Light-weight audiovisual quality assessment of mobile video: ITU-T Rec. P.1201.1

Kazuhisa Yamagishi #, Shan Gao *

NTT Network Technology Laboratories, NTT Corporation
Tokyo, 180-8585 Japan
yamagishi.kazuhisa@lab.ntt.co.jp

* Media Laboratory, Huawei Technologies Co., Ltd.
Shenzhen, P.R.China
simon.gaoshan@huawei.com

***Abstract*—Significant progress has been made in the recent years in the development of technologies for encoders, decoders and networks. As a result, content, network, and Internet service providers can deliver video content over IP networks. To provide a high-quality service, the quality of experience (QoE) is becoming much more important. Since QoE is affected by such factors as the audiovisual content, encoding and decoding techniques, and network performance, service providers should monitor the QoE of a communication service in real time to confirm its status. To do this, a quality monitoring tool is necessary. The International Telecommunication Union - Telecommunication Standardization Sector Study Group 12 (ITU-T SG12) has studied the parametric non-intrusive assessment of audiovisual media streaming quality (P.NAMS). The P.NAMS – lower resolution application area was finally standardized as ITU-T Recommendation P.1201.1 in October 2012. The P.1201.1 model can be used for estimating audio, video, and audiovisual quality for mobile audiovisual media streaming using packet headers. Since the model analyzes only packet header information, the computational power of the model is very light, and the model can be applied to encrypted packets. This paper describes the P.1201.1 model and its performance.**

I. Introduction

Internet protocol television (IPTV) is now widely used. To provide high-quality IPTV, quality management is becoming much more important. The quality of IPTV is generally affected by a processing chain composed of audio and video compression, a transmission scheme (i.e., user datagram protocol (UDP) or transmission control protocol (TCP)) and transmission behavior (i.e., packet loss or delay), and client behavior (i.e., packet loss concealment (PLC) or buffering). To manage the quality of IPTV, it is necessary for service providers to monitor audio, video, and audiovisual quality in real time. This requires a real-time quality monitoring tool.

Monitoring the QoE at the head end is important because quality degradation influences the QoE of all users. In these cases, monitoring technique needs to detect even minor degradations, so full reference (FR) media-layer models that take uncompressed source and encoded media signals as input are suitable for monitoring quality. In contrast, it is difficult to use media signals to monitor the QoE at the network or end-user terminal because the analysis needs to be implemented in terminals such as mobile terminals, home gateways, and set-top boxes (STBs). In addition, uncompressed source signals cannot be used for monitoring the QoE. From these reasons, it is preferable to analyze packets using a method with low computational load at the network or end-user terminal. In these cases, no-reference (NR) packet-layer models are suitable for estimating QoE from IP packet header information.

There have been studies on packet-layer models. Several models that use the bit rate for estimating video quality affected by compression have been proposed [1], [2], [3]. In general, Yamagishi et al. [4] argued that video quality affected by compression depends on video content, which indicates that compression affecting video quality cannot be taken into account using only the bit rate. That is, the model requires more information to take into account the impact of video content on video quality. Several models for estimating video quality by using the packet-loss ratio have been proposed [2], [3]. Consecutive IP packets are often lost by a network. In such a case, the video quality degraded by packet loss cannot be estimated based only on random packet-loss characteristics [5], [6], [7]. According to Masuda et al. [8], the video quality affected by packet loss depends on the positions of lost video-frame types (i.e., I-, P-, and B-frames). The number of damaged video frames depends on the video-frame type that has the lost packet and on the group of picture (GoP) structure. That is, it is essential to take into account the number of damaged video frames to estimate video quality degraded by packet loss. From these investigations, the International Telecommunication Union – Telecommunication Standardization Sector Study Group 12 (ITU-T SG12) has studied the parametric non-intrusive assessment of audiovisual media streaming quality (provisional code: P.NAMS) [9] – lower resolution (LR: i.e., quarter common intermediate format (QCIF, 176 × 144 pixels), quarter video graphics array (QVGA, 320 × 240 pixels), or half VGA (HVGA, 320 × 480 pixels)) [10], [11] and higher resolution (HR: i.e., standard definition (SD, 720 × 480 pixels) and high definition (HD, 1280 × 720 or 1920 × 1080 pixels)) [12] application areas that take packet headers as input. The P.NAMS models can be

MMSP'13, Sept. 30 - Oct. 2, 2013, Pula (Sardinia), Italy. 978-1-4799-0125-8/13/$31.00 ©2013 IEEE.

applied in the quality monitoring of UDP-based streaming. The quality monitoring of TCP-based streaming is outside the scope of the P.NAMS. ITU-T SG12 is now investigating quality monitoring of TCP-based streaming, which is known as P.NAMS-PD (progressive download).

