PAPER *Special Section on Image Media Quality*

Non-intrusive Packet-Layer Model for Monitoring Video Quality of IPTV Services

Kazuhisa YAMAGISHI[†a)] ***and*** **Takanori HAYASHI**[†], ***Members***

SUMMARY Developing a non-intrusive packet-layer model is required to passively monitor the quality of experience (QoE) during service. We propose a packet-layer model that can be used to estimate the video quality of IPTV using quality parameters derived from transmitted packet headers. The computational load of the model is lighter than that of the model that takes video signals and/or video-related bitstream information such as motion vectors as input. This model is applicable even if the transmitted bitstream information is encrypted because it uses transmitted packet headers rather than bitstream information. For developing the model, we conducted three extensive subjective quality assessments for different encoders and decoders (codecs), and video content. Then, we modeled the subjective video quality assessment characteristics based on objective features affected by coding and packet loss. Finally, we verified the model's validity by applying our model to unknown data sets different from training data sets used above.

key words: *QoE, IPTV, quality monitoring, monitoring point*

1. Introduction

Advances in broadband Internet protocol (IP) networks and video encoders and decoders (codecs) have enabled content, network, and Internet service providers to deliver TV content over IP networks. This is called Internet protocol television (IPTV).

The quality of experience (QoE) [1] of IPTV services is affected by many factors. We focus on the video quality of IPTV services as a part of the QoE. Video-related quality factors can be summarized as follows: first, how the source video content is encoded before transmission; second, how the IP packets are carried over networks; and finally, how the encoded video is decoded and displayed at the client terminal (e.g., set-top box (STB) and mobile terminal).

International Telecommunication Union - Telecommunication standardization sector (ITU-T) Recommendation G.1081 [2] defines performance monitoring points for IPTV services that will allow the service provider and/or network operator to monitor the performance of the complete IPTV service delivery to the end user, as shown in Fig. 1. There are five performance monitoring points: source media and metadata are monitored to confirm an appropriate quality at point 1 (PT1); encoded and packetized source media is monitored at point 2 (PT2); the packet transmission characteristics are monitored at point 3 (PT3); the received packets are monitored to check that they are sufficient to provide the required QoE at the client terminal at point 4 (PT4); and displayed media is checked at the client terminal at point 5 (PT5).

Several models [4], [5], which monitor the QoE at PTs 1 and 2, have been studied and standardized by ITU. Monitoring QoS parameters (e.g., bit rate and packet-loss ratio) at PT 3 is suitable from the view point of the computational load because many streams pass through IP networks. However, little attention has been given to models for monitoring the QoE at PTs 4 and 5.

To monitor the QoE of each user at the client terminal (PTs 4 and 5), a no reference (NR) objective quality assessment model, which estimates the QoE by using information obtained only from a client terminal, is essential. Objective quality assessment models are categorized from the viewpoint of input information [3]. Media-layer models [4]–[10] take video signals, packet-layer models [11]–[13] take packet headers, bitstream-layer models [14], [15] take packet payloads (i.e., bitstream information), and hybrid models [16] take a combination of video signals, packet headers, and packet payloads as input.

In general, media-layer, packet-layer, bitstream-layer, or hybrid models can be applied to the QoE monitoring at the client terminal because video signals, packet headers, and packet payloads can be obtained from the client terminal. However, estimating the QoE with a low computational load is desirable at the client terminal because such a function usually needs to be implemented in client terminals such as home gateways (HGWs) and STBs. In this case, packet-layer and bitstream-layer models are suitable for estimating the QoE by using transmitted packet information because media-layer models require a high computational load for estimating the QoE by scanning video signals. Bitstream information (e.g., motion vector and quantization parameter) is often encrypted to protect copyrights. Packet-layer models are applicable even if the transmitted packet payload is encrypted because it uses transmitted packet headers rather than payloads. Therefore, packet-layer models are suitable for passively monitoring the QoE at an HGW and STB.

Packet-layer models are used to estimate the average video quality over typical video content using transmitted packet headers (e.g., IP, user datagram protocol (UDP), real-time transport protocol (RTP), transport stream (TS), and packetized elementary stream (PES) headers) that exclude video-related bitstream information. These models do not have access to bitstream and codec information (e.g., codec

Manuscript received March 5, 2009.
Manuscript revised July 28, 2009.

