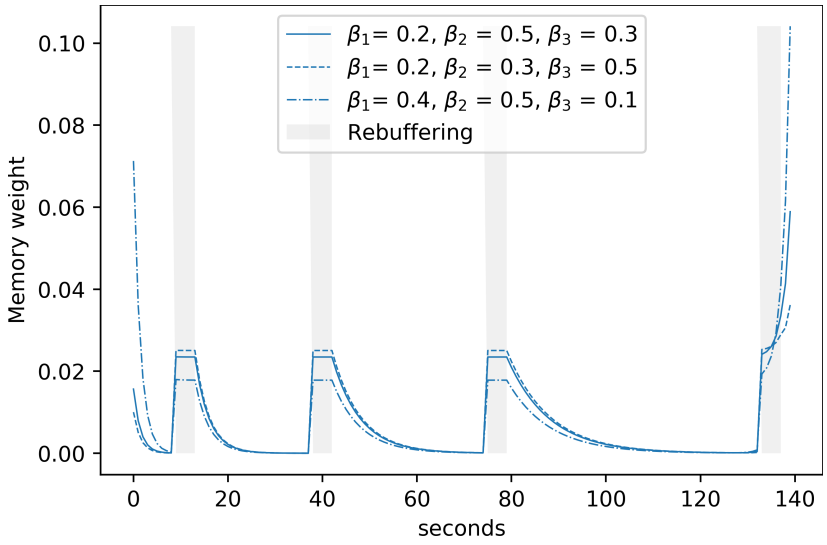


FIGURE 5.4 An example of the memory weight in a session under different values of parameters β_1, β_2, and β_3

where $\beta_1, \beta_2, \beta_3$ respectively determine the contribution of primacy effect, recency effect and repetition to the memory weight.

Figure. 5.4 shows that when an rebuffering event occurs near the end of the session, the recency effect has a stronger effect on human perception. Therefore, in this period of times, the end user's QoE will drops dramatically. In addition, the forgetting rate of a specific interruption is also smaller than those of previous ones, determining the characteristics of forgetting behavior and repetition. Therefore, the proposed memory weight potentially reflects the intensity of human memory over time during a streaming session.

5.2.3 Degree-of-Interest

For modeling QoE, there have been numerous studies that take into account video content-related factors (e.g., type of video, the complexity of video, etc.). However, most of them neglected the user's interest, in other words, DoI. In fact, influenced by video content and viewer preferences, the user possibly has different DoI on different videos or different parts of a video. Intuitively, the user seems to provide higher QoE scores for the video with interesting content and vice versa. Typically, Degree-of-Interest (DoI) [63] is defined as the interestingness of the video content, or the ability of the video content to attract the user and keep the user's interest [59].

To make this clear, we investigate the correlation between DoI and the overall QoE by conducting a subjective test. In this test, 18 undistorted videos from the LFOVIA Database

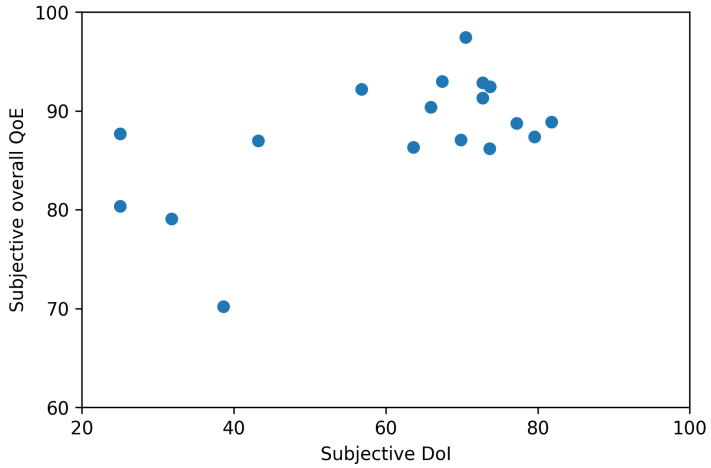


FIGURE 5.5 Scatter plot between the mean of subjective DoI scores and the subjective overall QoE obtained in the database.

[34] were utilized. The video content varied upon nature, wildlife, outdoor, marine, sports, animation, and gaming [34] among every video, each of which the duration is 120 seconds. This guaranteed that the subjects would retain their interests as they watched. The referenced videos were randomly divided into 6 collections and encoded using FFmpeg [64] under the default settings with the resolution of 1920 x 1080 and were displayed on a 15-inch monitor with a resolution of 1920 x 1080 and a black background. The Absolute Category Rating (ACR) [65] method was used and there were 60 subjects agreed to participate in this experiment. Each video was assessed by at least 10 subjects. At the end of each video, the subject was asked to give an overall score representing his/her interest in the entire video content, ranging from 1 (worst or not at all interested) to 5 (best or extremely interested), following the general principle of the ITU-T recommendation P.913 [65]. A 3-minute break was provided to each subject between each video to minimize the effects of viewer fatigue. The average of subjects' scores or Mean Opinion Score (MOS) for each video was utilized as the DoI of video. These values were then linearly scaled up to the range of 0 to 100 and compared with the corresponding overall QoE in the LFOVIA Database.

Figure 5.5 illustrates the obtained correlation between DoI and the overall QoE, which achieved the Pearson Correlation Coefficient (PCC) of **0.601**. The correlation was modest. We speculate this as the small number of subjects participating in the experiment. Yet, it is shown that the DoI has an influence on the final decisions of the users when they provide the overall QoE. In the future, a larger number of subjects will be considered for further investigation. Based on the conclusion of this experiment, we introduce DoI as one of the potential influence factors in the proposed cumulative QoE model.

5.2.4 Cumulative QoE model

Through the investigation of the above human-related influence factors, the proposed cumulative QoE model is generally presented in Eq. 5.7. In this model, to quantify how each of the user's past experiences influences the cumulative perception, the instantaneous QoE needs to be weighted by the memory effect from the beginning of playback to the investigated time point t within a streaming session. According to our proposed model, the procedure of estimating cumulative QoE is described as follows: Firstly, the instantaneous QoE is predicted by LSTM-QoE model [1], and stored into vector $Q_t = (q_0, q_1, \ldots, q_t)$. Secondly, the memory weight is calculated by the Eq. 5.6 to form vector $W_t = (w_0, w_1, \ldots, w_t)$.

$$CQ_t = \lambda_1 \left(Q_t \times W_t^T \right) + \lambda_2 DoI \tag{5.7}$$

where λ_1, λ_2 are correlation coefficients which respectively determine the contribution of the user's past experience and user's interest in video content to the predicted cumulative QoE CQ_t at time instant t.

