

Recent Activities of QoE-related Standardization in ITU-T SG12

Yoichi Matsuo, Kazuhisa Yamagishi, and Masanori Koike

Abstract

This article introduces recent standardization activities related to the evaluation of the quality of experience (QoE) of speech, audiovisual, and other new services such as extended reality and chatbots, focusing on the activities of the Study Group 12 of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T SG12), which is responsible for standardization work on performance, quality of service, and QoE.

Keywords: quality of experience, adaptive bitrate streaming, extended reality, chatbots

1. ITU-T Study Group 12

The Study Group 12 of the International Telecommunication Union - Telecommunication Standardization Sector (ITU-T SG12) is a lead SG on network performance, quality of service (QoS), and quality of experience (QoE) in the worldwide standardization of speech and video quality evaluation that takes into account achievements in regional standardization bodies such as ETSI (European Telecommunications Standards Institute) and ATIS (Alliance for Telecommunications Industry Solutions). Since standardization work is carried out in various standardization organizations regarding network performance parameters, all these organizations have confirmed that their work matches that of SG12.

2. Full-band and super-wideband E-model (G.107.2)

Recommendation G.107, called the E-model, has been standardized as a quality-planning tool for telephony services and is used worldwide. For instance, in Japan, JJ201.11, which specifies the quality of Internet Protocol (IP) telephony services, uses R values calculated on the basis of the E-model. In Question 15/12 (Parametric and E-model-based planning, prediction and monitoring of conversational speech quality), the extension of the E-model was studied for

evaluating full-band (20 to 20,000 Hz) speech communication services, and the basic algorithm was standardized as Recommendation G.107.2 in 2019. As the next step, the revision of G.107.2 was studied for enabling the full-band E-model to be used under various conditions. Therefore, G.107.2 was revised by updating the calculation algorithm of the effective equipment impairment factor (l_e , eff), delay impairment factor (l_d), basic signal-to-noise ratio (R_o), and simultaneous impairment factor (l_s) so that it can handle background noise, burst packet loss, and delay.

3. Quality-estimation model and degradation-analysis procedure for adaptive bitrate streaming (P.1203, P.1204, and P.DiAQoSE)

For monitoring adaptive-bitrate-streaming quality, Recommendation P.1203, which specifies quality-monitoring techniques for high-definition resolution coded in H.264/AVC (Advanced Video Coding), and Recommendations P.1204.3, P.1204.4, and P.1204.5, which specify quality-monitoring techniques for 4K video and H.265/HEVC (High Efficiency Video Coding), have been standardized. Adaptive bitrate streaming using AV1 codec (AOMedia Video 1) has been increasing. Thus, the extension of P.1203 and P.1204 to support the new codec is being studied.

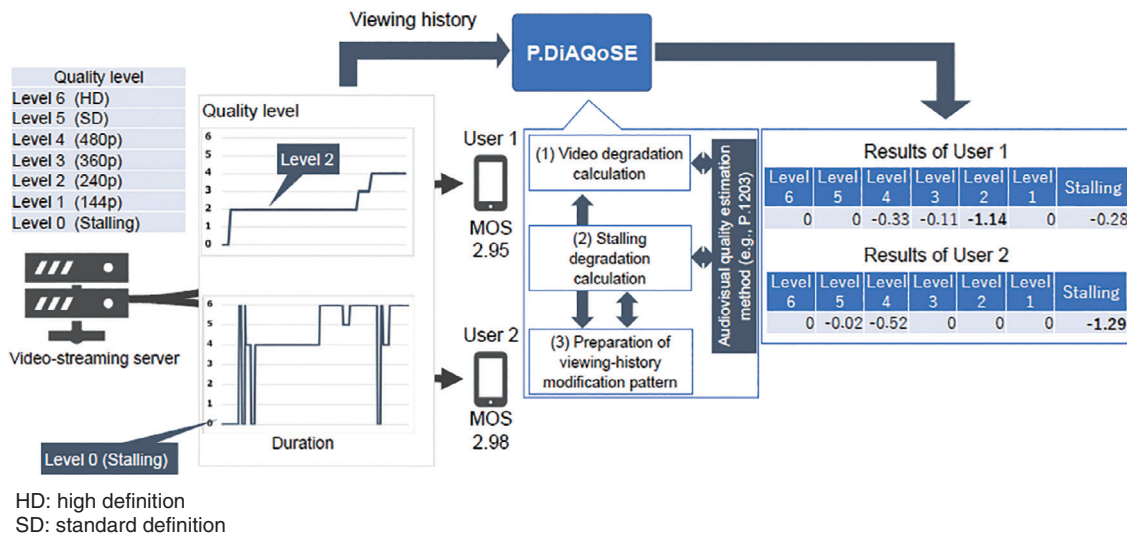


Fig. 1. Overview of P.DiAQoSE.