[†]The authors are with NTT Service Integration Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
a) E-mail: yamagishi.kazuhisa@lab.ntt.co.jp

DOI: 10.1587/transfun.E92.A.3297

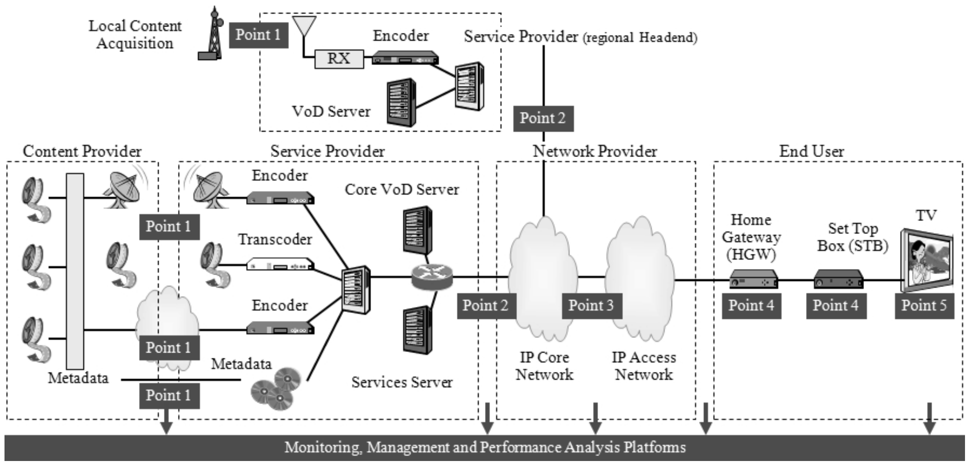


Fig. 1 QoE monitoring points for IPTV services [2].

type and codec implementation), so they cannot take into account video quality dependence on video content and codec. Therefore, they must make some assumptions with respect to video content and codec characteristics.

Conventional packet-layer models and bitstream-layer models were proposed in Refs. [12], [13] and [14], [15], respectively. The relationships between bit rate and random packet-loss ratio are taken into consideration when using several packet-layer models [12], [13]. However, consecutive IP packets are often lost by the network. In such a case, the video quality degraded by packet-loss cannot be estimated based only on random packet-loss characteristics [11]. Therefore, to improve quality estimation accuracy, packet-loss patterns need to be taken into account when using packet-layer models. Packet-layer models cannot take into account video quality dependence on video content because they do not take bitstream information and media signals as input, while bitstream-layer models can take into account video quality dependence on video content because they take bitstream information (e.g., coding parameters: frame rate, motion vector, and quantization parameters) as input. However, bitstream information is often encrypted to protect copyright. Packet-layer models are applicable even if the transmitted packet payload is encrypted because they use transmitted packet headers rather than payloads, as mentioned above. ITU-T SG12 started investigating such a model for IPTV services. This project is provisionally called Non-intrusive parametric model for the assessment of performance of multimedia streaming (P.NAMS) and is expected to be standardized as a packet-layer model in 2010.

We first propose a framework of a packet-layer model to estimate the average video quality of IPTV. We then identify video-quality characteristics affected by coding and packet loss derived from subjective quality assessments. We develop a packet-layer model for estimating the average video quality affected by coding and packet loss. We show that the mathematical equation forms of the proposed model can be unchanged for various codecs and groups of video content. After that, we verify that our model has sufficient quality-estimation accuracy for unknown data sets different from the training data sets used above. Finally, we conclude with a summary and mention further studies.

2. Proposed Framework

2.1 Concept

The packet-layer model is suitable for in-service quality monitoring at client terminals (PT4) because it can be used to estimate the QoE by using transmitted packet headers (e.g., IP, UDP, RTP, TS, and PES headers). Our proposed model for monitoring the video quality of IPTV services is shown in Fig. 2.

Our proposed model handles transmitted packet headers of TS/UDP/IP and TS/RTP/UDP/IP as input. Although there are many protocol stacks for IPTV services (e.g., TCP/IP, UDP/IP, RTP/UDP/IP, TS/UDP/IP, and TS/RTP/UDP/IP) protocol stacks of TS/UDP/IP and TS/RTP/UDP/IP are mainly used in broadcasting, retransmission, and VoD of IPTV services.