The P.NAMS-LR (P.1201.1) model can be used for estimating quality degradation due to coding, packet loss, and/or rebuffering artifacts. In contrast, the P.NAMS-HR (P.1201.2) model can be used for estimating quality degradation due to coding artifacts and/or packet loss. However, the model cannot be used to estimate the quality if it is affected by rebuffering.

These models have been developed based on 39 databases (16 for P.NAMS-LR and 23 for P.NAMS-HR), and the validity of the models have been verified in terms of the quality estimation accuracy. These two models were standardized by the ITU-T in October, 2012.

The P.1201.1 (i.e., P.NAMS-LR) model, which can be used to estimate the quality of mobile audiovisual media streaming services, is introduced here, as it is relevant because of the recent rapid growth of mobile IPTV services. The P.1201.1 model was developed based on individual proposals from NTT Corporation and Huawei Technologies Co., Ltd.

The remainder of this paper is structured as follows. The application of the P.1201.1 model is explained in Section II. The P.1201.1 model itself is explained in Section III. A summary of verification tests is given in Section IV. The validation results of the model are explained in Section V. Finally, we summarize and suggest possible directions for future studies in Section VI.

II. Application of P.1201.1 model

The P.1201.1 model can be used for in-service quality monitoring of audiovisual, video, and audio UDP-based streaming (i.e., RTP (real-time transport protocol) /UDP/IP). The model can be applied to an encrypted or unencrypted stream. The audio and video codecs the model can process are AMR-NB (adaptive multi-rate narrowband), AMR-WB+ (extended adaptive multi-rate wideband), AAC-LC (advanced audio coding low complexity), HE-AACv1 (high-efficiency AAC, version 1), and HE-AACv2 for audio, and MPEG-4 Part 2 and H.264/AVC (MPEG4 Part 10) for video. The applicable video resolutions are QCIF, QVGA, and HVGA.

The model can evaluate quality degradations due to coding artifacts, packet-loss artifacts, and rebuffering artifacts (also called freezing without skipping). Packet-loss artifacts depend on a PLC (packet-loss concealment) scheme, so packet-loss artifacts can be categorized as follows; 1) slicing artifacts are introduced when packet losses are concealed using the PLC scheme of the receiver in trying to repair erroneous video slices or frames; 2) freezing artifacts are introduced when the PLC scheme of the receiver replaces the erroneous frames (either due to packet loss or error propagation) with the previous error-free frame until a decoded picture without errors has been received (also called freezing with skipping). Note that the audio quality estimation model cannot estimate quality degradation caused by rebuffering. Also note that the rebuffering length and timing are calculated by the client, so the calculation of the rebuffering is beyond the scope of Recommendation ITU-T P.1201.1.

III. Algorithm description of P.1201.1 model

A block diagram of the P.1201.1 model is shown in Fig. 1. The model takes packet headers (i.e., IP, UDP, RTP packet headers) and side information (e.g., encoder and decoder (codec), PLC, and rebuffering behavior) as input. Note that rebuffering behavior needs to be derived by the terminal or other system (i.e., the derivation of rebuffering behavior is outside the scope of the P.1201.1 model).

The parameter-extraction modules (P-E) extract audio- and video-related parameters using RTP headers (i.e., RTP timestamp, sequence number, marker bit, and payload length) and rebuffering-related parameters such as the rebuffering start time and length. With these parameters, parameter calculation modules derive parameters that are used by quality estimation modules. Finally, the quality estimation modules output individual estimates of the audio, video, and audiovisual quality in terms of the 5-point absolute category rating (ACR) mean opinion score (MOS) scale defined in ITU-T P.910 [13].

