QoE Evaluation of Multimedia Services Based on Audiovisual Quality and User Interest

Jiarun Song, Fuzheng Yang, *Member, IEEE*, Yicong Zhou, *Senior Member, IEEE*,
Shuai Wan, *Member, IEEE*, and Hong Ren Wu

***Abstract*—Quality of experience (QoE) has significant influence on whether or not a user will choose a service or product in the competitive era. For multimedia services, there are various factors in a communication ecosystem working together on users, which stimulate their different senses inducing multidimensional perceptions of the services, and inevitably increase the difficulty in measurement and estimation of the user's QoE. In this paper, a user-centric objective QoE evaluation model (QAVIC model for short) is proposed to estimate the user's overall QoE for audiovisual services, which takes account of perceptual audiovisual quality (QAV) and user interest in audiovisual content (IC) amongst influencing factors on QoE such as technology, content, context, and user in the communication ecosystem. To predict the user interest, a number of general viewing behaviors are considered to formulate the IC evaluation model. Subjective tests have been conducted for training and validation of the QAVIC model. The experimental results show that the proposed QAVIC model can estimate the user's QoE reasonably accurately using a 5-point scale absolute category rating scheme.**

***Index Terms*—Audiovisual quality, audiovisual services, quality of experience (QoE), user interest, viewing behavior.**

I. Introduction

WITH the development of advanced multimedia and network technologies, audiovisual services have become more accessible to users than ever before, accompanied by increasing competition between product or system manufactures as well as service providers which has inspired various audiovisual applications in audiovisual communications, broadcasting, entertainment, and recreation audio, video and photography. Since user perception and satisfaction are crucial determinants for the success of a product, application and service in the marketplace, more and more multimedia service providers steer their focus on assessment and prediction of the user's quality of experience (QoE) for audiovisual services [1], [2].

In the new era of customization and pervasive computing, all users expect customized or personalized service whenever a requested service is being delivered [3]. Therefore, knowing what the user wants, desires, and appreciates is the highest realm of a service where providers can provide customized service delivery. Taking the audiovisual service as an example, the user perception and experience are always different for countless audiovisual clips with various contents and quality levels. By estimating the user experience, providers can provide users with more customized services according to their preference, in terms of content recommendation, quality improvement, price adjustment, and so on. Generally, the user experience can be used as a guideline for service providers to improve their services.

Traditionally, technology-centric quality metrics were used to measure and monitor the audiovisual services in terms of quality of service (QoS), such as the packet loss, delay, and available bandwidth, etc. [4], to meet the needs or requirements of the users and applications. From the application point of view, video quality, audio quality, and audiovisual quality were usually employed as the metrics to evaluate multimedia services [5]. However, these quality metrics primarily target at improving product, system or service quality with respect to network- and application-level technical parameters, lacking sufficient consideration of user's actual perceptions and feelings [6]. To solve this problem, the user-centric notion of QoE was introduced to measure the service quality as perceived by users, which gradually became a research focus in both academia and industry [2]. QoE was defined by ITU-T, SG 12 (International Telecommunication Union Telecommunication Standardization Sector, Study Group 12) as "the overall acceptability of an application or service, perceived subjectively by the end-user" [7]. A conceptual decomposition framework for QoE in communication ecosystem was introduced in [2] to clarify and crystalize the concept of QoE and to facilitate research efforts in the area. A hypothesis of quality formation process was presented and the definition of QoE was further revised in [8] as "the degree of delight or annoyance of the user for an application or service, and it resulted from the fulfillment of the user's expectations with respect to the utility and/or enjoyment of the service in the light of the user's personality and current state". QoE is a fast emerging multidisciplinary field based on social psychology, cognitive science, economics, and engineering science, with a focus on understanding overall human quality requirements [9]. For service providers, it seems to make more sense to migrate

Manuscript received May 11, 2015; revised August 29, 2015 and January 13, 2016; accepted January 14, 2016. Date of publication January 21, 2016; date of current version February 18, 2016. This work was supported in part by the NSF of China under Grant 61371089 and Grant 61571337, and in part by the 111 Project B08038. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Christian Timmerer.

J. Song and F. Yang are with the Collaborative Innovation Center of Information Sensing and Understanding and ISN, Xidian University, Xi'an 710071, China (e-mail: jrsong@stu.xidian.edu.cn; fzhyang@mail.xidian.edu.cn).

Y. Zhou is with the Department of Computer and Information Science, University of Macau, Macau, China (e-mail: yicongzhou@umac.mo).

S. Wan is with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: swan@nwpu.edu.cn).

H. R. Wu is with the School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC 3001, Australia (e-mail: henry.wu@rmit.edu.au).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes a demo of the experiment results. This material is 1.6 MB in size.

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

Digital Object Identifier 10.1109/TMM.2016.2520090

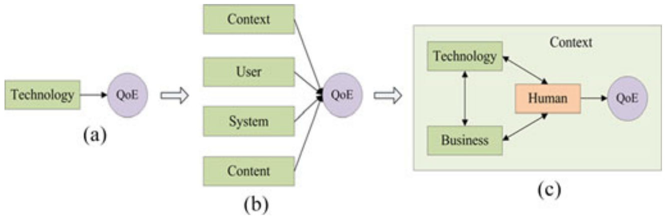


Fig. 1. Evolution of the QoE assessment. (a) QoE evalution by technology factor. (b) Taxonomy of QoE. (c) Communication ecosystem.

from the traditional quality evaluation methods which focus on the service performance to the QoE evaluation which aims at the user's perception and needs.

Generally, the assessment of QoE can be done subjectively or objectively [10]. For the subjective assessment, it extends beyond the user-perceived media quality to include measures such as user interest and user satisfaction, which can be obtained via questionnaires and rating scales [11]. Although the subjective evaluation may be the only way to obtain the QoE closest to the "ground truth", it is extremely expensive and time-consuming to perform subjective evaluation. The vast majority of users are reluctant to enter explicit ratings about their perception during the services because it can interrupt normal patterns of watching and may impose an additionally cognitive load on them [12]. On the other hand, the objective measures of QoE use different models of human perceptions, and try to estimate the performance of audiovisual services to approximate the subjective QoE measure in an automated manner without human involvement [10]. Compared with the subjective methods, objective QoE assessment is more efficient, and is of significant importance for service providers.

To date, there have been several objective methods to evaluate QoE. Most of them directly map the QoS parameters or media-related parameters to QoE using a certain function (e.g., the exponential model) [13], [14], and usually cannot evaluate the QoE accurately because these methods do not consider user perception. It was emphasized in [15] that QoE cannot be understood as a singular objective quality parameter of the services, but must take into consideration of every factor that contributes to the service quality perception by subjects. The taxonomy of QoE which considers influential factors was proposed to video services, including context, user, technical system and content [15]. In 2008, QoE was examined for the first time as a central concept for analysis of the entire communications ecosystem [2]. Generally, the communications ecosystem covers a huge area from technical issues to business models and human behaviors. In the context of communications ecosystem, a holistic and unified QoE model was then proposed in [8], where the user's QoE was interactively affected by various factors in different domains such as technology, business, context, and human. As shown in Fig. 1, the evolution of the QoE assessment can be deduced from the single technological parameter estimation to multiple interactive parameters estimation and from technology-centric to user-centric. With well considerations of the communication ecosystem, the aforementioned research publications have clarified the concept of QoE. Nevertheless, they have not provided a practical solution to objectively estimate QoE. The objective assessment of QoE is nontrivial due to the fact that multiple interactive factors are involved in the complicated service context. How to objectively evaluate QoE remains as an open problem and a challenge.

In this paper, an objective QoE assessment model (QAVIC model) for audiovisual services is proposed with the idea of communication ecosystems, where the influence of the technology, content, and human domains are analyzed. Different from those methods directly mapping influential factors to QoE [e.g., Fig. 1(b)], where the factors are considered to be unrelated to each other, the proposed framework is based on the user-centric framework where the influential factors in each domain are interactive with the human domain (e.g., Fig. 1(c), where the business domain is replaced by the content domain). The user's QoE is evaluated mainly by analyzing user perceptions with regard to different dimensions, such as the perception of audiovisual quality and user interest in content and/or story. These perceptions are usually stimulated by multiple influential factors. For example, the perception of audiovisual quality is mostly influenced by the factors in both the technology and user domains, while the user interest in audiovisual content is affected by the factors in both the content and the user domains. Each perception will be analyzed respectively and finally combined together to form the QoE. The contributions of this work include: 1) a prediction model for user interest in audiovisual content (IC) based on analysis of general user's viewing behaviors, and 2) a user-centric objective QoE evaluation model (QAVIC) to estimate the user's overall QoE for audiovisual service, which takes account of perceptual audiovisual quality (QAV) and user interest in audiovisual content, rather than focusing on the user satisfaction with technological quality levels like in the traditional methods.

The remainder of this paper is organized as follows. Section II introduces the QoE assessment framework for audiovisual services adopted for this investigation. Section III describes the experiment design, including how to construct a living lab as a test platform and how to detect users' viewing behaviors. In Section IV, different perceptions of the audiovisual services by users are analyzed and a new QoE evaluation model is proposed based on the perceptual QAV metric and IC prediction model. The experimental results are provided in Section V. Conclusion are drawn in Section VI.

II. Framework of QoE Evaluation System

A well-constructed QoE evaluation framework can help the service providers to identify, understand, and quantify the most influential aspects which affect user perceptions [16]. It also helps to clarify the relationship between the influential factors, user perceptions, and the resultant user's QoE. Here, a user-centric framework for the objective QoE assessment model is proposed, as shown in Fig. 2, which focuses on the actual generation process of user perception in audiovisual services. In this framework, user perceptions are affected by both external and internal factors, and formed in different dimensions by a complex psychological process with the influences of these factors. The composition of this evaluation framework is discussed in detail as follows.