Our proposed model outputs the average video quality (Vq), which is averaged over assumed sets of video content, rather than video quality per video content. To estimate QoE accurately, it is ideal to make use of all usable information. However, this model does not have systematic access to information about the codec type (e.g., MPEG2, MPEG4 and H.264), codec implementation (e.g., motion-detection algorithm and rate-distortion algorithm), coding parameters (e.g., frame rate, group of picture, and video format), and video content because they are not included in the transmitted packet headers. Therefore, our proposed model must have some a-priori information with respect to these conditions.

The a-priori information (i.e., codec type, codec implementation, coding parameters, and video content) can be provided, for example, by an IPTV-service provider because it must know such information.

A set of video content is selected based on the qual-

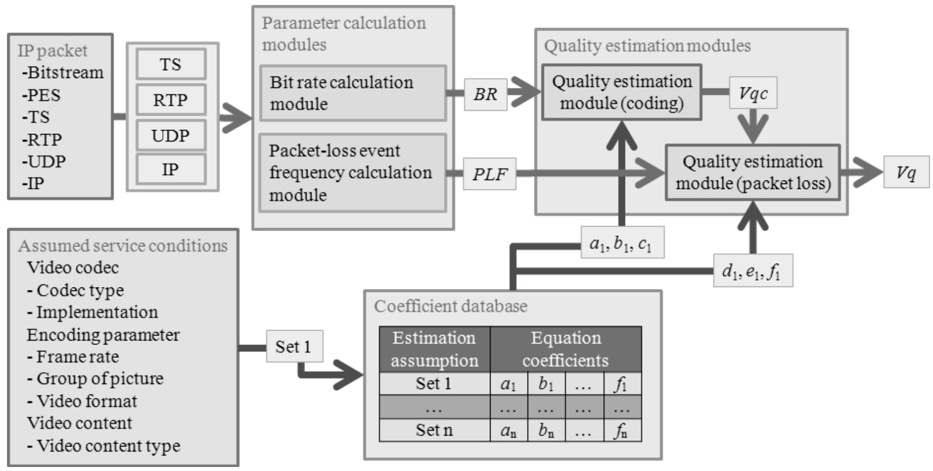


Fig. 2 Packet-layer model for monitoring video quality of HD-based IPTV services.

ity monitoring policy of the IPTV-service provider. For example, when they would like to detect even a little quality degradation for video coding, a set of video content that has high video coding difficulty should be selected.

The proposed model handles video-quality degradations affected by bit rate and packet loss. The proposed model calculates packet loss by taking into account packets discarded by a jitter buffer and packets recovered by a forward error correction (FEC) if the model is incorporated into the client terminal. Automatic repeat request (ARQ) is out of this paper's scope, so this model does not take into account quality degradations affected by re-buffering.

In general, video-quality degradations depend on assumed service conditions [17], even if bit rate (*BR*) and packet-loss ratio (*PLR*) (e.g., BR = 10 Mbps and PLR = 0.1%) of each video content are the same. However, the tendency of video-quality degradations with respect to each quality parameter (e.g., *BR* and *PLR*) does not depend on assumed service conditions. For example, video quality increases with increasing bit rate and video quality degrades with increasing packet-loss ratio. For these reasons, the mathematical equation forms of proposed model were unique while the model's coefficients were optimized for each assumed service condition.

2.2 Function

Quality degradations affected by coding and packet loss mainly depend on *BR* and packet loss (*PL*). Therefore, parameter calculation modules calculate *BR* and *PL*. First, video packets are detected by PID (See. ITU-T Rec. H.222) in the TS header. Second, *BR* is calculated based on the number of detected video TS packets. Third, *PL* is calculated based on the RTP sequence number in an RTP header and/or the continuity counter (See. ITU-T Rec. H.222) in the TS header.

The quality-estimation modules output Vq by using *BR* and *PL*. As described above, the mathematical equation's form (e.g., logistic function and exponential function) of the proposed model can be unique for all the assumed service conditions.

The mathematical equation's coefficients of the proposed model are optimized from subjective data sets under the assumed service conditions. These coefficients are stored in a coefficient database. To create the coefficient database, service providers need to conduct subjective tests for their own service systems and then optimize the coefficients using nonlinear regression analysis. For example, when packet loss occurs, video quality under different GoP lengths differs, so coefficients need to be selected for each assumed coding parameter. In addition, by definition, a packet-layer model cannot take into account video quality dependence on video content, so this model needs to assume the video content type. That is, if the characteristics of video content used for the training data sets differ from those of monitored video content in terms of spatial-temporal information (e.g., spatial detail and motion), the coefficients need to be changed.