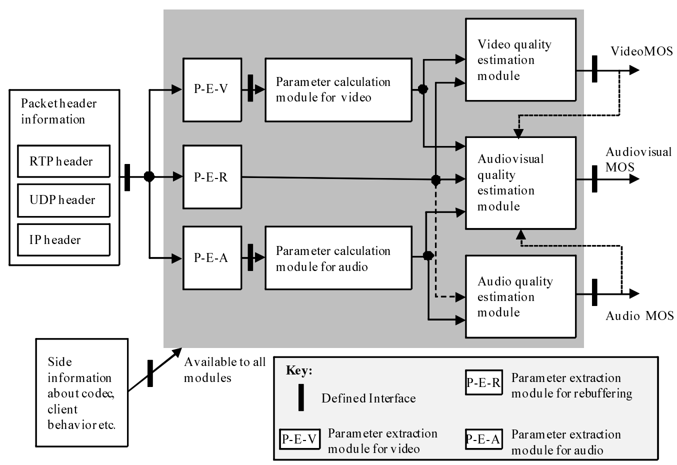

Fig. 1. Block diagram of P.1201.1 model [10]

A. Audio quality estimation module

Audio quality is affected by the codec type, coding bit rate, packet loss, and rebuffering, so it is necessary to model the relationship between the following quality factors and the subjective audio quality.

- Effect of audio codec (i.e., AMR-NB, AMR-WB+, AAC-LC, HE-AACv1, and HE-AACv2) on audio quality
- Effect of coding bit rate on audio quality
- Effect of lost audio frame length due to packet loss on audio quality
- Effect of rebuffering on audio quality (under study)

The parameter extraction module for audio (P-E-A) extracts the RTP timestamp, sequence number, and payload length. On the basis of the audio RTP timestamp and clock rate, P-E-A calculates the measurement time for audio (A_MT). On the basis of the audio RTP sequence number, P-E-A extracts the packet-loss length per packet-loss event (A_PLLk) and the audio packet-loss-event frequency (A_PLEF), as shown in Fig. 2. On the basis of the audio RTP payload, P-E-A counts the number of received audio RTP packets and the audio payload per RTP packet ($A\_receivedBytes_i$). To compensate for the lost audio payload due to packet loss, P-E-A estimates the audio payload per RTP packet ($A\_lostBytes_j$) for the lost packet in bytes using the average of the current and previous received audio RTP payloads (i.e., $A\_receivedBytes_i$ and $A\_receivedBytes_{i-1}$).

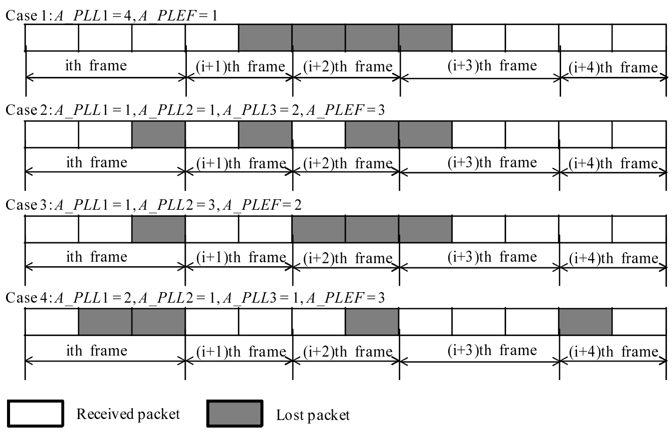

Fig. 2. Examples of calculating A_PLLk and A_PLEF [10]