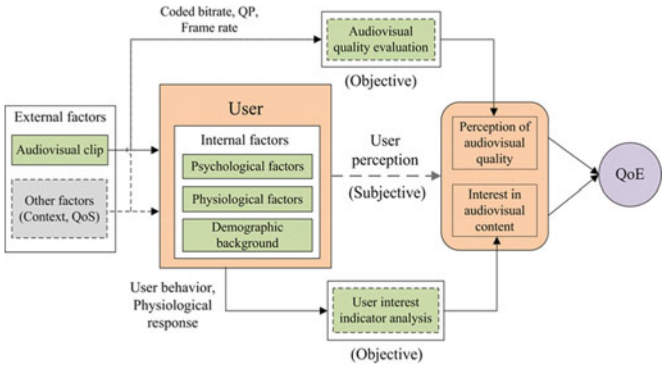

Coded bitrate, QP, Frame rate
Audiovisual quality evaluation
(Objective)
External factors
Audiovisual clip
Other factors (Context, QoS)
User
Internal factors
Psychological factors
Physiological factors
Demographic background
User perception
(Subjective)
Perception of audiovisual quality
Interest in audiovisual content
QoE
User behavior, Physiological response
User interest indicator analysis
(Objective)

Fig. 2. QoE assessment framework for audiovisual services.

A. *Influential Factors of QoE*

The external and internal factors are two main aspects which predominantly influence user's QoE. For audiovisual services, the primary external factor is the audiovisual clip, which contains contents carried by audiovisual signals and directly stimulates sensors and faculties of users to perceive. Other external factors can be further extended with respect to different multimedia application scenarios. For example, QoS factors and context have to be considered for audiovisual streaming over networks, and business factors need to be taken into account for the multimedia services which require payment. In the current work, the main focus will be on the basic aspects of audiovisual services such as the audiovisual quality and content, while the other external factors are kept constant wherever and whenever possible (e.g., constant context, free services and so on).

The internal aspect contains the psychological and physiological factors as well as demographic background in the human domain. These factors are closely related to users themselves and most influential in deliverance or outcomes of users' QoE. The psychological factors include the user's mood, curiosity, and spirit, and so on. The physiological factors include the user's health, fatigue states, and so on. The demographic background includes the user's age, gender, education background, and so on. These external and internal factors interact with each other to stimulate the user perceptions of the services [9].

B. *QoE Measurement From Multiple Dimension Perceptions*

Considering user's diverse concerns about the services, user perceptions can be divided into multiple dimensions of the interactions between users and the services. According to [17], the user experience of a service or system can be evaluated from pragmatic and hedonic aspects. To better estimate the user's QoE of audiovisual services, two fundamental aspects/factors which affect QoE in human domain for almost all audiovisual services are considered in the proposed QoE prediction model, namely, the perception of audiovisual quality and user interest in audiovisual content.

As illustrated in Fig. 2, user perception can be estimated directly by the internal and external influential factors. However, the user's perceptive process is quite difficult to comprehend, and it is usually hard to determine the relationship between the influential factors and user perception (illustrated by the dashed arrow in Fig. 2). Moreover, it is extremely difficult, if not impossible, to quantify some of the user's internal factors such as life experience, disposition and so on, which further aggravates the difficulty of the objective QoE evaluation. To formulate the proposed QoE assessment model, a feasible way (illustrated by the solid arrow in Fig. 2) to estimate user perception is adopted based on psychophysical theory. Two aforementioned aspects/factors which affect user's QoE in human domain are discussed as follows.

1) Perception of Audiovisual Quality: As a performance metric in technology domain, audiovisual quality is one of the most essential of multimedia services. It may be affected in varying degrees by the coding strategy, network transmission, and device performance. When users subscribe to multimedia services, perception of the audiovisual quality will directly influence their QoE, and therefore it is indispensible for QoE evaluation.

In the past two decades, a number of audiovisual quality assessment models have been proposed targeting different application scenarios [18]–[20]. The audiovisual quality is mainly evaluated by the video quality, audio quality, and interaction and integration of the two metrics [5]. More specifically, the video and audio quality degradations are usually caused in the coding and transmission processes, and these distortions can be evaluated using the application and network-level parameters. With respect to the interaction between the audio and video quality, it has been estimated based on a multiplicative model with audio quality and video quality as input variables [21]–[23]. The synchronization between audio and video (e.g., lip sync) is also an important interaction issue when assessing the audiovisual quality. The influence of asynchrony to the audiovisual quality has also been widely investigated [24]. Additionally, the human characteristics are also taken into consideration to estimate perceptual quality. For example, modeling the spatial and temporal just-noticeable-difference (JND) has been reported based on the sensitivity of HVS to the luminance contrast [25]. With the development of computational techniques for visual attention modeling [26], the attention properties have also been considered in image and video coding applications [27], [28]. Recently, there is a new trend to evaluate user perception of audiovisual quality using electroencephalography (EEG) and other physiological measurement devices [29], and experiment results show high correlation values between subjective and physiological data. There have been intense research efforts and standardization activities regarding audiovisual quality assessment and the related standards can be found in [18], [19].

The main emphasis of this research is to identify the relationship between the audiovisual quality and QoE, while the perceptual audiovisual quality is estimated in virtue of the subjective test. Considering the effects of coding distortion may be different from that caused by other sources, only the influence of coding distortion will be considered here for simplicity, while distortions caused by other sources, may be investigated further based on the proposed model.

2) User Interest in Audiovisual Content: Content is another basic attribute of audiovisual services. The user interest in audiovisual content indicates how the content/story appeals to the user. It also decisively influences the user's QoE from the hedonic point of view [30]. However, as one of human internal states, user interest is usually difficult to measure quantitatively. How to estimate user interest in audiovisual content is a key issue to formulate the QoE evaluation model.

In psychology, drive theory discusses how a person's internal state affects a person's behavior while incentive theory discusses how an external stimulus affects a person's behavior [31]. User interest in something is associated with a number of user's bodily behaviors such as laughing, more fixations, few blinks, lively movements of shoulders and head-nods [32]. Accordingly, the user interest may be estimated using these self-behaviors. For audiovisual services, it was proposed in [33] that the user interest can be estimated using the "gazing rate". In [34], the user's operational behaviors to audiovisual services were collected to evaluate the user interest, including the watching duration, duration of fast forward, and so on. There have also been other reports to estimate the user interest and emotion level with more elaborated physiological responses, e.g., the brain waves, galvanic skin responses, and so on [35].

In this paper, the term "user interest" is defined as a physically expressed state of concentration which can be visually recognized when a user is involved in the audiovisual content/story [33]. Generally, a person's internal states tend to be most likely expressed by his eyes [33]. The line of sight and interval between blinks has been used as a feature to estimate a person's concentration in daily life [36], [37]. Thus, a number of common viewing behaviors such as blink, fixation, and saccade are analyzed and used in this paper to indicate the user interest in audiovisual content. Detail analyses of user interest and viewing behaviors will be presented in Section IV.

III. Design of Experiments

A. *Test Platform Design*

Test platform is indispensable for the QoE research. In traditional QoE evaluation of audiovisual services, the procedure and environment of subjective experiments are strictly controlled, leading to a rough estimation of user's QoE. There is an urgent need for studying QoE in real-life and realistic settings. Recently, the concept of the living lab is emerging to address this issue. According to [38], [39], the living lab can be defined from different points of view. It may refer to 1) a user-centric test and environment platform based on the real world settings, or 2) a research and development methodology where innovations are created and validated with user participation. Because the living lab has multiple roles and functions, it can be employed for innovating, sensing, prototyping, validating and refining complex solutions in evolving real life contexts [38].

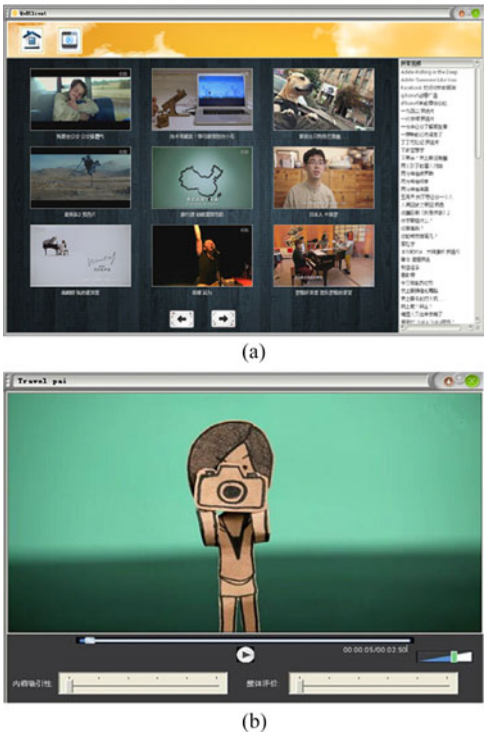


Fig. 3. Interface of the client software. (a) Main window. (b) Display window.

In this paper, the living lab is a test platform. A living lab (e.g., a study) was established to simulate the environment in the real world where audiovisual services are usually enjoyed. The living lab designed for QoE evaluation experiments contained a bookcase, a computer, a table, a light, as well as some flowers and pictures. Room luminosity was between 300–400 Lux. The noise of the living lab was 40–50 dB. Different from the traditional strictly controlled test platform, the living lab provided a familiar context of daily life for users (e.g., surrounding and lighting). The users could choose a comfortable way to enjoy the audiovisual services as they like without any restriction with regard to, e.g., viewing distance, posture or order of viewing audiovisual clips.

A number of audiovisual clips with a variety of quality and content were stored on a local server, and the users could enjoy these audiovisual clips by logging into a client computer. The interface of the designed client software is shown in Fig. 3. Different from the traditional subjective test, the user could operate the software to control the play progress and to rate his/her interest in the content and QoE values. Meanwhile, a high-definition camera with the frame rate of 30 fps (frame per second) was installed in the front of the user to capture videos of his/her face area, which was used to detect the user's viewing behaviors such as blinks and eye movement. The information of viewing behaviors and the scores of the subjective QoE were sent to the server for statistics collection and analysis.

B. *Subjective Test Design*

Several subjective tests were carried out to obtain the rating data on audiovisual quality, user interest and QoE. There were two separate sessions, i.e., the QoE assessment and audiovisual quality assessment. The QoE assessment was carried out in the living lab. This is because the traditional controlled laboratory environment may influence the user's true perceptions of the services (e.g., the viewing distance and posture are strictly controlled, the order of viewing clips is pre-designed, the user is in a closed environment, and so on). The audiovisual quality assessment was carried out following the specifications of ITU-T recommendation P.913 [40]. Moreover, considering that the user interest and QoE may be affected if the audiovisual clip had been watched before, the QoE assessment session was carried out before the audiovisual quality session.