5.3 Performance Evaluation and Discussion

In this section, we start with the explanation of the proposed model's establishment where the necessary parameters including $\{\alpha_P, \alpha_R, \alpha_{RP}\}$, $\{\beta_1, \beta_2, \beta_3\}$ and $\{\lambda_1, \lambda_2\}$ are numerically determined. Afterward, we briefly evaluate and discuss the prediction performance of our model. The evaluation is two-fold. First, the prediction performance of the proposed model is quantitatively and qualitatively assessed on test videos in a specific database [34]. Second, a subjective test is conducted in order to evaluate how well the predicted cumulative QoE correlates with subjective cumulative evaluation at different moments of a streaming session. Finally, the complexity of the proposed model is also analyzed for real-time cumulative QoE prediction.

5.3.1 Model Establishment

The parameters of the proposed model was computed according to a four-step procedure as follows:

1) A specific publicly available database was employed for establishing and evaluating the proposed model.

2) An LSTM-QoE model [1] was trained to predict the instantaneous QoE values.

3) The memory effects' parameters $\{\alpha_P, \alpha_R, \alpha_{RP}\}$ were computed to form the memory weight vector.

4) The coefficients of memory weight $\{\beta_1, \beta_2, \beta_3\}$ in Eq. 5.6 and the parameters of the proposed model $\{\lambda_1, \lambda_2\}$ in Eq. 5.7 were determined through the predicted instantaneous QoE values and the subjective DoI collected from the experiment in subsection 3.3.

The details of each step are described in the next sub-subsections.

5.3.1.1 Database description

Our model was established and evaluated based on a set of 36 distorted videos in LFOVIA Video QoE Database [34]. These videos have different playout patterns distorted by bitrate switching and rebuffering events. In this database, the overall QoE and the time-varying instantaneous QoE scores for those videos were obtained are in the range [0, 100], with score 0 being the worst and 100 being the best. The set of distorted videos was divided into training and testing sets with a training:testing ratio of 80:20. Accordingly, there were 28 videos in the training set and 8 videos in the testing set. The training and testing set were respectively used to obtain the model parameters described in sub-subsection 5.3.1.3 and evaluate the prediction performance of the model presented in subsection 5.3.2.

5.3.1.2 Instantaneous QoE Prediction by LSTM-QoE

The instantaneous QoE values were estimated by the LSTM-QoE model [1]. The model was trained on the training set with 28 distorted videos driven by 4 features $STSQ$, PI, NR and TR. The performance of this model was then quantified on the 8 test videos using the Pearson Correlation Coefficient (PCC) and Spearman Rank Order Correlation Coefficient (SROCC). Consequently, the model achieved high accuracy with PCC of **0.9946** and SROCC of **0.8870**. The performance of the trained model is illustrated in Figure 5.6, demonstrating high accurate prediction.

5.3.1.3 Parameters Selection

As discussed in subsection 5.2.2, the parameters $\{\alpha_P, \alpha_R, \alpha_{RP}\}$ indicate how memory factors impact the perceived video quality over time. The larger $\{\alpha_P, \alpha_R, \alpha_{RP}\}$ are, the easier it is for the user to forget. According to [52, 8], the effects of primacy and recency gradually decrease within 15 to 20 seconds. Therefore, the deteriorating time was set to 15 seconds. Since the user usually recalls unpleasant events when providing a QoE score, the effect of

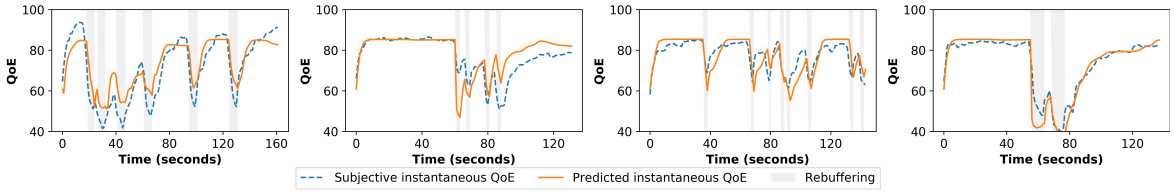


FIGURE 5.6 Some examples of instantaneous QoE prediction performance obtained from the LSTM-QoE model on different test videos of the database.

TABLE 5.1 Parameters of the primacy and recency effect, forgetting curve and repetition

α_P	α_R	α_{RP}
0.6807	0.6807	0.3404

repetition is larger and remain longer than primacy and recency. As a result, the values of α_{RP} must be smaller than α_P and α_R. The effect of repetition will remain within 30 seconds. The function *solve* in MATLAB [66] was employed to compute the parameters α_P, α_R, and α_{RP} according to Eq. 5.3, 5.4, and 5.5, respectively. Consequently, the obtained values of parameters $\{\alpha_P, \alpha_R, \alpha_{RP}\}$ are shown in Table 5.1.

Thereby, the parameters $\{\beta_1, \beta_2, \beta_3\}$ and $\{\lambda_1, \lambda_2\}$ of weight memory and the proposed cumulative QoE model are now can be estimated. Considering a streaming session with a video in the training set of L seconds, the cumulative QoE from the beginning to the end of the streaming session was calculated as follows:

$$\begin{aligned} CQ_L &= \lambda_1 \left(Q_L \times W_L^T\right) + \lambda_2 DoI \\ &= \lambda_1 \sum_{i=0}^{L} w_i q_i + \lambda_2 DoI \\ &= \lambda_1 \sum_{i=0}^{L} \left(\beta_1 f_P(i) + \beta_2 f_R(i) + \beta_3 f_{RP}(i)\right) q_i + \lambda_2 DoI \end{aligned} \tag{5.8}$$

where, Q_L is the vector of instantaneous QoE $(q_0, q_1, \ldots, q_L)$, W_L is the memory weight vector $(w_0, w_1, \ldots, w_L)$.

TABLE 5.2 Parameters of memory weight and the cumulative QoE model

β_1	β_2	β_3	λ_1	λ_2
0.0284	0.8492	0.1177	0.9809	0.0800

As mentioned in subsection 5.2.4, the cumulative QoE at the end of the session CQ_L is also considered as the overall QoE. Therefore, we first need to minimize the least square error:

$$\boldsymbol{J} = \|CQ_L - Q_{overall}\|^2 \tag{5.9}$$

where $Q_{overall}$ is the subjective overall user's QoE obtained from the database. A curve fitting is performed using *lsqcurvefit* in MATLAB [66] with 28 training videos to obtain the memory weight parameters $\{\beta_1, \beta_2, \beta_3\}$ and the cumulative QoE parameters $\{\lambda_1, \lambda_2\}$. The numerical values of those parameters are shown in Table 5.2.