To take into account the case in which one audio RTP packet contains several audio frames, the parameter calculation module for audio (P-C-A) estimates the lost audio frame length per audio RTP packet (A_LFLpP) using the audio RTP timestamp and clock rate. At the same time, P-C-A calculates the number of audio packets per RTP timestamp (A_NPpTS). Then, the lost audio frame length (A_LFL) in milliseconds is calculated using A_PLEF, A_LFLpP, the average audio burst packet loss length (A_ABPLL), A_NPpTS, and the audio frame length ($audioFrameLength$) as follows:

$$A\_LFL = A\_PLEF \cdot \max(audioFrameLength, K), \quad (1)$$

$$K = A\_LFLpP \cdot \frac{A\_ABPLL + A\_NPpTS - 1}{A\_NPpTS}. \quad (2)$$

The audio bit rate (A_BR) in kbps is calculated per RTP packet audio payload (A_Bytes_i), and the estimated amount of lost audio data ($A\_lostBytes_j$) is calculated as

$$A\_BR = \frac{8 \cdot 10^{-3}}{A\_MT} \cdot \left(\sum_{i=1}^{A\_RP} A\_Bytes_i + \sum_{j=1}^{J} A\_lostBytes_j\right), \quad (3)$$

where A_RP represents the total received number of RTP audio packets and J represents the total number of lost RTP audio packets.

Finally, the audio quality estimation module estimates audio quality (A_MOS) as follows:

$$A\_MOS = 1 + (A\_MOSC - 1) \cdot MA, \quad (4)$$

$$A\_MOSC = 1 + (a1 - \frac{a1}{1 + (A\_BR/a2)^{a3}}), \quad (5)$$

$$MA = (1 - a4) \cdot \exp(-\frac{10 \cdot A\_LFL}{a5 \cdot A\_MT}) + a4 \cdot \exp(-\frac{10 \cdot A\_LFL}{a6 \cdot A\_MT}), \quad (6)$$

where A_MOSC is audio quality due to compression, and the coefficient values ($a1$ to $a6$) can be found in [10].

B. Video quality estimation module

As with audio quality, video quality is affected by the codec type, coding bit rate, packet loss, and rebuffering. Video quality is also affected by the number of bits per video frame type because it varies depending on the spatio-temporal information of the video content. Therefore, it is necessary to model the relationship between the following quality factors and the subjective video quality.

- Effect of video codec (i.e., MPEG-4 and H.264/AVC) on video quality
- Effect of video resolution (i.e., QCIF, QVGA, and HVGA) on video quality
- Effect of coding bit rate [14], frame rate, and ratio between total bit count and I-frame (intra-coded frame) bit count on video quality
- Effect of the number of packet-loss events [15], number of damaged video frames [14], and the video frame area damaged by the packet losses on video quality
- Effect of the number of rebuffering events, average rebuffering length, and influence of multiple rebuffering events (i.e. average interval between rebuffering events) on video quality

The parameter extraction module for video (P-E-V) extracts the video RTP timestamp, sequence number, market bit, and payload.

The parameter calculation module for video (P-C-V) calculates the video packet-loss length ($V\_lostPackets$) based on the video RTP sequence number and the number of lost bytes for lost video RTP packets ($V\_lostBytes$) using the same method as that of P-E-A.

The number of lost video frames ($V\_lostFrames$) between two consecutive received video RTP packets (i.e., current and previous video RTP packets) is calculated based on the video RTP timestamp and clock rate, and the video frame rate. The marker bit and video RTP timestamp are used to identify the video frame boundary between video RTP packets. P-C-V estimates the video frame type (i.e., I- or P-frame) based on the number of bytes per video frame. The detailed frame type estimation method can be found in [10].

P-C-V calculates the average number of bytes per I-frame (V_ABIF) based on the number of I-frames and the total bytes of all I-frames. In addition, the impairment rate due to the packet loss and spatial error propagation per video frame (V_IRpF) is calculated, as shown in Fig. 3.

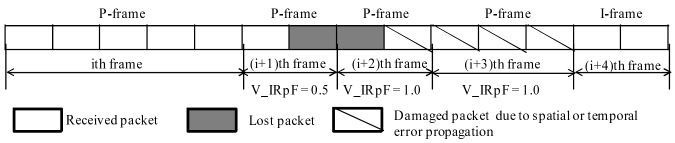

Fig. 3. Examples of calculating V_IRpF

Quality parameters for coding artifacts are calculated as follows. The measurement time for video, in seconds, is calculated based on the total number of video frames and the video frame rate ($videoFrameRate$). The video bit rate (V_BR) is calculated based on the total number of bytes and the measurement time (V_MT) as

$$V\_BR = \frac{\sum_{i=1}^{V\_TVF} V\_TBpFi}{V\_MT}, \tag{7}$$

where V_TVF represents the total number of video frames and V_TBpFi represents the number of bytes per video frame. Since the spatial quality per video frame depends on the video bit rate and video frame rate, the normalized video bit rate (V_NBR) is calculated as

$$V\_NBR = \frac{V\_BR \cdot 8 \cdot 30}{1000 \cdot \min(30, videoFrameRate)}. \tag{8}$$