A total of 120 audiovisual sequences with a spatial resolution of 672 × 378 pixels and an aspect ratio of 16:9 were downloaded from the online video site Youku (www.youku.com) and employed in the experiments, including a variety of contents such as advertisement, sport, short film, speech, animation, movie trailer, and so on. This resolution was usually used for video applications subject to heavy network traffic and low transmission bandwidth. The length of each audiovisual clip was about 5 to 10 minutes. These clips were randomly divided into four groups (30 videos per group), and marked with TG1, TG2, TG3 and TG4, respectively. The videos in different groups had different video quality levels modified by coding with quantization parameters (QP) 37, 32, 27 and 24, respectively, using the reference software of FFmpeg 0.4.9 with $\times .264$ library.[1] The frame rate of each video was 25 frames per second and the number of reference frame was 4. Meanwhile, the corresponding audio sequences were coded by the HE-AAC (high-efficiency advanced audio coding) codec with a constant bit-rate of 64 kbps [41]. Each video sequence was encapsulated into FLV (flash video) format with the corresponding audio sequence, and audio and video signals were synchronous. The monitor for display was a 22-in LCD flat panel, and the video clip was shown in its original resolution on the screen.

There were 60 people aged 20 and 32 years old in the subjective tests (N $=60$, average age $=26.5$, standard deviation $=$ 2.3), including 28 females (46.7%) and 32 (53.5%) males. All participants were university students in different grades and were screened for visual acuity and color blindness. Users were classified into two types according to the amount of their spare time. For the users who had sufficient spare time, they were invited to randomly watch two non-adjacent groups of audiovisual clips from TG1 to TG4. Others were invited to watch one group only. Consequently, 40 users (66.7%) watched two groups of audiovisual clips and 20 users (33.3%) watched one group. For the clips in each group, there were 25 users to watch them. For the users who watched two groups, the videos were mixed together and the order of each clip was randomly set. The procedures of the two experiments are illustrated as follows.

1) During the QoE assessment session, both user interest in audiovisual content and QoE were rated. Users evaluated the audiovisual services in the living lab, where the environment was closer to the real world. They were all voluntary for the test to enjoy the audiovisual services in their leisure time. To avoid potential visual fatigue, the maximum duration of each viewing was 30 minutes and the users could stop watching at any time if they felt tired. Each user could attend the test at most 3 times a day, and the interval between two tests was set as at least 4 hours. It took several days for each user to finish the test. When users were enjoying the audiovisual services, their viewing behaviors were captured and stored in the database together with the related technical parameters for QoE prediction.

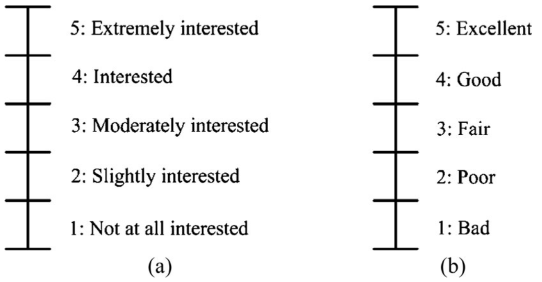


Fig. 4. Five-point scales for user's interest and QoE. (a) User interest. (b) QoE.

To the best of our knowledge, there was no specific recommendation with regard to methodology for the subjective test or assessment of user interest and QoE. Thus, two 5-point scale rating schemes were defined and used in the experiments to classify the user interest and QoE, following the general principle of the ITU-T recommendation P.913 [40], as illustrated in Fig. 4. Before the formal test, the users were invited to watch four examples with different contents and QP settings to get familiar with the control software and services. During the formal test, the users were instructed to watch each audiovisual clip once, and to rate both the user interest and QoE values immediately after watching the clip. There was no specific rating order for users to follow, i.e., either user interest or QoE could be rated first. To check the consistency of the rating results, 10 users were randomly chosen from all participants, and asked to rate the values of user interest and QoE of 5 clips again $(\mathrm{N}=50)$ after one week from the end of their last formal test, using the clips which they had watched previously. If the rating values of user interest and QoE are the same with that rated in the formal QoE test, the rating results will be regarded as consistent. It was found that the consistency of the results could reach about 90% on average.

2) In a few days after the QoE assessment session, the users were asked to evaluate the audiovisual quality of the clips in the same groups which were used for their QoE assessment. In this session, they were required to focus only on the audiovisual quality with a viewing distance of 4H. All the clips were shortened to 20 seconds from the beginning of the original clips. There were no full episodes in these clips. The test environment was strictly controlled and set following the guidance of ITU-T recommendation P.913 [40]. A single-stimulus Absolute Category Rating (ACR) method with a 5-point scale was used for audiovisual quality assessment [42].

All subjective tests were conducted within a two month period, and yielded 3000 rating samples of the audiovisual quality, user interest, and QoE, respectively, which were check for their accuracy by the observer screening procedure specified by the ITU-T to form a training set for QoE assessment modeling. Here, the users were screened with regard to the accuracy of their rating values in terms of the audiovisual quality. The standard exclusion procedures were followed as specified in [43]. After this screening process, the rating samples of 4 out of 60 subjects (6.7%) were discarded. The rating samples of 1 subject were discarded and 24 users were left in each group. There were

[1] "FFmpeg," [Online]. Available: http://sourceforge.net/projects/ffmpeg

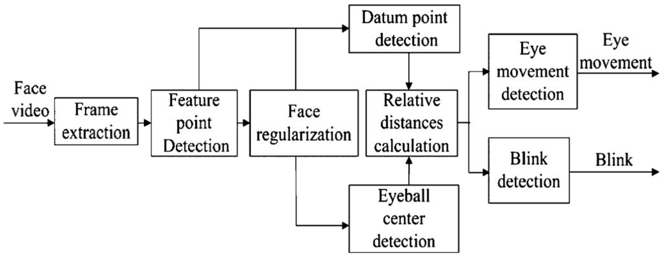


Fig. 5. Framework of viewing behavior detection.

2880 valid rating samples of the audiovisual quality, user interest, and QoE, respectively. Additionally, the audiovisual quality of each clip was measured in terms of the average score of all users, also known as the Mean Opinion Score (MOS) [40]. As a result, there were 120 MOS values for the audiovisual quality and 2880 rating values for the user interest and QoE, respectively.

C. Viewing Behavior Detection

Eyes play an essential role in daily life communications and convey the person's attentions, emotions, feelings, and so on [44]. As one of the most significant indicators for user's inner world, the viewing behaviors were analyzed in this paper to better understand user interest. Generally, viewing behaviors include eye movements and blinks. The eye movement can be further divided into fixation, smooth pursuit and saccade [45].

In the experiment, a viewing behavior detection method was proposed using an ordinary video camera, which was simple to deploy in real life and robust to deal with situations such as head moving and tilting. The framework of the behavior detection method is illustrated in Fig. 5, which consists of a number of steps. At first, each frame was extracted from the recorded face video. The feature points in each frame were then detected by the scale-invariant feature transform (SIFT) algorithm to determine the position of the datum point and to regularize the face image. The eyeball center was detected using the regularized face image. The relative distance between the eyeball center and the datum point was calculated to detect the eye movement and blink.

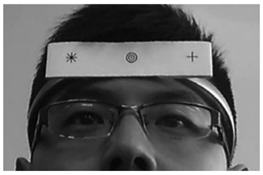


Fig. 6. Artifact used in the experiment.

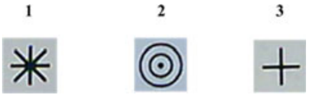


Fig. 7. Signs with different features.

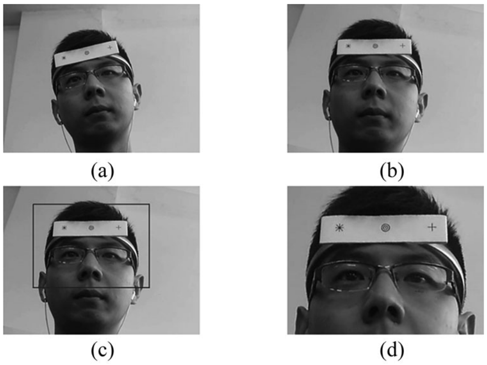


Fig. 8. Face regularization. (a) Original face frame. (b) Frame after affine transformation. (c) Clip region in the alignment frame. (d) Regularized face frame.

1) Datum Point Detection and Face Regularization: A datum point serves as a reference to measure other quantities [46], which is indispensable to describe the eye movement. Generally, there are two principles which govern the selection of the datum point: First, it must be a constant point that is measureable. Second, it must be easy to recognize and detect. The inherent features of the human face (e.g., eye corners and eyebrows) may vary with relaxation and contraction of facial muscles, which are not suitable to serve as a datum point. To solve this problem, the users in the experiment were asked to wear a lightweight artifact when they used audiovisual services, as shown in Fig. 6. According to the answers to a post-questionnaire, 81.6% of the users indicated that the artifact did not influence their watching experience. The other users reported that it felt a little strange but was still acceptable. On the artifact, there were three symbols with different features. They were numbered 1, 2, and 3, respectively, and placed horizontally with the equal intervals, as shown in Fig. 7. These symbols were used for two purposes. One was to determine the datum point by matching the feature points in the face image. The other was to create the rotation matrix to regularize a tilted face. The middle position between Symbol 1 and Symbol 3 was defined as the datum point, and Symbol 2 was reserved to calibrate the datum point if Symbol 1 or Symbol 3 incurred a mismatch.

First of all, the symbols in each frame should be detected. In the experiment, the SIFT methodology [47], which is widely used in computer vision to detect and describe local features in images, was employed to detect these symbols (using the C source code provided by OpenCV 2.4.5).[2] The accuracy of matching results in the test can achieve to 98.4% $(\mathrm{N} = 3000)$.