5.3.2 Performance Evaluation on Testing Videos

After obtaining the necessary parameters for the proposed model, we quantitatively and qualitatively evaluate its prediction performance on 8 distorted videos in the testing set. Alternatively, the discussion on the results is also performed.

To quantitatively assess the prediction performance, the correlation between the subjective overall QoE obtained in the LFOVIA Video QoE Database [34] and our predicted cumulative QoE at the end of each video was computed. It is crucial to note that the subjective overall QoE can be considered as the cumulative perception of the user at the end of streaming session. There were three evaluation metrics utilized for evaluation: 1) Pearson Correlation Coefficient (PCC), 2) Spearman Rank Order Correlation Coefficient (SROCC), and 3) Root Mean Square Error (RMSE). Typically, PCC and SROCC quantify how well the predicted QoE tracks the actual QoE scores in the database, whereas, RMSE indicates the closeness between them. We also compared our proposed model with a reference method of [67], using the same training set and testing set. The cumulative QoE model in [67] is characterized by the following equation:

$$Q_t = \gamma Q_{t-1} + (1-\gamma) q_t \tag{5.10}$$

where q_t is the instantaneous user experience at moment t, Q_{t-1} is the cumulative QoE at the previous moment $t-1$, and γ is the memory strength parameter. The correlation between the predicted cumulative QoE obtained from this model and the subjective overall QoE in LFOVIA database was also investigated through PCC, SROCC and RMSE metrics. We reported the performance of our model and the reference method in Table 5.3. This result shows a superior prediction performance of our model. Figure 5.7 additionally emphasizes the competitive performance of our model. Thereby, the proposed model has effectively assessed cumulative perception over multiple scenarios in testing videos.

TABLE 5.3 Prediction performance of the reference model and our proposed model over training and testing set

		PCC	SROCC	RMSE
Training	[67]	0.7413	0.6420	10.6187
	Proposed model	**0.9441**	**0.8604**	**4.1525**
Testing	[67]	0.2777	0.2381	7.5135
	Proposed model	**0.7664**	**0.7857**	**4.6538**

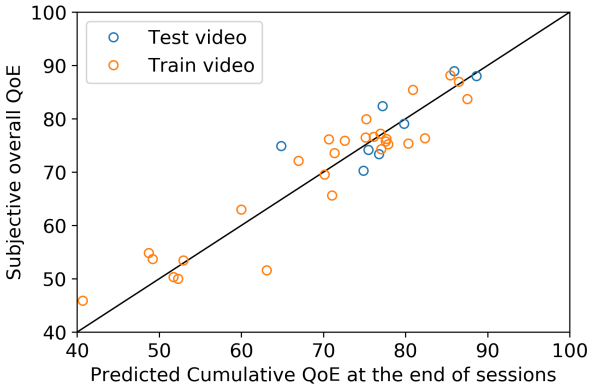


FIGURE 5.7 Correlation between subjective overall QoE and predicted cumulative QoE at the end of streaming session

In qualitative evaluation, our purpose is to validate the impact of memory effects and DoI on the cumulative QoE prediction over multiple scenarios on testing videos. Thereby, the prediction performance of the proposed model in a short period and longer period can be assessed. For that reason, in Figure 5.8, we plot the predicted cumulative QoE in comparison with both subjective instantaneous QoE and subjective overall QoE which are obtained from the database. From now, the terms of subjective instantaneous QoE and subjective overall QoE will be referred to as instantaneous QoE and overall QoE for short.

In general, the predicted cumulative QoE precisely reacts to any interruption at any moment while being close to the overall QoE at the end of the streaming session. For the initial interruption, it is always witnessed a significant deterioration in predicted cumulative QoE. Nevertheless, when such unpleasant events continuously occur, the predicted cumulative QoE tends to decrease at a lower rate. Additionally, a lower recovering rate is subsequently introduced after each event.

5.3.2.1 Impacts of memory effects

In pattern #2, #5, #7, there is only one interruption with short duration occurring near the beginning of streaming sessions. As a result, the predicted cumulative QoE introduces a

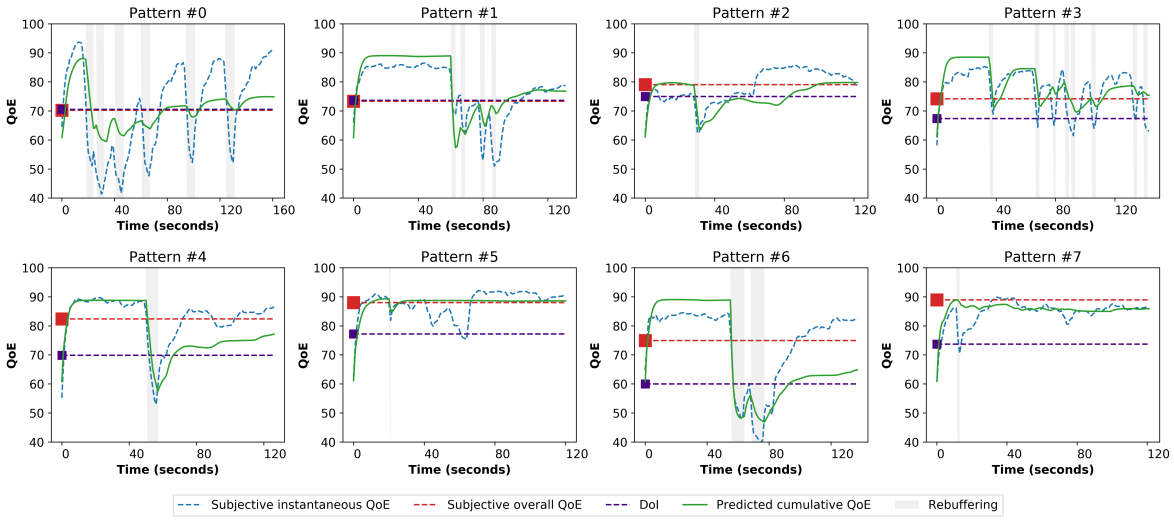


FIGURE 5.8 Predicted cumulative QoE in comparison with the subjective overall and instantaneous QoE over eight different playout patterns

slight decrease, following by a gradual recovery and convergences at the values as close to the overall QoE. These trends are consistent with those of instantaneous QoE. It means that the prediction accurately demonstrates the role of forgetting curve characteristic as well as the recency effect. More concretely, after the finishing of the interruption, the memory intensity about such event starts to exponentially decay, leading to the recovery in perceived video quality. At the end of streaming sessions, there is a possibility that the decay has completely finished, in other words, the memory of distorted events is vanished. Therefore, the recency effect becomes dominant, leading to the consistency among the predicted cumulative QoE, instantaneous QoE, and overall QoE.