Then, the video content complexity factor (V_CCF) is calculated based on the video bit rate and the average number of bytes per I-frame as

$$V\_CFF = \min(\sqrt{\frac{V\_BR}{V\_ABIF}}, 1.0). \tag{9}$$

Quality parameters for the packet loss are calculated as follows. The average impairment rate of the video frame (V_AIRF) is the sum of the impairment rate per video frame (V_IRpF) divided by the number of damaged video frames (V_NDF) that includes those damaged by temporal error propagation. The impairment rate of the video stream (V_IR) is the number of damaged video frames (V_NDF) divided by the total number of video frames (V_TNF). The method of counting the video packet-loss event frequency (V_PLEF) is different for the PLC scheme. For the PLC to generate slicing artifacts, the video packet-loss event frequency (V_PLEF) is equal to the number of video frames actually damaged by packet loss (i.e., video frames damaged by temporal error propagation are not counted). For the PLC to generate freezing artifacts, the video packet-loss event frequency (V_PLEF) must be equal to the number of damaged group of pictures (GoPs). Although the parameter extraction module for rebuffering (P-E-R) is not a part of P-E-V or P-C-V, we nevertheless explain it because the quality parameters for rebuffering are used in estimating video quality.

Quality parameters for rebuffering that are calculated by the parameter extraction module for rebuffering (P-E-R) are as follows. The number of rebuffering events (NRE) is counted. The average rebuffering length (ARL) is the average value of the total length of the rebuffering event. The multiple rebuffering events effect factor ($MREEF$) is the average value of all the rebuffering intervals between two consecutive rebuffering events, and it is calculated when $NRE > 1$. ARL and $MREEF$ are calculated as follows:

$$ARL = \sum_{i=1}^{NRE} bL_i / NRE \tag{10}$$

$$MREEF = \sum_{i=1}^{NRE-1} (bST_i - b_{i-1})/(NRE-1) \tag{11}$$

$$b_{i-1} = (bST_{i-1} + bL_{i-1}), \tag{12}$$

where bL_i represents the buffer length, bST_i represents the buffer start time, and i denotes the current rebuffering event.

The video quality estimation module estimates video quality (V_MOS) as follows. Video quality due to compression (V_MOSC) is calculated as
IF $videoFrameRate \geq 24$

$$V\_MOSC = 5 - V\_DC, \tag{13}$$

ELSE IF $videoFrameRate < 24$

$$V\_MOSC = (5 - V\_DC) \cdot F, \tag{14}$$

$$F = 1 + v1 \cdot V\_CFF - v2 \cdot V\_CFF \cdot \log(\frac{1000}{videoFrameRate}).$$

Video distortion quality due to compression (V_DC) is calculated as

$$V\_DC = 4/(1+G), \tag{15}$$

$$G = \frac{V\_NBR}{(v3 \cdot V\_CFF + v4)^{(v5 \cdot V\_CFF + v6)}}.$$

Video quality due to packet loss (V_MOSP) is calculated as

$$V\_MOSP = V\_MOSC - V\_DP. \tag{16}$$

Video distortion quality due to packet-loss (V_DP) is calculated as
IF $videoPLC = SLICING$

$$V\_DP = (V\_MOSC - 1) \cdot \frac{H \cdot I}{1 + H \cdot I}, \tag{17}$$

$$H = (\frac{V\_AIRF \cdot V\_IR}{v7 \cdot V\_CFF + v8})^{v9},$$

$$I = (\frac{V\_PLEF}{v10 \cdot V\_CFF + v11})^{v12}.$$

ElSE IF $videoPLC = FREEZING$

$$V\_DP = (V\_MOSC - 1) \cdot \frac{J \cdot K}{1 + J \cdot K}, \tag{18}$$

$$J = (\frac{V\_IR}{v7 \cdot V\_CFF + v8})^{v9},$$

$$K = (\frac{V\_PLEF}{v10 \cdot V\_CFF + v11})^{v12}.$$

Video quality due to rebuffering (V_MOSR) is calculated as

$$V\_MOSR = Video\_Quality - V\_DR, \tag{19}$$

where $Video\_Quality = V\_MOSP$ when packet loss occurs, but $Video\_Quality = V\_MOSC$ when packet loss does not occur. Video distortion quality due to rebuffering (V_DR) is calculated as