During watching, users may adjust their head or body unconsciously and the position of these symbols in each frame may be varying with the movement correspondingly, as shown in Fig. 8(a). In such a case, the datum point in each frame will not keep fixed, and may introduce errors in viewing behavior detection. To solve this problem, the face frame was regularized by an affine transformation to eliminate the influence of rotation and scaling. The affine transformation function can be expressed as follows:

$$\Theta = H \times \Pi = \begin{pmatrix} \Phi & \Psi \\ 0^T & 1 \end{pmatrix} \times \Pi \tag{1}$$

[2] [Online]. Available: http://opencv.org/

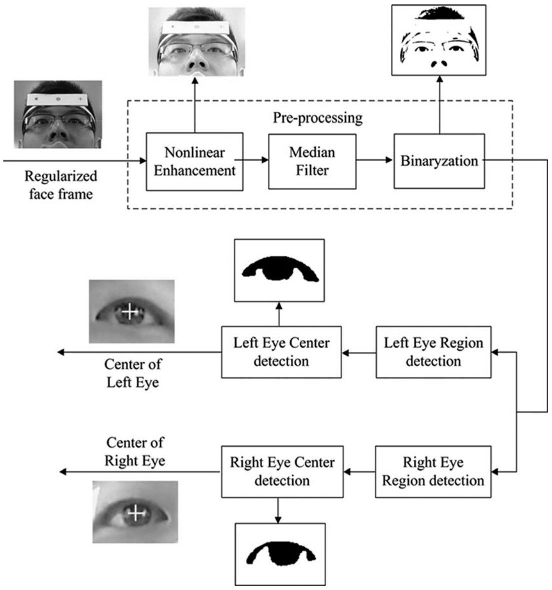


Fig. 9. Procedures of eyeball center detection.

where Θ is the point in the frame after the affine transformation, H is the affine transformation matrix, Π is the point in the original face frame, $\mathbf{0}^{T}$ is a 1×2 vector whose elements are 0, Φ is a 2×2 matrix, and Ψ is a 2×1 vector. The elements of Φ and Ψ can be determined using the method in [47].

After the affine transformation, the original face frame was rotated to alignment, as shown in Fig. 8(b). To ensure that the position of symbols is fixed in each frame, the rotated face frame was further regularized by clipping and resizing. According to (1), the position of each symbol in Fig. 8(b) was obtained. The face region was then clipped from Fig. 8(b) with the size of 400×300, where the left border of the clipped face region was 100 pixels far from Symbol 1 and the top border was 80 pixels far from Symbol 1, as illustrated in Fig. 8(c). The clipped region was finally resized to the regularization frame with the size of 640×480, as shown in Fig. 8(d). Considering the symmetry of Symbol 1 and Symbol 3, the datum point in Fig. 8(d) located at the fixed point (320, 128), which coincided with the center of Symbol 2.

2) Eyeball Center and Viewing Behavior Detection: Among various eyeball center detection methods, the projection function method is one of the simplest and the most effective. It is more suitable for the regularized face after the affine transformation which has better symmetry [48]. The projection method described in [48] was employed to detect the eyeball, whose procedures are illustrated in Fig. 9. It is worth noting that when users were blinking, the eyeball center was not detected by this method. In such a case, the eyeball centers and the datum point were assumed in the same horizontal line, namely, the fixed point (192, 128) for the right eyeball center and (448, 128) for the left eyeball center.

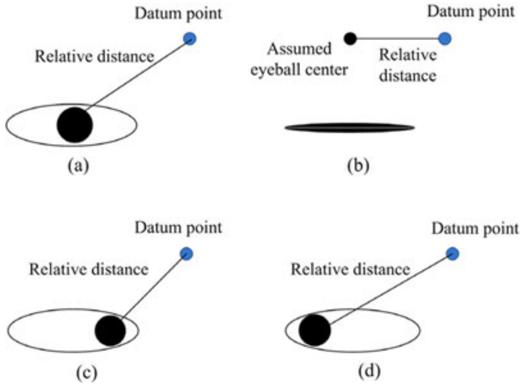


Fig. 10. Relative distances between the eyeball center and datum point for different viewing behaviors. (a) Look straight. (b) Eye closed. (c) Look left. (d) Look right.

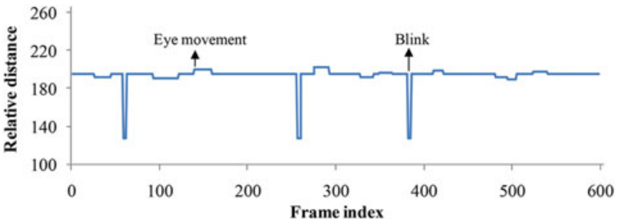


Fig. 11. Illustration of the relative distance in each frame for a recorded video.

The eye movement was detected by comparing the relative distance between the datum point and the eyeball center over time. Considering that movements of two eyes of a person are usually consistent, the right eye was taken as an example for further analysis. The relative distance $D(i)$ of the ith frame was calculated as

$$D(i) = \sqrt{(y_e(i) - y_d(i))^2 + (x_e(i) - x_d(i))^2} \tag{2}$$

where $x_e(i), y_e(i), x_d(i), y_d(i)$ were the horizontal and vertical coordinates of the datum point and eyeball center in the ith frame, respectively. Fig. 10 illustrates the relative distance for different viewing behaviors. It is obvious that when the user eyes move, the relative distance between the two points changes at the same time. Thus, both the eye movement and the blink can be detected by analyzing the relative distances.

Consequently, the number of eye movement, the duration and the amplitude of each eye movement can be obtained according to Fig. 11. Moreover, when users were blinking, the eyeball center was set at a fix point. In such a case, the corresponding value of the relative distance in the blink frame was significantly smaller than that in other frames (as shown by the sharp declines of the curve in Fig. 11). This difference can be used to distinguish blinking from other eye movements. Accordingly, the blink frequency and the duration between two blinks were obtained for the recorded video as well.

IV. QoE Evaluation of Audiovisual Services

In this section, we are going to analyze the influences of QoE for the audiovisual services considering both the audiovisual quality and user interest. Then, an objective model (QAVIC model) is proposed for QoE evaluation which can be used to help service providers to monitor and predict user perception and needs.

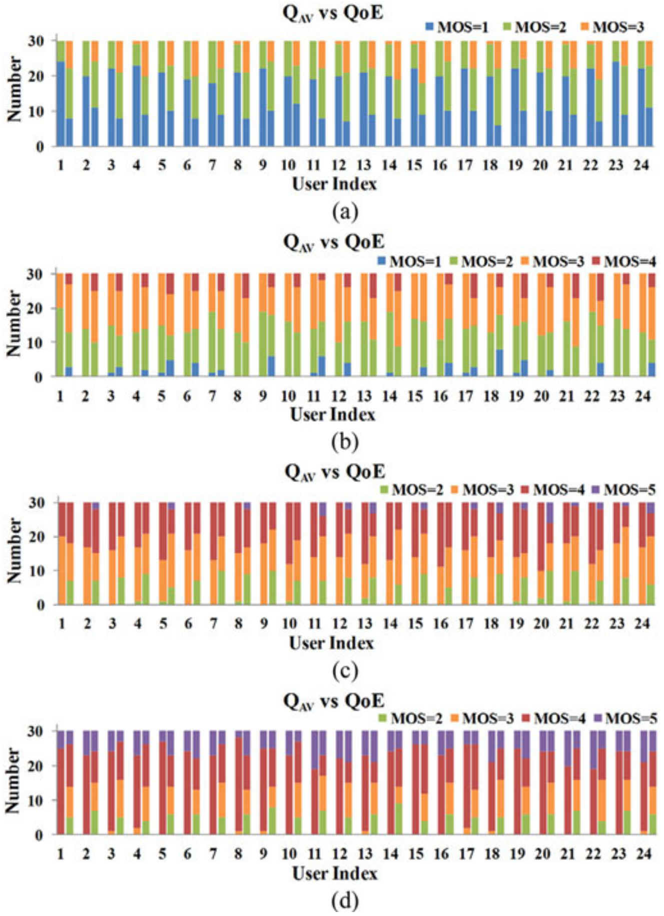


Fig. 12. Distributions of the audiovisual quality $Q_{\rm AV}$ and QoE in each group. (a) TG1 (F-test, $\rm F = 18.91,\ p < 0.01$). (b) TG2 $(\rm F = 26.25,\ p < 0.01)$. (c) TG3 $(\rm F = 30.38, p < 0.01)$. (d) TG4 $(\rm F = 8.41, p < 0.01)$.

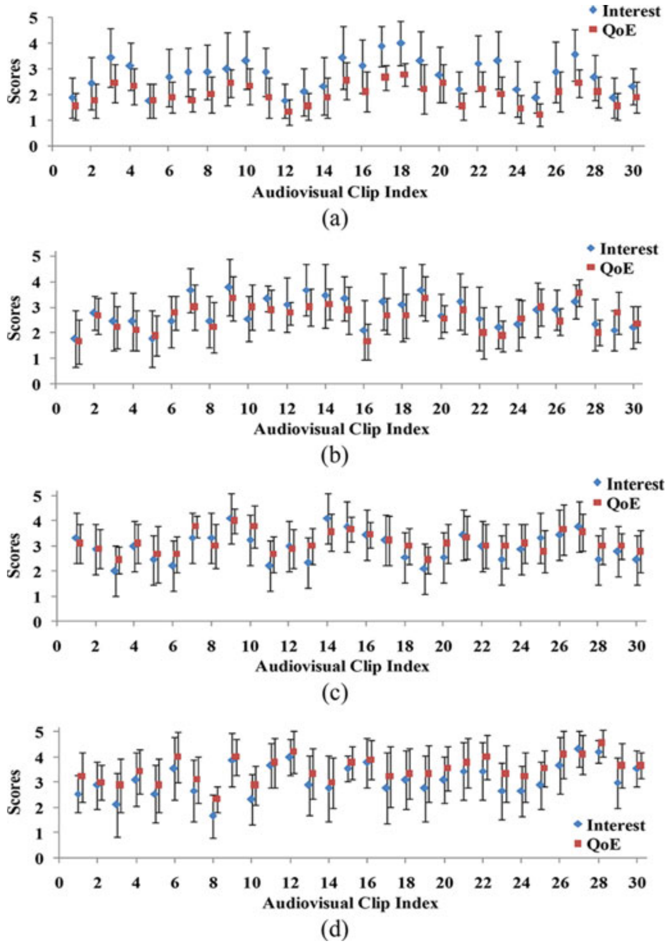


Fig. 13. Relationship between user interest and QoE in each group. (a) TG1 (F-test, $\rm F = 333.13,\ p < 0.01$). (b) TG2 $(\rm F = 218.32,\ p < 0.01)$. (c) TG3 $(\rm F = 206.59, p < 0.01)$. (d) TG4 $(\rm F = 758.86, p < 0.01)$.