In pattern #0 and #1, rebuffering event repeatedly occurs in the middle of streaming sessions. While the predicted cumulative QoE is consistent with the overall QoE at the end of sessions, the instantaneous QoE tends to continuously increase, creating a big gap to the overall QoE. At first sight, one might think that the overall QoE must be as high as the instantaneous QoE at the end of streaming sessions due to the recency effect. This inference is understandable because the moment at which the last interruption occurs is quite far from the end of sessions, thus, the recency effect would have become dominant, resulting in the consistency among those QoE evaluations. However, when the interruption repeats many times, the impact of repetition characteristic become significantly obvious. Consequently, the user tends to provide an overall evaluation whose value is lower than the instantaneous QoE. On the other hand, by considering the recency effect and repetition characteristic, our proposed model can effectively provide the prediction consistency with the overall QoE.

According to the hysteresis effect [49], the user is highly sensitive to a single unpleasant event and provides poor QoE scores immediately. However, when the interruption occurs many times as in pattern #0, #1 and #3, the impact of the hysteresis effect will be shared with the repetition characteristic. This makes the user behaves in the consideration of past annoying events to avoid the aggressive reaction. In addition, under the impact of repetition characteristic, such the events are stuck in the user's memory and are recalled when the user provides the overall assessment. However, the instantaneous QoE always aggressively reacts to the distorted events, by dramatically decreasing and quickly recovering during a short period. This is because the instantaneous QoE is estimated locally without considering the global views of the streaming session. Oppositely, by weighting the instantaneous QoE by the memory effects (especially repetition characteristic), the predicted cumulative QoE can react calmly, and, eventually, correlates perfectly with the overall QoE.

Interestingly, the predicted cumulative QoE also indicates a special behavior in human perception which cannot be found in the instantaneous QoE and overall QoE. We call such the behavior as the *persistent evaluation* where the user seems to familiar with the distorted event and to accept it. The user does not even want to deteriorate their evaluation score or to quit from the streaming session. For instance, pattern #0 and #3 visualizes that the cumulative QoE dramatically falls after the occurrence of the first rebuffering event. However, it decreases with a significantly lower amplitude on the ones happening subsequently.

5.3.2.2 Impacts of DoI

As we mentioned in subsection 5.2.3, the correlation between DoI and subjective overall QoE is modest. However, the contribution of DoI on prediction performance is well recognized in some cases as shown in pattern #2, #3, #5 and #7 which share a common characteristic where the predicted cumulative QoE correctly meets the overall QoE. To be honest, without DoI ($\lambda_2 = 0$), the predicted cumulative QoE would have been much lower than the overall QoE. Especially, in each pattern #5 and #7 where contains only one super-short rebuffering event near the beginning of streaming sessions, the memory intensity about this event must be completely vanished, followed by the dominance of the recency effect, resulting in very high overall QoE. However, the contents of these two videos might not sufficiently interesting to the users, leading to the deterioration in their evaluation. Therefore, when the contribution of DoI is precisely recognized, our proposed model provides an extremely high accurate prediction. However, in pattern #4 and #6, there exists long duration interruptions in the middle of streaming sessions, creating significantly high intensity memory about those events. As a result, the predicted cumulative QoE dramatically decreases and slowly recovers. However, the insufficiently accurate contribution of DoI has curbed the recovering rate.

TABLE 5.4 Prediction performance of reference model and the proposed model over subjective experiment

	PCC	SROCC	RMSE	OR (%)
[67]	**0.5418**	0.3917	9.1318	33.3
Proposed model	0.5405	**0.5146**	**9.0922**	**25.0**

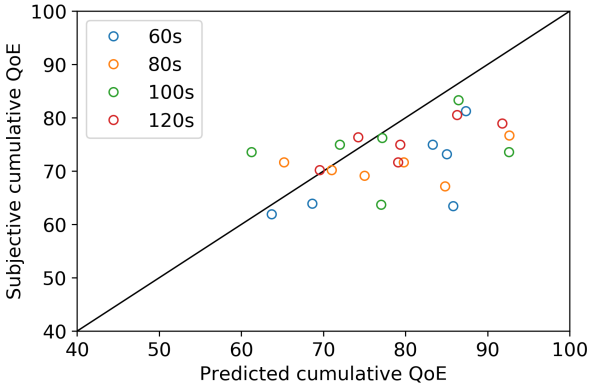


FIGURE 5.9 Scatter plot of predicted cumulative QoE and subjective cumulative QoE.

Consequently, the predicted cumulative QoE cannot catch up with the overall QoE at the end of streaming sessions. This emphasizes the lack of generalization in DoI coefficient λ_2. We believe that the original reason is the insufficient number of participated subjects in the subjective evaluation in subsection 5.2.3 where each video was watched and evaluated by only 10 subjects. In the future, a larger number of participants must be involved in this experiment.

5.3.3 Subjective Evaluation

In this subsection, a subjective evaluation is conducted to assess the accuracy of the proposed model aligning with ground truth QoE scores provided by a number of subjects. The performance of QoE prediction using the proposed model is evaluated by relying on the following four measures: 1) PCC, 2) SROCC, 3) RMSE and 4) Outage Rate (OR) [6]. While PCC and SROCC quantify the correlation between predicted cumulative QoE and the subjective cumulative QoE, the closeness between predicted scores and the ground truth scores is numerically obtained by using RMSE and OR. In particular, OR measures the frequency of times when the prediction p_i falls outside twice the confidence interval of subjective scores s_i, which is defined as the following equation:

$$OR = \frac{1}{N}\sum_{i}^{N} \mathbb{1}(|p_i - s_i| > 2CI_{s_i}) \tag{5.11}$$

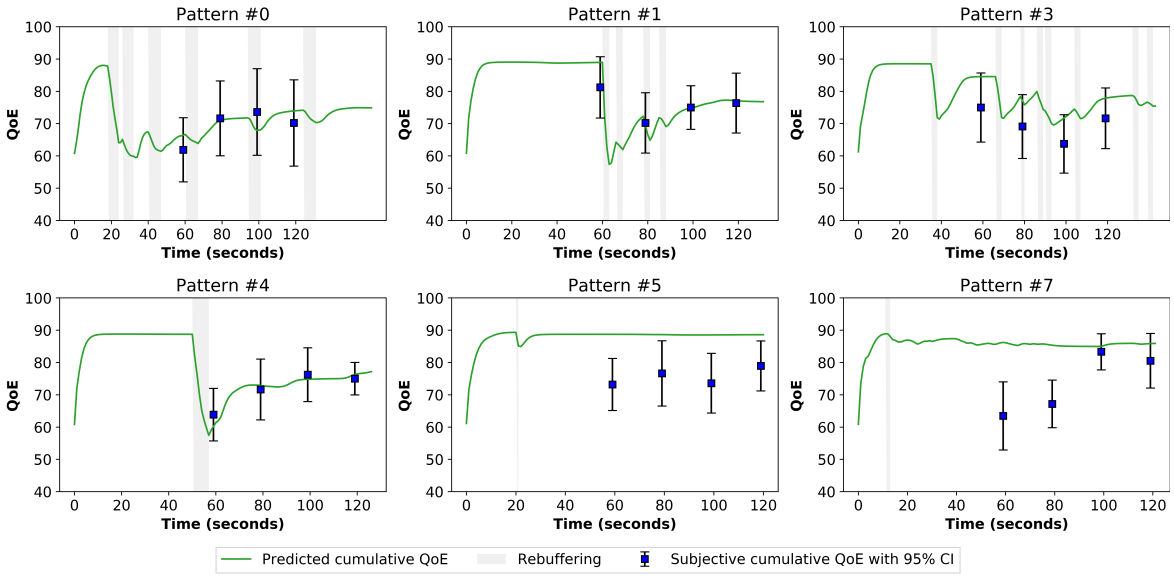


FIGURE 5.10 Performance of our predicted cumulative QoE in comparison with the subjective cumulative QoE.

where $\mathbb{1}(\cdot)$ is the indicator function

To conduct the subjective test, 6 distorted videos from the testing set of LFOVIA database (pattern #0, #1, #3, #4, #5, #7) were selected. We selected those videos because they have different contents, thus, the role of DoI in our model can be potentially assessed. Each distorted video was cropped into 4 small videos with starting timestamps of 00:00:00 and different length (60, 80, 100, and 120 seconds) using FFmpeg [64]. The purpose is to ask the subjects to provide subjective cumulative evaluations at the time points of 60, 80, 100, and 120 seconds of each distorted video. The correlations between subjective cumulative QoE and predicted cumulative QoE obtained from our model and reference model were assessed. The cropped videos were divided into 6 collections with different video content and displayed on a 15-inch screen with a resolution of 1920x1080 and a black background. Each video was rated by at least 18 subjects and there were totally 120 participants. Note that these subjects are different from those in "DoI" experiment in subsection 3.3. The Absolute Category Rating method was used in our experiment [65]. The subjects give a rating score at the end of each cropped video with the score ranging from 1 (worst) to 5 (best) based on the perceived quality and video content, following the general principle of the ITU-T recommendation P.913 [65]. The average of subjects scores, associated with 95% confidence interval, for each cropped video, was utilized as the subjective cumulative QoE. These values were linearly rescaled so that the scores lay in the range [0, 100] and then compared with the predicted cumulative QoE.

Figure 5.9 illustrates the obtained correlation between the predicted cumulative QoE and subjective cumulative QoE. The comparison in QoE prediction performance between our model and reference model is tabulated in Table 5.4. Accordingly, we observe that the proposed model provides a competitive performance in terms of SROCC, RMSE and OR against the reference model. On the other hand, Fig. 5.10 shows a reasonable prediction performance of our model in comparison with subjective cumulative QoE at four discrete moments (at time points of 60, 80, 100, and 120 seconds) within a streaming session. In general, the proposed model performs extremely well when the high frequent and long duration rebuffering occur. It means that our model is capable of cumulatively capturing the effects of all the occurred unpleasant events on human perception. However, the model performance in pattern #5 and #7 are poorer, as compared to other patterns (#0, #1, #4) even though they have only one short rebuffering event. This can be explained that in pattern #5 and #7, the users' perception seems to be significantly affected by the video content. In other words, the effect of DoI become dominant in their evaluation, which is not precisely captured by our model.

5.3.4 Computational Complexity

The computational complexity of the proposed model is determined by the computational complexity of forming the instantaneous QoE vector $Q_t = (q_0, q_1, ..., q_t)$ predicted by the LSTM-QoE model. It is important to note that at the time instant t, the previous instantaneous QoE values $\{q_0, q_1, ..., q_{t-1}\}$ have already been predicted and cached in the memory. Since the LSTM-QoE model takes up only a very small computational overhead to predict q_t in order to form the vector Q_t, the cumulative QoE of each second CQ_t can be predicted in real-time. To demonstrate this, we calculated the required computing time for training LSTM-QoE model and predicting the instantaneous QoE at the end of a session q_L. All the timing experiments were carried out on a 18.04 Ubuntu LTS Intel i7-8750H @ 2.20GHz and 16GB RAM system. The LSTM-QoE model took 620.740 seconds to train and 0.4917 milliseconds to predict q_L. Furthermore, the cumulative QoE CQ_L prediction took 0.5103 milliseconds. Thus the proposed model is suitable for real-time cumulative QoE prediction.

5.3.5 Overall Evaluation

In evaluation section, we assessed the performance of the model on a publicly available database and the subjective test. By doing this way, we can validate the predicted cumulative QoE in both quantitative and qualitative manners. Typically, the model can precisely provide cumulative QoE prediction in different scenarios. Therefore, the model promisingly provides

an alternative and reliable approach in modeling QoE towards QoE based control and management.

5.4 Summary

In this chapter, the cumulative QoE prediction model was presented. The model successfully and effectively incorporated the impacts of human-related influence factors to predict the cumulative perceived video quality. In different scenarios, the proposed model achieved impressive performance, outperforming the reference model. Additionally, it was shown that the introduced memory weight accurately mimicked human memory during a streaming session, especially when unpleasant events repeatedly occurred. Although the correlation between DoI and subjective overall QoE was not so high due to the small number of subjects involved in the experiment, the user's interest in video content can be considered as a potential influence factor in predicting QoE.

Chapter 6

Discussion

This chapter discusses the QoE prediction performance of the three QoE models proposed in this thesis. Section 6.1 discusses the performance of two instantaneous QoE prediction models which are BiLSTM-QoE and CNN-QoE. Section 6.2 summarized the advantages and remaining issues of the cumulative QoE prediction model.