$$V\_DR = (Video\_Quality - 1) \cdot \frac{M \cdot N \cdot O}{1 + M \cdot N \cdot O}, \tag{20}$$

$$M = (NRE/v13)^{v14},$$
$$N = (ARL/v15)^{v16},$$
$$O = (MREEF/v17)^{v18}.$$

The coefficient values ($v1$ to $v18$) can be found in [10].

C. Audiovisual quality estimation module

Audiovisual quality is calculated based on audio and video quality. Audiovisual quality due to compression (AV_MOSC) is calculated as

$$AV\_MOSC = av1 \cdot V\_MOSC + av2 \cdot A\_MOSC + av3 \cdot V\_MOSC \cdot A\_MOSC + av4. \tag{21}$$

Audiovisual quality due to packet loss (AV_MOSP) is calculated as

$$AV\_MOSP = AV\_MOSC - AV\_DP. \tag{22}$$

Audiovisual distortion quality due to packet loss (AV_DP) is calculated as

$$AV\_DFV = \frac{V\_MOSC - Video\_Quality}{V\_MOSC}, \tag{23}$$

$$AV\_DFA = \frac{A\_MOSC - Audio\_Quality}{A\_MOSC}, \tag{24}$$

$$AV\_DF = \frac{av5 \cdot AV\_DFV + av6 \cdot AV\_DFA}{1 + av5 \cdot AV\_DFV + av6 \cdot AV\_DFA}, \tag{25}$$

$$AV\_DP = (AV\_MOSC - 1) \cdot AV\_DF, \tag{26}$$

where $Video\_Quality = V\_MOSP$ and $Audio\_Quality = A\_MOS$ when packet loss occurs, but $Video\_Quality = V\_MOSC$ and $Audio\_Quality = A\_MOSC$ when packet loss does not occur. Audiovisual quality due to rebuffering (AV_MOSR) is calculated as

$$AV\_MOSR = Audiovisual\_Quality - AV\_DR, \tag{27}$$

$$AV\_DR = (Audiovisual\_Quality - 1) \cdot S,$$
$$S = \frac{P \cdot Q \cdot R}{1 + P \cdot Q \cdot R},$$
$$P = (NRE/av7)^{av8},$$
$$Q = (ARL/av9)^{av10},$$
$$R = (MREEF/av11)^{av12},$$

where $Audiovisual\_Quality = AV\_MOSP$ when packet loss occurs, but $Audiovisual\_Quality = AV\_MOSC$ when packet loss does not occur. Coefficient values ($av1$ to $av12$) can be found in [10].

IV. Summary of verification test

Sixteen subjective tests were conducted for audio, video, and audiovisual sequences that were generated by varying the codec type, coding bit rate, packet-loss pattern, packet-loss concealment, and rebuffering pattern. The subjective tests were conducted for each medium and video resolution. A summary of the test conditions is given in Table I, and the detailed test plan and processing chain for generating processed audio and video can be found in [16], [17].

TABLE I
TEST FACTORS AND CODING TECHNOLOGIES [9]

Test factors the models have been validated for
Encoding (compression) degradation of audio and video with variety of bitrates Video: 40–6000 kbps Audio: 4.75–576 kbps
Packet loss degradation of audio and video (both random and bursty packet loss patterns)
Rebuffering degradation (audio-only re-buffering not validated)
Video content with different spatio-temporal complexity
Different video keyframe and frame-rates Frame rates: 5–30 fps GOP lengths (1 / keyframe rate): 2–10 sec
Different video resolutions: HVGA, QVGA, QCIF
Different decoder-side packet loss concealment strategies (freezing with skipping, one slice per RTP packet/frame)
Coding technologies models have been trained on
Video: MPEG4 Part 2, H.264 (MPEG4 Part 10)
Audio: AMR-NB/WB+, AAC-LC, HE-AACv1/v2

In the subjective quality assessment, the quality was evaluated using ACR with a 5-point scale [13]. The subjects were required to rate the quality within 5 seconds after a processed sequence was presented. The presentation order of processed sequences was randomized in these tests. The subjective score was represented as MOS.