A. QoE Analysis With Audiovisual Quality and User Interest

Audiovisual quality is a significant factor that may affect the user's QoE. Generally, a high-quality audiovisual service is more likely to provide users with a good experience. However, it is only a qualitative conclusion on their relationship, which is quite imprecise and unavailable for QoE evaluation. Here, we analyze the relationship between the audiovisual quality and QoE from the quantitative point of view.

The user's QoE is a personal property that is distinctive for different users. To accurately evaluate QoE, it is necessary to analyze the performance of all individual users. Fig. 12 shows the distribution of the QoE values and audiovisual quality $\rm Q_{AV}$ for each user in groups TG1, TG2, TG3 and TG4, respectively. Each user corresponds to two bars, the left bar indicates the number of audiovisual quality and the right bar indicates the number of QoE rated by the user. For each bar, there are several sub-bars with different colors that denote different rating scores. The height of each sub-bar indicates the number of audiovisual quality or QoE under a certain rating score. For instance, the left bar of User 1 in Fig. 12(a) means that, among all 30 values of the audiovisual quality in group TG1, there are 24 audiovisual clips rated 1 and 6 clips rated 2 for User 1. The right bar of User 1 in Fig. 12(a) indicates that there are 8 audiovisual clips rated 1, 14 clips rated 2, and 8 clips rated 3 among all QoE values rated by User 1. From Fig. 12(a)–(d), it can be found that the audiovisual quality in groups TG1, TG2, TG3 and TG4 gradually increase from $1 \sim 2$ to $4 \sim 5$. Meanwhile, the QoE values in groups TG1, TG2, TG3 and TG4 increase as well from $1 \sim 3$ to $2 \sim 5$. It shows that the audiovisual quality generally has a positive impact on QoE.

Additionally, according to the left and right bars of an individual user in each group, it is obvious that the audiovisual quality values are concentrated in a small range with a consistent level (two rating scores in each group), while the QoE values are quite distinct for different clips (up to four rating scores in each group). This difference indicates that the audiovisual quality itself is not sufficient for accurately estimating the user's QoE.

To further check the diversity of QoE, Fig. 13 presents the average values and standard deviations (SD) of all user interest and QoE values for each audiovisual clip. It can be found that the change of QoE values is in good accordance with that of user interest when the audiovisual quality is at a uniform level. Therefore, the user interest in content should also be taken into consideration for QoE evaluation. Moreover, for a given audiovisual clip, the user interest and QoE rating values usually vary significantly for different users (i.e., the SDs of user interest

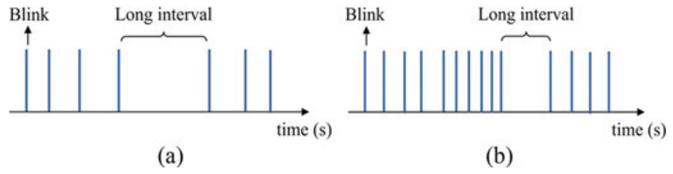


Fig. 14. Blink distributions of two users. (a) User 1. (b) User 2.

and QoE rating values are usually considerably large), reflecting diversity of their background and experience. It is necessary to evaluate the individual user interest and QoE combining his properties.

B. Viewing Behaviors and User Interest

In psychology, human's internal states can be evaluated using the human's behaviors on the basis of the drive theory and incentive theory. Here, some common viewing behaviors that tend to be unconsciously expressed are analyzed to indicate the user interest, including blink, fixation, and saccade. However, it is found that the eye movements like fixation and saccade are closely related to the temporal complexity or motion activity of a video. For example, if the temporal complexity of a video is low, such as the scene of news or concert, the user's fixations density usually tends to be higher and the saccade frequency is correspondingly lower than that of a video with high temporal complexity, such as the scene of sports and action movies. With respect to the blink, recent study has reported that the blink synchrony occurs only when subjects have to follow a storyline by extracting information from a stream of visual events, and the blink usually occurs during scenes that required less attention such as at the conclusion of an action, during the absence of the main character, during a long shot and during repeated presentations of a similar scene and so on [49]. Therefore, the blink may be more close to the user interest in audiovisual content. Next, the relationship between the blink and user interest will be analyzed. Particularly, here we just focus on the general features of the viewing behaviors for user communities.

When users are viewing an interesting audiovisual clip, fewer eye blinks will be drawn to prevent temporal loss of critical visual information [49]. Correspondingly, the time interval between two adjacent blinks will be longer. A long time interval between two adjacent blinks is usually a significant explicit indicator to measure user interest [37], [49]. However, the blink frequency is the inherent feature of individual, and leads to a difference in the average time interval between two adjacent blinks for different users. Fig. 14 illustrates the blink distributions for two users whose blink frequencies are different. It is obvious that the average time interval of blinks of User 1 is much larger than that of User 2. Considering that the long time interval between two adjacent blinks of a user is a value relative to the average blink interval, here we define it as follows:

$$\mathrm{Flag}_{\mathrm{LB}}(i) = \begin{cases} \text{True}, & \text{if } T_B(i) \geq T_B + m \cdot \sigma_B \\ \text{False}, & \text{otherwise} \end{cases} \tag{3}$$

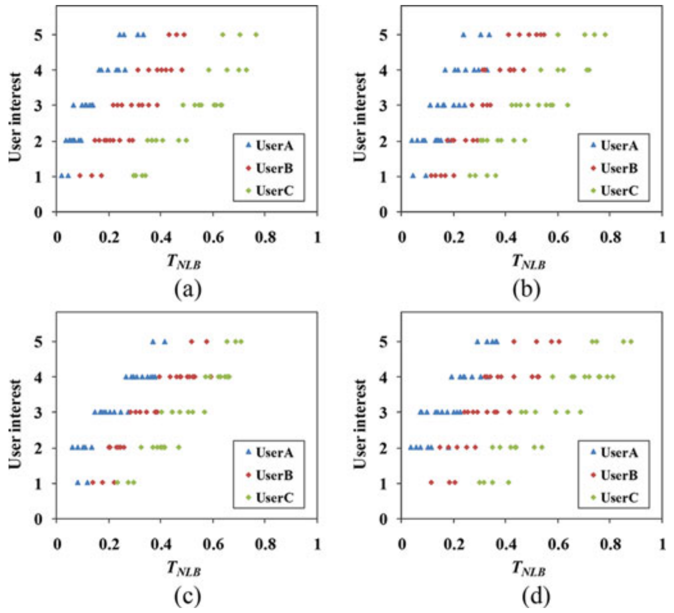


Fig. 15. Relationship between T_{NLB} and user interest in each group. (a) TG1. (b) TG2. (c) TG3. (d) TG4.

where $\mathrm{Flag}_{\mathrm{LB}}(i)$ is a flag to identify a long blink interval, T_B is the average interval of adjacent blinks and $T_B(i)$ is the interval between the ith and $(i-1)$th blinks, is the variance of the interval of adjacent blinks, m is a constant set to 3, which is empirically obtained by observations and experimental statistics. It is worth noting that the values of T_B for a user across different groups do not have significant differences (F-test, $\mathrm{p} > 0.05$). The value of T_B for each individual user is calculated by averaging the blink intervals in all video sequences. Correspondingly, the user's average blink frequency F_B is equal to $1/T_B$. There is one average blink interval and blink frequency for each user. As illustrated in Fig. 14, the long blink intervals for different users are not constant because the average blink intervals are diverse. For the users with a lower blink frequency, the long blink intervals are generally longer than the blink intervals of those with a higher blink frequency.

Accordingly, the total long blink interval in a recorded video sequence can be obtained by accumulating all long blink intervals in it. Considering that the lengths of recorded video sequences are different, the total long blink interval is normalized with the length of the recorded video sequence, which can be expressed as follows:

$$T_{\mathrm{NLB}} = \sum_{i=0}^{n} T_{\mathrm{LB}}(i) \Big/ T \tag{4}$$

where T_{NLB} is the total long blink interval normalized by the sequence length, $T_{\mathrm{LB}}(i)$ is the long interval between the ith and $(i-1)$th blinks, T is the duration of the recorded video, and n is the number of the long blink interval.

Fig. 15 illustrates the relationships between T_{NLB} and user interest for different users in each group, where each point indicates an audiovisual clip. It is obvious that the values of user interest are gradually increasing with the increment of T_{NLB}. Moreover, the ranges of T_{NLB} for different users are not the same. If a user has a high average blink frequency, the range of the total long blink interval is relatively smaller than that of a user with a low average blink frequency. For example, the average blink frequency F_B of User A and User C in Fig. 15(a) are 0.65 and 0.09, respectively, while the range of T_{NLB} of User A and User C are $0.02 \sim 0.33$ and $0.29 \sim 0.76$, respectively. The same phenomenon also appears in other groups. It seems that the diversity of T_{NLB} may be related to the difference of the individual's average blink frequency F_B.