6.1 QoE prediction performance of BiLSTM-QoE and CNN-QoE models

In term of accuracy, the BiLSTM-QoE and CNN-QoE models are both outperforms the existing studies, as shown in Section 3.3 and 4.4. Table 6.1 tabulated the comparison in QoE prediction accuracy between our models and reference models over the LIVE Netflix Video QoE Database. Accordingly, the CNN-QoE model provides a competitive performance in terms of PCC and SROCC against the BiLSTM-QoE model. It should be noted that the BiLSTM-QoE model was evaluated on only one database. In contrast, three different

TABLE 6.1 QoE prediction accuracy of the BiLSTM-QoE and CNN-QoE over the LIVE Netflix Video QoE Database.

	PCC	SROCC	RMSE
BiLSTM-QoE	**0.894**	**0.830**	7.43
CNN-QoE	0.848	0.733	**6.97**
LSTM-QoE [1]	0.802	0.714	7.78
NLSS-QoE [20]	0.655	0.483	16.09
NARX [18]	0.621	0.557	8.52

databases were used to assess the performance of the CNN-QoE model. Section 4.4 showed that the CNN-QoE model can perform consistently well across the QoE databases.

Moreover, we also introduce several improvements to the CNN-QoE architecture to overcome the computational complexity drawbacks of LSTM-based QoE models. These improvements helped the model run faster than the reference models, leading to real-time QoE prediction advantages. Therefore, the CNN-QoE model can be an excellent choice for predicting the instantaneous QoE.

6.2 Cumulative QoE prediction model

The results of the cumulative QoE prediction model validated the impact of memory effects and DoI on the user's QoE. Moreover, the model can quickly and precisely estimate the cumulative QoE in the experiment. However, it still has some limitations. First, the model relies on an instantaneous QoE prediction model to predict the user's cumulative QoE due to the lack of data on subjective QoE evaluations. Second, the correlation between DoI and the user's QoE was not so high, hence, the prediction accuracy of the model is perhaps not sufficient. Finally, the model should be evaluated in multiple databases to understand how well the model will perform across diverse scenarios of video streaming.

Chapter 7

Conclusion and Future Work

7.1 Summary

In this thesis, three QoE prediction models for video streaming was presented in order to improve the QoE prediction performance in terms of both prediction accuracy and computational complexity. The BiLSTM-QoE model was first introduced to enhance QoE prediction accuracy by utilizing the advantages of BiLSTM networks. However, BiLSTM increases the model complexity which makes it not suitable for real-time QoE monitoring. Thus, the CNN-QoE model was then proposed to leverage the parallel processing in CNN architecture. The model achieved not only the state-of-the-art QoE prediction accuracy but also the high reduction in computational complexity. Since human-related influence factors play an important role in QoE modeling, we introduced the cumulative QoE prediction model that predicts the user's cumulative perception which takes into account the impact of past events during a streaming session. The cumulative QoE prediction model provides a promisingly alternative and reliable approach in modeling QoE towards QoE-based control and management.

7.2 Future Work

In the future, we plan to extend this thesis by focusing on the following research:

7.2.1 Develop a cumulative QoE database

It was difficult to conduct a medium to large scale experiment for gathering cumulative QoE evaluations. Further studies should be carried out to use a larger cumulative QoE database in order to develop more accurate prediction models and obtain more data for analyzing the impacts of human-related factors on the user's perception.

7.2.2 Consider more QoE influence factors

In this thesis, human-related factors were considered in QoE modeling and show promising results in improving the QoE prediction accuracy. However, there are many QoE influence factors (e.g., context, content-related) that have not been taken into account since it is challenging to obtain this information from the users. In order to accurately measure the user's QoE, further studies should investigate more QoE influence factors and apply those factors in QoE modeling.

References

[1] N. Eswara, S. Ashique, A. Panchbhai, S. Chakraborty, H. P. Sethuram, K. Kuchi, A. Kumar, and S. S. Channappayya. Streaming video qoe modeling and prediction: A long short-term memory approach. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2019. ISSN 1051-8215. doi: 10.1109/TCSVT.2019.2895223.

[2] Cisco Visual Networking Index. Cisco visual networking index: Forecast and methodology, 2016–2021. *Complete Visual Networking Index (VNI) Forecast*, 12(1):749–759, 2017.

[3] Chanh Minh Tran, Tho Nguyen Duc, Phan Xuan Tan, and Eiji Kamioka. Qabr: A qoe-based approach to adaptive bitrate selection in video streaming services. *International Journal of Advanced Trends in Computer Science and Engineering*, 8:138–144, 09 2019. doi: 10.30534/ijatcse/2019/2181.42019.

[4] Jie Liu, Xiaoming Tao, and Jianhua Lu. Qoe-oriented rate adaptation for dash with enhanced deep q-learning. *IEEE Access*, 12 2018. doi: 10.1109/ACCESS.2018.2889999,.

[5] Tobias Hoßfeld, Michael Seufert, Christian Sieber, and Thomas Zinner. Assessing Effect Sizes of Influence Factors Towards a QoE Model for HTTP Adaptive Streaming. *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 111–116, 2014. doi: 10.1109/QoMEX.2014.6982305.

[6] Chao Chen, Lark Kwon Choi, Gustavo de Veciana, Constantine Caramanis, Robert W. Heath, and Alan C. Bovik. Modeling the Time-Varying Subjective Quality of HTTP Video Streams With Rate Adaptations. *IEEE Transactions on Image Processing*, 23 (5):2206–2221, may 2014. ISSN 1057-7149. doi: 10.1109/TIP.2014.2312613. URL http://ieeexplore.ieee.org/document/6775292/.

[7] Deepti Ghadiyaram, Janice Pan, and Alan C. Bovik. Learning a Continuous-Time Streaming Video QoE Model. *IEEE Transactions on Image Processing*, 27(5):2257–2271, may 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2790347. URL http://ieeexplore.ieee.org/document/8247250/.

[8] Christos George Bampis, Zhi Li, Anush Krishna Moorthy, Ioannis Katsavounidis, Anne Aaron, and Alan Conrad Bovik. Study of Temporal Effects on Subjective Video Quality of Experience. *IEEE Transactions on Image Processing*, 26(11):5217–5231, nov 2017. ISSN 1057-7149. doi: 10.1109/TIP.2017.2729891. URL http://ieeexplore.ieee.org/document/7987076/.

[9] Tho Nguyen Duc, Chanh Minh Tran, Phan Xuan Tan, and Eiji Kamioka. Bidirectional lstm for continuously predicting qoe in http adaptive streaming. In *Proceedings of the 2019 2nd International Conference on Information Science and Systems*, pages 156–160, 2019.