V. Performance of P.1201.1 model

The audio, video and audiovisual quality estimation models were validated by using the cross-validation method as described in [19]. The verified results showed that no overtraining occurred in any of the three quality estimation models. Therefore, the performance of these models trained using 16 entire databases is provided. The root mean square error (RMSE) and Pearsons correlation coefficient (PCC) for audio, video, and audiovisual quality estimation models are listed in Tables II, III, and IV, respectively. From these statistical values, it can be said that the P.1201.1 model reaches a satisfactory level in terms of practical use of the in-service quality monitoring.

VI. Conclusion

The ITU-T Recommendation P.1201.1 model, which can be applied for in-service quality monitoring of mobile audiovisual media streaming services, was introduced. This model can evaluate audio, video, and audiovisual quality due to coding, packet loss, and rebuffering artifacts. The performance of

TABLE II
PERFORMANCE OF AUDIO QUALITY ESTIMATION MODEL

Codec	Degradation	RMSE	PCC	# files
AAC-LC	C, L	0.300	0.955	207
HE-AAC v1	C, L	0.354	0.960	24
HE AAC v2	L	0.429	0.808	16
AMRNB	C, L	0.263	0.755	161
AMRWBP	C, L	0.419	0.853	282
Overall		0.351	0.941	690

Note: C and L represent coding and packet loss.

TABLE III
PERFORMANCE OF VIDEO QUALITY ESTIMATION MODEL

Codec	Degradation	RMSE	PCC	# files
H264 (QCIF)	C, S, F, R	0.531	0.848	207
MPEG4 (QCIF)	C, S, R	0.461	0.829	184
H264 (QVGA)	C, S, F, R	0.529	0.851	375
MPEG4 (QVGA)	C, S, R	0.523	0.708	264
H264 (HVGA)	C, S, F, R	0.584	0.832	400
Overall		0.535	0.830	1430

Note: C, S, F, and R represent coding, slicing, freezing, and rebuffering.

audio, video, and audiovisual quality estimation models were validated based on 16 databases.

Further studies are needed to extend the model. Recently, a progressive download video service has been widely used due to advances in tablet computers and smartphones. There are two types of progressive download video (i.e., non-adaptive or adaptive streaming). For non-adaptive streaming, the P.1201.1 model needs to be extended so that it can be applied to TCP-based streaming. Since lower to higher resolutions and frame rate are used and the bit rate is adaptively changed in adaptive streaming, the model needs to be improved to evaluate such features.

REFERENCES

[1] K. Yamagishi and T. Hayashi, “Non-intrusive Packet-layer Model for Monitoring Video Quality of IPTV Services,” IEICE Transactions on Fundamentals, vol. E92-A, no. 12, pp. 3297–3306, Dec. 2009.

TABLE IV
PERFORMANCE OF AUDIOVISUAL QUALITY ESTIMATION MODEL

Codec	Degradation	RMSE	PCC	# files
H264 (QCIF), AAC-LC	C, S, R	0.360	0.951	87
H264 (QCIF), AMRNB	C, S, F, R	0.480	0.701	64
H264 (QCIF), AMRWBP	C, S, R	0.499	0.882	80
MPEG4 (QCIF), AAC-LC	C, S, R	0.420	0.754	48
MPEG4 (QCIF), AMRNB	C, S	0.379	0.917	48
MPEG4 (QCIF), AMRWBP	C, S	0.474	0.823	56
H264 (QVGA), AAC-LC	C, S, F, R	0.463	0.888	153
H264 (QVGA), AMRNB	C, S, F, R	0.443	0.811	71
H264 (QVGA), AMRWBP	C, S, F, R	0.639	0.729	100
MPEG4 (QVGA), AAC-LC	C, S, R	0.645	0.724	42
MPEG4 (QVGA), AMRNB	C, S	0.495	0.706	73
MPEG4 (QVGA), AMRWBP	C, S, R	0.395	0.749	104
H264 (HVGA), AAC-LC	C, S, F	0.422	0.927	80
H264 (HVGA), AMRNB	C, S, F	0.465	0.784	72
H264 (HVGA), AMRWBP	C, S, F	0.455	0.868	88
Overall		0.470	0.852	1166

Note: C, S, F, and R represent coding, slicing, freezing, and rebuffering.