Considering that the user interest is an ordinal outcome, the ordered logit regression (OLR) [50] is employed to determine the relationship between user interest, T_{NLB}, and F_B. It can be expressed as (5)

$$\begin{aligned}\mathrm{logit}(P(I_C > j)) &= \ln \frac{P(I_C > j)}{1 - P(I_C > j)}\\ &= \alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_j \Rightarrow P(I_C > j)\\ &= \frac{\exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_j)}{1 + \exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_j)}\\ &\quad \times (j = 1, 2, 3, 4) \qquad (5)\end{aligned}$$

where $P(\cdot)$ is the probability function, I_C denotes the user interest in audiovisual content, T_{NLB} and F_B are the explanatory variables. α_1 and α_2 are the logit coefficients. β_j is the threshold for user interest, which indicates the point (in terms of a logit) where user interest is predicted into the higher rating levels. The (5) further implies (6)

$$\begin{aligned}P(I_C = 1) &= \frac{1}{1 + \exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_1)}\\ P(I_C = j) &= \frac{\exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_{j-1})}{1 + \exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_{j-1})}\\ &\quad - \frac{\exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_j)}{1 + \exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_j)}\\ &\quad \times (j = 2, 3, 4)\\ P(I_C = 5) &= \frac{\exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_4)}{1 + \exp(\alpha_1 T_{\mathrm{NLB}} + \alpha_2 F_B - \beta_4)}. \qquad (6)\end{aligned}$$

It can be found that given the certain T_{NLB} and F_B, the probability of $I_C = j$ $(j = 1, 2, 3, 4, 5)$ is equal to $P(I_C = j)$. Here, the predicted value of I_C is determined according to the maximum probability. For example, if $P(I_C = 3)$ is the maximum probability under the certain T_{NLB} and F_B, the value of I_C is 3.

To obtain the logit coefficients α_1 and α_2 and threshold β_j $(j = 1, 2, 3, 4)$, the proposed OLR model is fitted using the SPSS (V.18.0) PLUM [51]. Table I illustrates the SPSS outputs for the OLR model. It includes the estimated coefficients for each variable and their standard errors, along with the Wald statistics and associated p-values (Sig.). According to the Wald statistics and p-values, the variables have a significant contribution to the prediction of the outcome, and their coefficients are different from zero. The effect of F_B and T_{NLB} are significant and positive, indicating that the larger values of F_B and T_{NLB} are more likely to achieve higher values of user interest. The pseudo R^2 $(\text{Nagelkerke} = 0.352)$ of the model is calculated and the score test for the proportional odds assumption is satisfied $(\chi^2 = 90.63, \mathrm{p} = 0.117)$, which indicates that the logit coefficients are consistent for all thresholds. Therefore, the OLR model of user interest (IC model) can be determined, where α_1 and α_2 are 7.682 and 3.434, respectively. The values of β_j $(j = 1, 2, 3, 4)$ are 0.678, 2.697, 4.349, 6.450, respectively.

TABLE I
PARAMETER ESTIMATES FOR USER INTEREST

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	$[I_C = 1]$	0.678	0.215	9.922	1	0.002	0.256	1.100
	$[I_C = 2]$	2.697	0.214	158.309	1	0.000	2.277	3.117
	$[I_C = 3]$	4.349	0.237	337.120	1	0.000	3.885	4.814
	$[I_C = 4]$	6.450	0.276	546.158	1	0.000	5.909	6.992
Location	T_{NLB}	7.682	0.391	385.835	1	0.000	6.916	8.449
	F_B	3.434	0.399	74.191	1	0.000	2.653	4.215

TABLE II
PARAMETER ESTIMATES FOR QOE MODEL

							95% Confidence Interval	
		Estimate	Std. Error	Wald	df	Sig.	Lower Bound	Upper Bound
Threshold	$[QoE = 1]$	2.427	0.271	79.892	1	0.000	1.894	2.959
	$[QoE = 2]$	4.612	0.296	242.342	1	0.000	4.031	5.192
	$[QoE = 3]$	6.764	0.339	398.468	1	0.000	6.100	7.429
	$[QoE = 4]$	8.992	0.376	572.257	1	0.000	8.255	9.729
Location	Q_{AV}	0.835	0.071	137.429	1	0.000	0.696	0.975
	I_C	1.028	0.056	332.490	1	0.000	0.918	1.139

C. QoE Evaluation Model

Here, we further analyze the relationship among the audiovisual quality, user interest and QoE, and finally establish an objective QoE evaluation model (QAVIC model) combining all aspects. The OLR model is employed to illustrate the relationship among them. It can be expressed as follows:

$$\begin{aligned}\mathrm{logit}(P(QoE > j)) &= \ln \frac{P(QoE > j)}{1 - P(QoE > j)}\\ &= \lambda_1 Q_{\mathrm{AV}} + \lambda_2 I_C - \mu_j \ (j = 1, 2, 3, 4)\\ &\Rightarrow P(QoE > j)\\ &= \frac{\exp(\lambda_1 Q_{\mathrm{AV}} + \lambda_2 I_C - \mu_j)}{1 + \exp(\lambda_1 Q_{\mathrm{AV}} + \lambda_2 I_C - \mu_j)} \qquad (7)\end{aligned}$$

where the *QoE* values are the dependent variables, and Q_{AV} and I_C are the explanatory variables. All of them are subjective data in training set. λ_1 and λ_2 are the logit coefficients. μ_j $(j = 1, 2, 3, 4)$ is the threshold for QoE. It denotes the point (in terms of a logit) where the user's QoE is predicted into the higher rating level. The value of *QoE* is also determined according to the maximum probability. For example, if the value of $P(QoE = 3)$ is the maximum probability under the certain Q_{AV} and I_C, the value of *QoE* is 3.

Table II illustrates the SPSS outputs for the OLR model. It can be found that both Q_{AV} and I_C provide a significant and positive effect to QoE. The larger values of Q_{AV} and I_C

Fig. 16. Illustration of recorded video sequence for validation.

are more likely to achieve higher values of the user's QoE. Moreover, the pseudo R^2 $(\text{Nagelkerke} = 0.732)$ of the model is calculated and the test of parallel lines is also carried out by SPSS. The score test for the proportional odds assumption is satisfied $(\chi^2 = 134.08, \mathrm{p} = 0.073)$. Accordingly, the OLR model of QoE (QAVIC model) can be determined, where λ_1 and λ_2 are 0.835 and 1.028, respectively. The values of $\mu_j (j = 1, 2, 3, 4)$ are 2.427, 4.612, 6.764, 8.992, respectively.

Above all, the QAVIC model is established considering both the audiovisual quality and user interest for the first time, which provides a convenient way to better evaluate user perceptions. Because the user's common behavior characteristics are analyzed and employed, the QAVIC model is suitable for different users.

V. Experimental Results

In this section, the accuracy of the blink and eye movement detection will be firstly studied. The performance of our proposed QoE assessment model will be then verified.

A. Accuracy of Blink and Eye Movement Detection

A total of 12 randomly recorded videos from different users were employed to validate the accuracy of the blink and eye movement detection, as illustrated in Fig. 16. The blink detector was evaluated by comparing its detected intervals with the actual blink intervals. The actual blink interval was measured by manually counting blinks in each subject's video. To verify the performance of the saccade detection, we checked each saccade by comparing its detection interval with actual saccade interval as well, where the actual saccade was collected by manual tracking using a computer mouse of the eyeball positions in a recorded video sequence.

Table III lists the performance of the proposed blink and eye movement detection method compared with other common methods using single video camera [33], [45] and the commercial Eye Tribe Tracker (EET).[3] It can be found that the proposed method obtains a superior performance than other methods using single video camera, and the accuracy of the blink and eye movement detection can achieve to 96.6% and 94.5%, respectively. As to the EET method, it can accurately detect the movement of the pupil with sub-millimeter precision.

TABLE III
Performance Comparison of Eye Movement and Blink Detection

	Index	Error detection	Miss detection	Total Number	Accuracy
Takahashi [33]	blink	59	6	654	91.2%
	Eye movement	829	538	13021	89.5%
Peng [45]	blink	54	4	654	90.0%
	Eye movement	1117	1305	13021	81.4%
ETT[3]	blink	14	2	654	97.6%
	Eye movement	245	172	13021	96.8%
Proposed method	blink	18	4	654	96.6%
	Eye movement	497	139	13021	94.5%

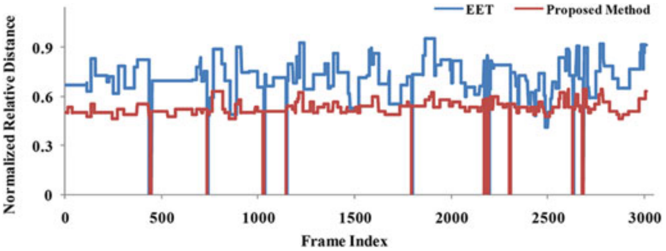


Fig. 17. Test results of eye movement detection.

Fig. 17 illustrates the detection results of the eye movement for both the proposed method and the ETT method. The vertical axis is the normalized relative distance. The changes of this value indicate the different viewing behaviors. For example, the fluctuations indicate the eye movements, while the sharp declines denote the blink. It is obvious that these two methods have a good consistency in the detection results.

B. Performance of the Proposed QoE Evaluation Model

To verify the performance of the proposed QoE assessment model, a different set of 60 audiovisual clips with a resolution of 672 × 378 pixels (16:9) were employed in our experiments, including a variety of contents. The duration of each clip was 5 to 10 minutes. All clips were divided into four different validation groups (VG1, VG2, VG3 and VG4). The videos in different groups had different video quality levels. They were coded under different QPs of 37, 32, 27, and 24 respectively, using the software of FFmpeg 0.4.9 with x.264 library. Each video sequence was encapsulated into the FLV format with the corresponding audio sequence. 48 users participated in the validation test, including 22 females and 26 males. Among these users, 24 users watched two groups of clips, and the others watched one group according to the amount of their spare time. The procedures of the subjective test were the same with those in the training set. There were 1035 rating simples of user interest, audiovisual quality and QoE, respectively.