A procedure for interpreting the quality-degradation factors is also being studied for monitoring the adaptive-bitrate-streaming services (Fig. 1). A work item, P.DiAQoSE, calculates the degree to which quality parameters (i.e., bitrate, resolution, frame rate, and stalling information), which are inputs to audiovisual quality-estimation models such as P.1203, P.1204.3, P.1204.4, and P.1204.5, reduces the audiovisual quality estimated using audiovisual quality-estimation models. Specifically, the amount of degradation due to the quality parameters for a given session is calculated by distributing the difference between the maximum estimated audiovisual quality, which is calculated by selecting the highest quality parameters, and the current estimated audiovisual quality to each quality parameter on the basis of Shapley theory. In the example in Fig. 1, the mean opinion score (MOS) estimated by P.1203 for the video viewing of Users 1 and 2 are 2.95 and 2.98, respectively. However, when the degradation amount for each quality parameter is calculated using P. DiAQoSE, the MoS of User 1 is more affected by video-quality level 2, while the MOS of User 2 is significantly reduced by stalling. Thus, in addition to audiovisual quality-estimation models, the amount of degradation for each quality parameter makes it easier to take action to improve service quality.

4. Object-recognition-rate-estimation model in surveillance video of autonomous driving (P.obj-recog)

SAE International (formerly, Society of Automotive Engineers) defines six levels of automated driving, from Levels 0 to 5, depending on the driver and area where the vehicle can be driven [1]. For Level 2, a remote monitoring center detects objects on the basis of video from an in-vehicle camera and assists driving in emergency situations. In this service, an observer in a remote monitoring center checks for objects on the road on the basis of the video from the in-vehicle camera that is encoded and transmitted to the remote monitoring center. Therefore, the video quality transmitted from the in-vehicle camera should be clear enough for observers to recognize objects. To confirm that the video is always transmitted in sufficient quality for object recognition, P.obj-recog was launched as a work item to study a technique to derive the probability that an observer can recognize an object from the encoded video, as shown in Fig. 2. This will enable quality monitoring of video transmitted from in-vehicle cameras.

5. QoE factors for augmented reality services and objective quality-evaluation method for extended reality services (G.1036, PSTR-OQM XR)