[2] J. Gustafsson, G. Heikkila, and M. Pettersson, “Measuring Multimedia Quality in Mobile Networks with an Objective Parametric Model,” IEEE ICIP 2008, pp. 405–408, Oct. 2008.

[3] A. Raake, M. Garcia, J. Berger, F. Kling, P. List, J. Johann, and C. Heidemann, “T-V-Model: Parameter-based Prediction of IPTV Quality,” IEEE ICASSP 2008, pp. 1149–1152, Mar. 2008.

[4] K. Yamagishi, T. Kawano, and T. Hayashi, “Hybrid Video-Quality-Estimation Model for IPTV Services,” IEEE Globecom 2009, Nov. 2009.

[5] D. Hands and M. Wilkins, “A Study of the Impact of Network Loss and Burst Size on Video Streaming Quality and Acceptability,” Proc. of the 6th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services, vol. 1718, pp. 45–57, Oct. 1999.

[6] S. Mohamed and G. Rubino, “A Study of Real-Time Packet Video Quality Using Random Neural Networks,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 12, no. 12, pp. 1071–1083, Dec. 2002.

[7] K. Yamagishi and T. Hayashi, “Parametric Packet-Layer Model for Monitoring Video Quality of IPTV Services,” IEEE ICC 2008, pp. 110–114, May 2008.

[8] M. Masuda, K. Ushiki, T. Tominaga, T. Hayashi, A. Takahashi, and K. Kawashima, ”End-User QoE Estimation for Video Communication Services by Packet-Layer (in Japanese),” IEICE Transaction on Communications, vol. J94-B, no. 1, pp. 24–35, Jan. 2011.

[9] Recommendation ITU-T P.1201, “Parametric non-intrusive assessment of audiovisual media streaming quality,” Oct. 2012.

[10] Recommendation ITU-T P.1201.1, “Parametric non-intrusive assessment of audiovisual media streaming quality – lower resolution application area,” October 2012.

[11] K. Yamagishi and S. Gao, “Packet-header-based-quality-estimation Model for Mobile Audiovisual Media Streaming - Recommendation ITU-T P.1201.1 -,” Proc. of VPQM2013, pp. 52-57, Jannuary 2013.

[12] Recommendation ITU-T P.1201.2, “Parametric non-intrusive assessment of audiovisual media streaming quality – higher resolution application area,” October 2012.

[13] Recommendation ITU-T P.910, “Subjective Video Quality Assessment Methods for Multimedia Applications,” April 2008.

[14] K. Yamagishi, J. Okamoto, T. Hayashi, and A. Takahashi, “No Reference Video-Quality-Assessment Model for Monitoring Video Quality of IPTV Services,” IEICE Transactions on Communications, Vol.E95-B, No.2, February 2012.

[15] K. Yamagishi and T. Hayashi, “Non-intrusive packet-layer model for monitoring video quality of IPTV services,” IEICE Trans. Fundamentals, Vol.E92-A, No.12, pp. 3297-3306, December 2009.

[16] J. Gustafsson and A. Raake, “P.NAMS Test Plan,” ITU-T SG12, TD469rev1, Jannuary 2011.

[17] J. Gustafsson and A. Raake, “Processing Chain for P.NAMS/P.NBAMS Hypothetic Reference Circuits,” ITU-T SG12, TD644rev2, November 2011.

[18] Recommendation ITU-T P.1401, “Methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models,” July 2012.

[19] J. Gustafsson and A. Raake, “Proposed P.NAMS and P.NBAMS Optimization Procedure,” ITU-T SG12, TD895, May 2012.