Fig. 18 illustrates the subjective QoE (SQoE) and objective QoE (OQoE) obtained by the proposed QAVIC model for individual users in different groups. It can be found that the scores of SQoE and OQoE obtained by the QAVIC model are various for

[3] [Online]. Available: http://theeyetribe.com/

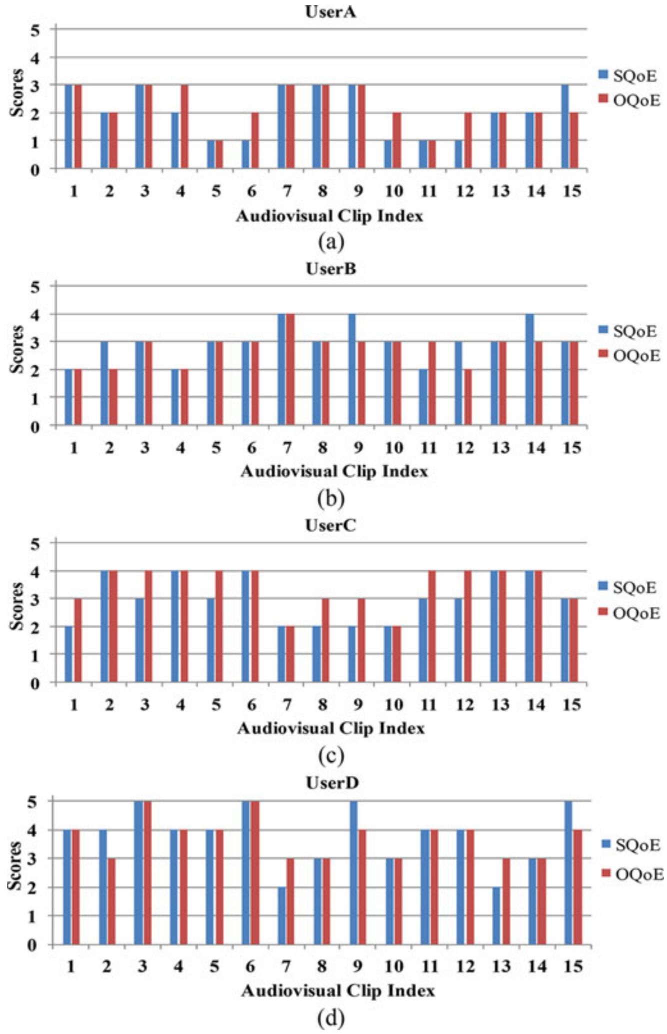


Fig. 18. Illustration of the performance of QAVIC model for individual users in each group. (a) VG1. (b) VG2. (c) VG3. (d) VG4.

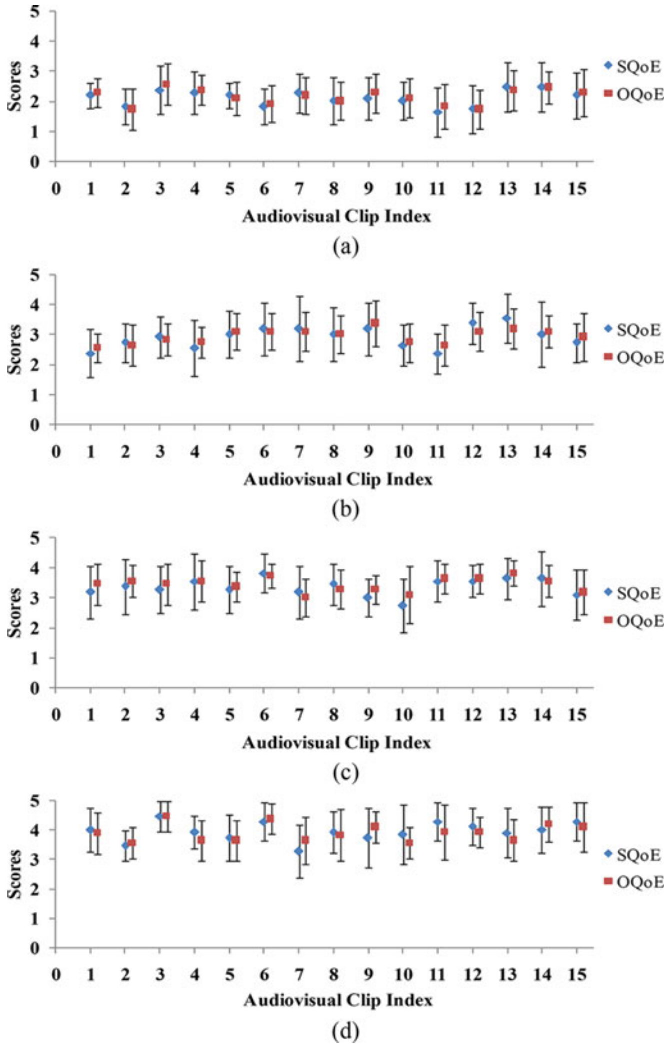


Fig. 19. Performances of the QAVIC model for all users in each group. (a) VG1. (b) VG2. (c) VG3. (d) VG4.

different audiovisual clips, but the scores of SQoE and OQoE are quite similar for a certain clip. Although different users may have different perceptions, the values of QoE can still be estimated through their responses to the services. Fig. 19 presents the average values and standard deviations of SQoE and OQoE of all users for each audiovisual clip. It is obvious that the average value of SQoE and OQoE are quite similar in all validation groups and the standard deviations of SQoE and OQoE for most audiovisual clips are in good agreement with each other.

To better verify the accuracy of the proposed model for the QoE evaluation, the confusion matrix of the QoE values predicted by the QAVIC model with the actual QoE values is provided in Table IV. This confusion matrix shows that the model predicts the largest percentage of correct outcome categories (73.4%) for 3 of the 5 categories, while the lowest percentage of correct outcome categories (53.2%) for 1 of the 5 categories. The average percentage of correct outcome category is 64.5%. The majority of cases in all categories (98.8%) were predicted to fall in either the correct outcome category or the adjacent category (i.e., ±1 category). Based on this observation, the model has demonstrated a satisfactory performance which vindicates the necessity of QoE evaluation taking account of multiple dimensions of human perceptions.

TABLE IV
CONFUSION MATRIX OF THE PREDICTED SCORES FROM THE QAVIC MODEL BY THE SUBJECTIVE SCORES

	OQoE, no.(%)					
SQoE	1	2	3	4	5	Total
1	41 (53.2)	36 (46.8)	–	–	–	77 (100)
2	11 (4.3)	154 (60.2)	91 (35.5)	–	–	256 (100)
3	–	44 (13.5)	240 (73.4)	43 (13.1)	–	327 (100)
4	–	12 (4.2)	89 (31.0)	176 (61.3)	10 (3.5)	287 (100)
5	–	–	–	31 (35.2)	57 (64.7)	88 (100)

VI. Conclusion

The user's perception and satisfaction are crucial determinants for the success of a particular service, knowing what is the user thinking about is the highest realm of the services in real life. However, the user's perception is the internal state and difficult to measure, which makes the progress on the objective QoE evaluation still limited and slow. Targeting at solving this problem, we have made a detailed analysis on the influence of technology, content, and user domains to QoE, and proposed an objective QoE evaluation model (QAVIC model) for the first time with a combination of the influences of both the perceptions of audiovisual quality and user interest in content. More specifically, the user interest is expressed using the common features of viewing behaviors (e.g., blinks). Experimental results have shown that the QAVIC model can well estimate the user's QoE.

It should be noted that there may be an impact on the correlation between audiovisual quality and QoE when the audiovisual quality assessment test and QoE assessment test were carried out by the same people under different conditions. Our future work will find out the relationship between the values of AV quality when the tests are performed by the same and different people, respectively. This will benefit the proposed QoE model. Moreover, we will consider the user perceptions in other dimensions and take more users' general explicit responses in the behavior and physiology into the QoE evaluation, such as physiological responses, expression behaviors, and commercial behaviors and so on.

References

[1] D. Geerts *et al.*, "Linking an integrated framework with appropriate methods for measuring QoE," in *Proc. IEEE Int. Workshop Quality Multimedia Experience*, Jun. 2010, pp. 158–163.

[2] K. Kilkki, "Quality of experience in communications ecosystem," *J. Universal Comput. Sci.*, vol. 14, no. 5, pp. 615–624, 2008.

[3] K. U. R. Laghari, N. Crespi, B. Molina, and C. E. Palau, "QoE aware service delivery in distributed environment," in *Proc. IEEE Workshop Int. Conf. Adv. Inform. Netw. Appl.*, Mar. 2011, pp. 837–842.

[4] H. J. Kim *et al.*, "The QoE evaluation method through the QoS-QoE correlation model," in *Proc. IEEE Int. Conf. Netw. Comput. Adv. Inform. Manage.*, Sep. 2008, pp. 719–725.

[5] J. You, U. Reiter, M. M. Hannuksela, and M. Gabbouj, "Perceptual-based quality assessment for audio-visual services: A survey," *Signal Process.: Image Commun.*, vol. 25, no. 7, pp. 482–501, 2010.

[6] N. Staelens, S. Moens, W. Van den Broeck, I. Mariën, and B. Vermeulen, "Assessing quality of experience of IPTV and video on demand services in real-life environments," *IEEE Trans. Broadcast.*, vol. 56, no. 4, pp. 458–466, Dec. 2010.

[7] *Definition of Quality of Experience (QoE)*, Rec. TD 109rev2 (PLEN/12) ITU-T, Geneva, Switzerland, 2007.

[8] K. Brunnström *et al.*, *"Qualinet white paper on definitions of quality of experience,"* Qualinet. (2013). [Online]. Available: http://www.qualinet.eu/images/stories/QoE_whitepaper_v1.2.pdf

[9] K. U. R. Laghari, N. Crespi, and K. Connelly, "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.

[10] P. Brooks and B. Hestnes, "User measures of quality of experience: Why being objective and quantitative is important," *IEEE Netw.*, vol. 24, no. 2, pp. 8–13, Mar.–Apr. 2010.

[11] R. Jain, "Quality of experience," *IEEE Multimedia Mag.*, vol. 11, no. 1, pp. 95–96, Jan.–Mar. 2004.

[12] M. Claypool, P. Le, M. Wased, and D. Brown, "Implicit interest indicators," in *Proc. ACM Int. Conf. Intell. User Interfaces*, 2001, pp. 33–40.

[13] M. Fiedler, T. Hossfeld, and P. Tran-Gia, "A generic quantitative relationship between quality of experience and quality of service," *IEEE Netw.*, vol. 24, no. 2, pp. 36–41, Mar.–Apr. 2010.

[14] H. J. Kim and S. G. Choi, "A study on a QoS/QoE correlation model for QoE evaluation on IPTV service," in *Proc. IEEE Int. Conf. Adv. Commun. Technol.*, Feb. 2010, pp. 1377–1382.

[15] M. Ries, P. Froehlich, and R. Schatz, "QoE evaluation of high-definition IPTV services," in *Proc. IEEE Int. Conf. Radio Elektronika.*, Apr. 2011, pp. 1–5.