Recommendation G.1036, which specifies QoE

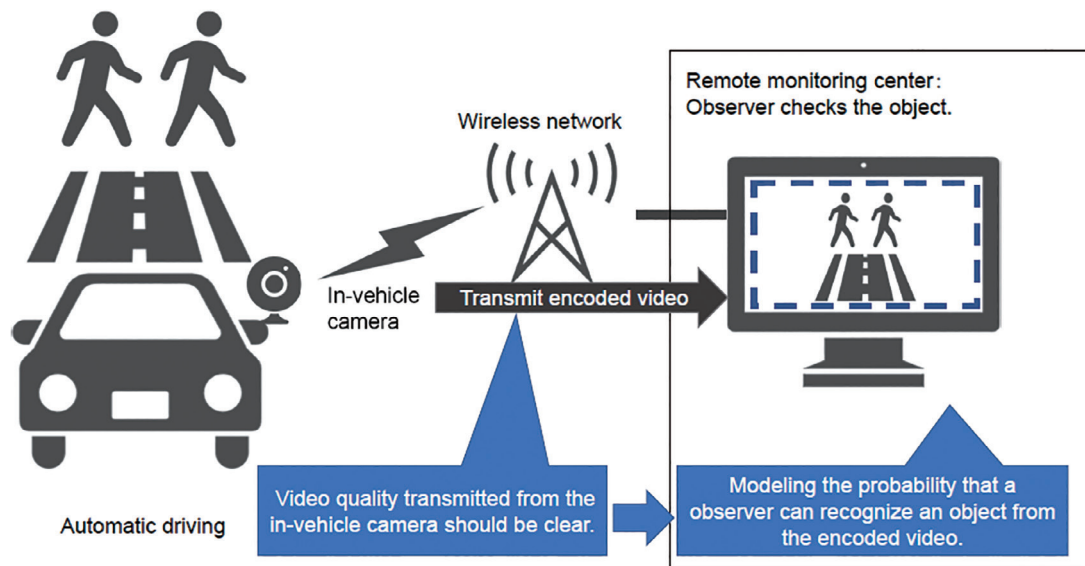


Fig. 2. Estimation method for object-recognition ratio in P.obj-recog.

factors for augmented reality (AR) services, was standardized to enable the evaluation of the quality of AR services. This Recommendation specifies QoE factors of two AR services that add objects around the user's face and objects to the real space. In addition to quality-influencing factors used in conventional audio and video quality factors such as bitrate, resolution, frame rate, transmission delay, and device size, the factors related to the degree of integration between virtual objects and objects in the real space, such as detection performance of user faces and objects in the real space, and response performance of AR services are described. The quality related to the interaction between the user and AR service is also specified.

In addition, PSTR-OQMXR has been launched as a work item for a technical report on objective quality modelling for extended reality (XR) services. The purpose of this technical report is to identify the current status and issues of current objective quality-assessment methods for XR services as well as issues that need to be addressed to construct an estimation model.

6. Guidance for the development of machine-learning-based solutions for QoS/QoE prediction and network-performance management in telecommunication scenarios (P. 1402)

Recommendation P.1402, which specifies guide-

lines for applying machine-learning methods to the prediction of QoS/QoE, has been standardized, enabling machine-learning methods to be applied to studied work items in SG12. This Recommendation describes the basics of using machine learning, including how to construct training and evaluation data, categorization of machine-learning methods, and how to avoid over-fitting. Specific use cases, which are input/output examples and methods when machine learning is used for voice and video QoS/QoE prediction, are also described such as in the P.565 series.

7. Subjective quality evaluation of text-based chatbots (P.852)

Recommendation P.852, which specifies a subjective-quality-evaluation method for text-based chatbots, has been standardized. This Recommendation describes the quality factors for chatbots, which are overall impression, system information, system behavior, and user impression of the system. It also describes the subjective experiments for chatbots and the evaluation procedure.

8. Outlook

This article described subjective assessment and quality-estimation models for speech, video streaming, XR, and chatbots. SG12 has recently studied the

extension of recommendations, such as support for a new codec for adaptive-bitrate-streaming services and revision of the full-band E-model. A study on the quality of transmitted monitoring video for automated driving services has also been launched.

Since various services are expected to be launched along with the deployment of the fifth-generation mobile communication system (5G)/6G, the design

and management of QoS/QoE for various services should be considered. Therefore, it will be important to investigate the activities of SG12.

Reference

- [1] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-road Motor Vehicles," J3016_201806, pp. 1–5, 2018.



Yoichi Matsuo

Research Engineer, NTT Network Technology Laboratories.

He received an M.E. and Ph.D. in applied mathematics from Keio University, Kanagawa, in 2012 and 2015. Since joining NTT in 2015, he has been engaged in research on network management. He has been contributing to ITU-T SG12 since 2019.



Masanori Koike

Researcher, NTT Network Technology Laboratories.

He received a B.E. in mathematical engineering and information physics in 2015 and M.S. in information science and technology in 2017 from the University of Tokyo. Since joining NTT in 2017, he has been engaged in research related to video quality. He has been contributing to ITU-T SG12 since 2023.



Kazuhisa Yamagishi

Senior Research Engineer, NTT Network Technology Laboratories.

He received a B.E. in electrical engineering from Tokyo University of Science in 2001 and M.E. and Ph.D. in electronics, information, and communication engineering from Waseda University, Tokyo, in 2003 and 2013. In 2003, he joined NTT, where he has been engaged in the development of objective quality-estimation models for multimedia telecommunications. He has been contributing to ITU-T SG12 since 2006. He has been a rapporteur of Question 13/12 since 2017, a vice-chair of Working Party 3 in SG12 since 2021, and is a vice-chair of SG12 for the 2022–2024 study period.