[10] T. N. Duc, C. T. Minh, T. P. Xuan, and E. Kamioka. Convolutional neural networks for continuous qoe prediction in video streaming services. *IEEE Access*, 8:116268–116278, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.3004125.

[11] Tho Nguyen Duc, Chanh Minh Tran, Phan Xuan Tan, and Eiji Kamioka. Modeling of cumulative qoe in on-demand video services: Role of memory effect and degree of interest. *Future Internet*, 11(8):171, 2019.

[12] Patrick Le Callet, Sebastian Möller, Andrew Perkis, et al. Qualinet white paper on definitions of quality of experience. *European network on quality of experience in multimedia systems and services (COST Action IC 1003)*, 3(2012), 2012.

[13] International Telecommunication Union. Recommendation ITU-T P,10/G.100 Vocabulary for Performance, Quality of Service and Quality of Experience. (P.10/G.100), 2017.

[14] International Telecommunication Union. Methods for subjective determination of transmission quality. *International Telecommunication Union, Geneva, Switzerland, ITU-Recommendation*, 800:22, 1996.

[15] N. Barman and M. G. Martini. Qoe modeling for http adaptive video streaming-a survey and open challenges. *IEEE Access*, 7:30831–30859, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2901778.

[16] M. N. Garcia, W. Robitza, and A. Raake. On the accuracy of short-term quality models for long-term quality prediction. In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 1–6, May 2015. doi: 10.1109/QoMEX.2015.7148123.

[17] Y. Shen, Y. Liu, Q. Liu, and D. Yang. A method of qoe evaluation for adaptive streaming based on bitrate distribution. In *2014 IEEE International Conference on Communications Workshops (ICC)*, pages 551–556, June 2014. doi: 10.1109/ICCW.2014.6881256.

[18] Christos G. Bampis, Zhi Li, and Alan C. Bovik. Continuous Prediction of Streaming Video QoE Using Dynamic Networks. *IEEE Signal Processing Letters*, 24(7):1083–1087, jul 2017. ISSN 1070-9908. doi: 10.1109/LSP.2017.2705423. URL http://ieeexplore.ieee.org/document/7931662/.

[19] Christos G Bampis and Alan C Bovik. An augmented autoregressive approach to http video stream quality prediction. *arXiv preprint arXiv:1707.02709*, 2017.

[20] Nagabhushan Eswara, Hemanth P. Sethuram, Soumen Chakraborty, Kiran Kuchi, Abhinav Kumar, and Sumohana S. Channappayya. Modeling Continuous Video QoE Evolution: A State Space Approach. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, jul 2018. ISBN 978-1-5386-1737-3. doi: 10.1109/ICME.2018.8486557. URL https://ieeexplore.ieee.org/document/8486557/.

[21] Yanwei Liu, Song Ci, Hui Tang, Yun Ye, and Jinxia Liu. Qoe-oriented 3d video transcoding for mobile streaming. *ACM Trans. Multimedia Comput. Commun. Appl.*, 8(3s), October 2012. ISSN 1551-6857. doi: 10.1145/2348816.2348821. URL https://doi.org/10.1145/2348816.2348821.

[22] W. Shi, Y. Sun, and J. Pan. Continuous prediction for quality of experience in wireless video streaming. *IEEE Access*, 7:70343–70354, 2019.

[23] Michael Seufert, Sebastian Egger, Martin Slanina, Thomas Zinner, Tobias Hoßfeld, and Phuoc Tran-Gia. A Survey on Quality of Experience of HTTP Adaptive Streaming. *IEEE Communications Surveys & Tutorials*, 17(1):469–492, 2015. ISSN 1553-877X. doi: 10.1109/COMST.2014.2360940. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6913491.

[24] R. K. P. Mok, E. W. W. Chan, and R. K. C. Chang. Measuring the quality of experience of http video streaming. In *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, pages 485–492, May 2011. doi: 10.1109/INM.2011.5990550.

[25] Zhiyong Cui, Ruimin Ke, Ziyuan Pu, and Yinhai Wang. Deep bidirectional and unidirectional lstm recurrent neural network for network-wide traffic speed prediction. *arXiv preprint arXiv:1801.02143*, 2018.

[26] Alex Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. Hybrid speech recognition with deep bidirectional lstm. In *2013 IEEE workshop on automatic speech recognition and understanding*, pages 273–278. IEEE, 2013.

[27] Zhou Yu, Vikram Ramanarayanan, David Suendermann-Oeft, Xinhao Wang, Klaus Zechner, Lei Chen, Jidong Tao, Aliaksei Ivanou, and Yao Qian. Using bidirectional lstm recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech. In *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pages 338–345. IEEE, 2015.

[28] R. Soundararajan and A. C. Bovik. Video quality assessment by reduced reference spatio-temporal entropic differencing. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4):684–694, April 2013. ISSN 1051-8215. doi: 10.1109/TCSVT.2012.2214933.

[29] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, volume 2, pages 1398–1402 Vol.2, Nov 2003. doi: 10.1109/ACSSC.2003.1292216.

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack. Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19(6):1427–1441, June 2010. ISSN 1941-0042. doi: 10.1109/TIP.2010.2042111.

[31] Christos G. Bampis and Alan C. Bovik. Feature-based prediction of streaming video QoE: Distortions, stalling and memory. *Signal Processing: Image Communication*, 68(May):218–228, 2018. ISSN 09235965. doi: 10.1016/j.image.2018.05.017. URL https://doi.org/10.1016/j.image.2018.05.017.

[32] Deepti Ghadiyaram, Janice Pan, and Alan C. Bovik. A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1):183–197, 2019. ISSN 10518215. doi: 10.1109/TCSVT.2017.2768542. URL http://ieeexplore.ieee.org/document/8093636/.

[33] David S Hands and SE Avons. Recency and duration neglect in subjective assessment of television picture quality. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 15(6):639–657, 2001.

[34] Nagabhushan Eswara, K. Manasa, Avinash Kommineni, Soumen Chakraborty, Hemanth P. Sethuram, Kiran Kuchi, Abhinav Kumar, and Sumohana S. Channappayya. A Continuous QoE Evaluation Framework for Video Streaming Over HTTP. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(11):3236–3250, nov 2018. ISSN 1051-8215. doi: 10.1109/TCSVT.2017.2742601. URL https://ieeexplore.ieee.org/document/8013810/.

[35] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *CoRR*, abs/1803.01271, 2018. URL http://arxiv.org/abs/1803.01271.