[16] T. De Pessemier, K. De Moor, W. Joseph, L. De Marez, and L. Martens, "Quantifying the influence of rebuffering interruptions on the user's quality of experience during mobile video watching," *IEEE Trans. Broadcast.*, vol. 59, no. 1, pp. 47–61, Mar. 2013.

[17] M. Hassenzahl, "The interplay of beauty, goodness, and usability in interactive products," *Human-Comput. Interaction*, vol. 19, no. 4, pp. 319–349, 2004.

[18] *Opinion Model for Video-Telephony Applications*, Rec. G.1070 ITU-T, Geneva, Switzerland, Apr. 2007.

[19] *Parametric Non-Intrusive Assessment of Audiovisual Media Streaming Quality*, Rec. P.1201 ITU-T, Geneva, Switzerland, 2012.

[20] S. Winkler and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," *IEEE. Trans. Multimedia*, vol. 8, no. 5, pp. 973–980, Oct. 2006.

[21] D. H. Hands, "A basic multimedia quality model," *IEEE. Trans. Multimedia*, vol. 6, no. 6, pp. 806–816, Dec. 2004.

[22] N. Kitawaki, Y. Arayama, and T. Yamada, "Multimedia opinion model based on media interaction of audio-visual communications," in *Proc. Int. Conf. MSAQN*, 2005, pp. 5–10.

[23] M. H. Pinson, W. Ingram, and A. Webster, "Audiovisual quality components," *IEEE Signal Process. Mag.*, vol. 2, no. 6, pp. 60–67, Nov. 2011.

[24] R. L. Van Eijk, A. Kohlrausch, and J. F. Juola, "Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type," *Percept. Psychophys.*, vol. 70, no. 6, pp. 955–968, 2008.

[25] Z. Chen and C. Guillemot, "Perceptually-friendly H. 264/AVC video coding based on foveated just-noticeable-distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 6, pp. 806–819, Jun. 2010.

[26] O. Le Meur, P. Le Callet, and D. Barba, "A coherent computational approach to model the bottom-up visual attention," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 28, no. 5, pp. 802–817, May 2006.

[27] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[28] A. P. Bradley and F. W. M. Stentiford, "Visual attention for region of interest coding in JPEG 2000," *J. Vis. Commun. Image Represent.*, vol. 14, pp. 232–250, Sep. 2003.

[29] K. De Moor, F. Mazza, I. Hupont, and M. R. Quintero, "Chamber QoE: A multi-instrumental approach to explore affective aspects in relation to quality of experience," in *Proc. SPIE-IS&T Electron. Imaging*, 2014, pp. 2458–2462.

[30] J. Lassalle, L. Gros, T. Morineau, and G. Coppin, "Impact of the content on subjective evaluation of audiovisual quality: What dimensions influence our perception?," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, Jun. 2012, pp. 1–6.

[31] S. N. Jeffrey, *Pyschology Concepts and Application*, 3rd ed., Boston, MA, USA : Houghton Mifflin, 2009.

[32] M. Argyle, *Bodily Communication*, London, U.K.: Routledge, 1988.

[33] M. Takahashi *et al.*, "An estimator for rating video contents on the basis of a viewer's behavior in typical home environments," in *Proc. IEEE Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2013, pp. 6–13.

[34] J. Song, F. Yang, and S. Wan, "QoE evaluation of video services considering users' behavior," in *Proc. IEEE Int. Conf. Multimedia Expo Workshop*, Jul. 2014, pp. 1–6.

[35] S. Möller and A. Raake, *Quality of Experience*, New York, NY, USA: Springer, 2014.

[36] H. Dibeklioğlu, M. O. Hortas, I. Kosunen, and P. Zuźnek, "Design and implementation of an affect-responsive interactive photo frame," *J. Multimodal User Interfaces*, vol. 4, no. 2, pp. 81–95, 2011.

[37] A. Frischen, A. P. Bayliss, and S. P. Tipper, "Gaze cueing of attention: Visual attention, social cognition, and individual differences," *Psychological Bulletin*, vol. 133, no. 4, pp. 694–724, 2007.

[38] P. Ballon, J. Pierson, and S. Delaere, "Test and experimentation platforms for broadband innovation: Examining European practice," in *Proc. Conf. Eur. Regional Conf. ITS*, 2005, pp. 4–6.

[39] J. Salminen, S. K. Laakso, M. Pallot, and B. Trousse, "Evaluating user involvement within living labs through the use of a domain landscape," in *Proc. IEEE Int. Conf. Concurrent Enterprising*, Jun. 2011, pp. 1–10.

[40] *Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in Any Environment*, Rec. P. 913 ITU-T, Geneva, Switzerland, 2014.

[41] *Coding of Audio-Visual Objects-Part 3: Audio (MPEG-4 Audio, 2nd Edition)*, Int. Std. 14496-3:2001 International Organization for Standardization/International Electrotechnical Commission, Geneva, Switzerland, 2001.

[42] Q. Huynh-Thu *et al.*, "Study of rating scales for subjective quality assessment of high-definition video," *IEEE Trans. Broadcast.*, vol. 57, no. 1, pp. 1–14, Mar. 2011.

[43] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, Rec. BT.500-13ITU-T, Geneva, Switzerland, 2012.

[44] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.

[45] W. T. Peng *et al.*, "Editing by viewing: Automatic home video summarization by viewing behavior analysis," *IEEE. Trans. Multimedia*, vol. 13, no. 3, pp. 539–550, Jun. 2011.

[46] J. Tao and K. Jiyong, "A 3-D point sets registration method in reverse engineering," *Comput. Indust. Eng.*, vol. 53, no. 2, pp. 270–276, 2007.

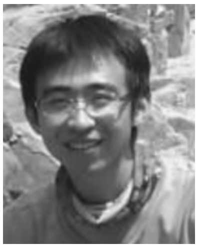
[47] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Con. Comput. Vis.*, Sep. 1999, vol. 2, pp. 1150–1157.

[48] Z. H. Zhou and X. Geng, "Projection functions for eye detection," *Pattern Recog.*, vol. 37, no. 5, pp. 1049–1056, 2004.

[49] T. Nakano, Y. Yamamoto, and K. Kitajo, "Synchronization of spontaneous eyeblinks while viewing video stories," in *Proc. R. Soc. B: Biol. Sci.*, 2009, pp. 3635–3644.

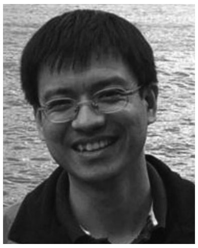
[50] M. Scott, *Applied Logistic Regression Analysis*, Thousand Oaks, CA, USA: Sage, 2002.

[51] R. Carver and J. Nash, *Doing Data Analysis With SPSS: Version 18.0*, Boston, MA, USA: Cengage Learning, 2011.



Jiarun Song received the B.S. degree in telecommunication engineering and the Ph.D. degree in communication and information system from Xidian University, Xi'an, China, in 2009 and 2015, respectively.

He is currently a Post-Doctorate with the State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include QoE, video quality assessment, and multimedia communication.



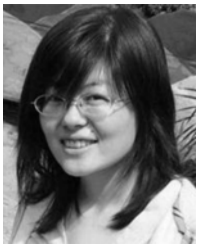
Fuzheng Yang (M'10) received the B.E. degree in telecommunication engineering, and the M. E. and the Ph.D. degrees in communication and information system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

He was a Lecturer and an Associate Professor with Xidian University in 2005 and 2006, respectively. From 2006 to 2007, he was a Visiting Scholar and Postdoctoral Researcher with the Department of Electronic Engineering, Queen Mary, University of London, London, U.K. He has been a Professor of Communications Engineering with Xidian University since 2012. His research interests include video quality assessment, video coding, and multimedia communication.

Yicong Zhou (S'08–M'10–SM'14) received the B.S. degree from Hunan University, Changsha, China, in 1992, and the M. S. and Ph.D. degrees from Tufts University, Medford, MA, USA, in 2008 and 2010, all in electrical engineering.

He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.

Dr. Zhou was a recipient of the third place prize of the Macau Natural Science Award in 2014.



Shuai Wan (M'08) received the B.E. degree in telecommunication engineering and the M.E. degree in communication and information system from Xidian University, Xi'an, China, in 2001 and 2004, respectively, and received the Ph.D. degree in electronic engineering from Queen Mary University of London, London, U.K., in 2007.

She is currently a Professor with Northwestern Polytechnical University, Xi'an, China. Her research interests include scalable/multiview video coding, video quality assessment, and hyperspectral image compression.

Hong Ren Wu received the B.Eng. and M.Eng. degrees from the University of Science and Technology Beijing (formerly Beijing University of Iron and Steel Technology), Beijing, China, in 1982 and 1985 respectively, and the Ph.D. degree in electrical and computer engineering from The University of Wollongong, Wollongong, NSW, Australia, in 1990.

From 1982 to 1985, he was an Assistant Lecturer with the Department of Industrial Automation, University of Science and Technology Beijing. He joined the Department of Robotics and Digital Technology, Chisholm Institute of Technology, Dandenong, VIC, Australian, as a Lecturer, and then became Faculty of Information Technology, Monash University, Clayton, VIC, Australia, in 1990, where he served as a Lecturer (1990 to 1992), Senior Lecturer (1992 to 1996), and Associate Professor of Digital Systems (1997 to 2005). He has been a Professor of Visual Communications Engineering with the Royal Melbourne Institute of Technology (RMIT University), Melbourne, VIC, Australia, since 2005, and concurrently served as the Head of Computer and Network Engineering from February 2005 to January 2010. He has authored or coauthored papers that have appeared in refereed journals, and is co-editor of the book *Digital Video Image Quality and Perceptual Coding* (Taylor and Francis, 2006). His research interests include the fields of signal processing, video and image processing and enhancement, perceptual coding of natural and medical images, digital video coding, compression and transmission, and digital picture quality assessment.

Prof. Wu was a Guest Editor for the Special Issue on Multimedia Communication Services of *Circuits, Systems and Signal Processing*, the Special Issue on Quality Issues on Mobile Multimedia Broadcasting of the IEEE TRANSACTIONS ON BROADCASTING, and the Special Issue on QoE Management in Emerging Multimedia Services of the *IEEE Communications Magazine*.