[36] Pierre Dutilleux. An implementation of the "algorithme à trous" to compute the wavelet transform. In *Wavelets*, pages 298–304. Springer, 1990.

[37] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[38] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[39] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, June 2016. doi: 10.1109/CVPR.2016.90.

[40] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[41] Tim Salimans and Durk P Kingma. Weight normalization: A simple reparameterization to accelerate training of deep neural networks. In *Advances in Neural Information Processing Systems*, pages 901–909, 2016.

[42] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[43] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks, 2017.

[44] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.

[45] Yazan Abu Farha and Jurgen Gall. Ms-tcn: Multi-stage temporal convolutional network for action segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3575–3584, 2019.

[46] J. De Vriendt, D. De Vleeschauwer, and D. Robinson. Model for estimating qoe of video delivered using http adaptive streaming. In *2013 IFIP/IEEE International Symposium on Integrated Network Management (IM 2013)*, pages 1288–1293, May 2013.

[47] Y. Liu, S. Dey, F. Ulupinar, M. Luby, and Y. Mao. Deriving and validating user experience model for dash video streaming. *IEEE Transactions on Broadcasting*, 61(4): 651–665, Dec 2015. ISSN 0018-9316. doi: 10.1109/TBC.2015.2460611.

[48] D. Zegarra Rodríguez, R. Lopes Rosa, E. Costa Alfaia, J. Issy Abrahão, and G. Bressan. Video quality metric for streaming service using dash standard. *IEEE Transactions on Broadcasting*, 62(3):628–639, Sep. 2016. ISSN 0018-9316. doi: 10.1109/TBC.2016.2570012.

[49] Kalpana Seshadrinathan and Alan C. Bovik. Temporal hysteresis model of time varying subjective video quality. In *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1153–1156. IEEE, may 2011. ISBN 978-1-4577-0538-0. doi: 10.1109/ICASSP.2011.5946613. URL http://ieeexplore.ieee.org/document/5946613/.

[50] Tobias Hoßfeld, Sebastian Biedermann, Raimund Schatz, Alexander Platzer, Sebastian Egger, and Markus Fiedler. The memory effect and its implications on Web QoE modeling. *2011 23rd International Teletraffic Congress (ITC)*, 2011. URL https://www.semanticscholar.org/paper/The-memory-effect-and-its-implications-on-Web-QoE-Ho{\T1\ss}feld-Biedermann/9512a8c21c4c38b6258c529b2db5932b9df364d2.

[51] Christos G Bampis and Alan C Bovik. Learning to predict streaming video qoe: Distortions, rebuffering and memory. *arXiv preprint arXiv:1703.00633*, 2017.

[52] A J. Greene. Primacy versus recency in a quantitative model: Activity is the critical distinction. *Learning & Memory*, 7:48–57, 01 2000. doi: 10.1101/lm.7.1.48.

[53] Christos G Bampis, Zhi Li, Ioannis Katsavounidis, Te-Yuan Huang, Chaitanya Ekanadham, and Alan C Bovik. Towards perceptually optimized end-to-end adaptive video streaming. *arXiv preprint arXiv:1808.03898*, 2018.

[54] Geoffrey R. Loftus. Evaluating forgetting curves. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, pages 397–406, 1985.

[55] Hermann Ebbinghaus. Memory: a contribution to experimental psychology. *Annals of neurosciences*, 20(4):155–156, oct 2013. ISSN 0972-7531. doi: 10.5214/ans.0972.7531.200408. URL https://www.ncbi.nlm.nih.gov/pubmed/25206041https://www.ncbi.nlm.nih.gov/pmc/PMC4117135/.

[56] G. Ghinea and J. P. Thomas. Qos impact on user perception and understanding of multimedia video clips. In *Proceedings of the Sixth ACM International Conference on Multimedia*, MULTIMEDIA '98, pages 49–54, New York, NY, USA, 1998. ACM. ISBN 0-201-30990-4. doi: 10.1145/290747.290754. URL http://doi.acm.org/10.1145/290747.290754.

[57] Jong Seok Lee, Francesca De Simone, and Touradj Ebrahimi. Subjective quality evaluation VIA paired comparison: Application to scalable video coding. *IEEE Transactions on Multimedia*, 13(5):882–893, 2011. ISSN 15209210. doi: 10.1109/TMM.2011.2157333.

[58] Michal Ries, Peter Froehlich, and Raimund Schatz. QoE evaluation of high-definition IPTV services. *Proceedings of 21st International Conference, Radioelektronika 2011*, pages 15–20, 2011. doi: 10.1109/RADIOELEK.2011.5936485.

[59] Patrick Le Callet and Jenny Benois-Pineau. Visual content indexing and retrieval with psycho-visual models. In *Visual Content Indexing and Retrieval with Psycho-Visual Models*, pages 1–10. Springer, 2017.

[60] Bennet B Murdock Jr. The serial position effect of free recall. *Journal of Experimental Psychology*, 64(5):482–488, 1962. ISSN 0022-1015(Print). doi: 10.1037/h0045106.

[61] Dewey Rundus. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89(1):63–77, 1971. ISSN 0022-1015(Print). doi: 10.1037/h0031185.

[62] J.A. Murakowski, P.A. Wozniak, and E.J. Gorzelanczyk. Two components of long-term memory. *Acta Neurobiologiae Experimentalis*, 55:301 – 305, 1995.

[63] Tobias Hoßfeld, Raimund Schatz, and Sebastian Egger. SOS: The MOS is not enough! *2011 3rd International Workshop on Quality of Multimedia Experience, QoMEX 2011*, pages 131–136, 2011. doi: 10.1109/QoMEX.2011.6065690.

[64] Fabrice Bellard. Ffmpeg multimedia system. *FFmpeg. [Last accessed: June 2019]. https://www.ffmpeg.org/about.html*, 2005.

[65] ITU-T. Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in any Environment. *Recommendation ITU-T P.913*, 2016. ISSN 13921215. doi: 10.1109/URSI-AT-RASC.2015.7303081. URL https://www.itu.int/rec/T-REC-P.913-201401-I/en.

[66] MATLAB. *version 9.6.0 (R2019a)*. The MathWorks Inc., Natick, Massachusetts, 2019.

[67] Jingteng Xue, Dong-Qing Zhang, Heather Yu, and Chang Wen Chen. Assessing quality of experience for adaptive HTTP video streaming. In *2014 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6. IEEE, jul 2014. ISBN 978-1-4799-4717-1. doi: 10.1109/ICMEW.2014.6890604. URL http://ieeexplore.ieee.org/document/6890604/.