2.3 Use of Model

Our proposed model works as follows. First, an IPTV-service provider sets an assumed service condition (e.g., set i (video codec: H.264, frame rate: 30 fps, Group of picture (GoP): M = 3 N = 15, video format: HD, video content: critical for the video quality) in Fig. 2), and then, IP packets (e.g., IP, UDP, RTP, TS headers) are input into our model. Next, parameters (i.e., *BR* and *PL*) are calculated based on packet headers, and the equation's coefficients (e.g., a_i, b_i, c_i, d_i, e_i, and f_i) from a coefficient database are selected based on the assumed service condition. After that, parameters and equation coefficients of our model are input into quality-estimation modules. The "quality-estimation module (coding)" estimates the average video quality affected by coding (Vqc) by using *BR*. Finally, the "quality-estimation module (packet loss)" outputs the average video quality (Vq), which is averaged over assumed sets of video content, by using Vqc and *PL*.

3. Subjective Quality Assessment Experiments

We built a viewing system for deriving video-quality characteristics necessary for developing the packet-layer model. In Sect. 2, we proposed that the model should estimate the average video quality, which is averaged over assumed sets of video content, by changing the model's coefficients, which were optimized for each assumed service condition, as shown in Fig. 2. Therefore, to verify the validity of the framework, we used two different types of H.264 codecs and two different groups of video content.

We conducted three different types of subjective quality assessments. As described in ITU-T Rec. P.910, video content that has various spatial-temporal characteristics (e.g., spatial detail and motion) needs to be selected for the subjective tests. We used 16 different types of video content [18] that lasted 10 seconds each, as listed in Table 1. Video contents were classified into two groups so that the ranges of criticality (See Fig. 5 of ITU-R Rec. BT.1210.3) would be almost the same. The video contents of Group A were used in Experiments 1 and 2, and the video contents of Group B were used in Experiment 3.

The coding parameters were video format, frame rate, and GoP, as listed in Table 2. The experimental parameters were *BR*, packet-loss-event frequency (*PLF*†), and averaged burst packet-loss length (*ABL*††), as listed in Table 2. To generate packet loss, we used a network emulator. Burst packet-loss lengths (*BL*s) were constant in Experiments 1 and 2, and *BL*s were varied by the uniform distribution in Experiment 3. Packet losses were generated at specified *BR*s, as shown in Table 2.

Two H.264 encoders (Products A and B) have different implementations. Product A was used in Experiments 1 and 3, and Product B was used in Experiment 2. One decoder without packet-loss concealment (PLC) used in Experiments 1 and 3 generated block noise when packet loss occurred. The other in Experiment 2 had PLC that generated freeze frames when packet loss occurred.

One IP packet was composed of seven TS packets, which were 7 × 188 bytes. Experiments 1, 2, and 3 had 880 (110 test conditions × 8 video sequences), 880 (110 test conditions × 8 video sequences), and 480 (60 test conditions × 8 video sequences) video sequences, respectively, which correspond to the number of combinations of the above-mentioned 3 experimental parameters and the number of video sequences.

Table 1 Video content for each group.

(a) Group A

No.	Title	Criticality [bit/pixel]
1	European market	0.3
2	Harbour scene	0.4
3	Whale show	0.7
4	Soccer action	0.8
5	Green leaves	1.1
6	Japanese room	0.2
7	Ice hockey	0.2
8	Weather report	0.1

(b) Group B

No.	Title	Criticality [bit/pixel]
1	Streetcar	0.2
2	Opening ceremony	1.0
3	Crowded crosswalk	0.3
4	Boy and toys	0.2
5	Buildings along the canal	0.3
6	Baseball	0.3
7	Summertime tanning	0.3
8	Flamingos	0.4

Table 2 Experimental settings for developing video-quality-estimation model (Experiments 1, 2, and 3).

(a) Codecs

Experiment 1	Product A without PLC
Experiment 2	Product B with PLC
Experiment 3	Product A without PLC

(b) Coding parameters

Video format	HD (1440 × 1080)
Frame rate	30 fps
GoP	M=3, N=15

(c) Bit rate *BR* [Mbps]

Experiment 1	Experiment 2	Experiment 3
20 p (Note 1)	20 p	20 p
18	18	17
16 p	16 p	15 p
14 p	14 p	13
12	12	11 p
10 p	10 p	9 p
8	8	7
6 p	6 p	6 p
4	4	4
2	3	2

(d) Packet-loss-event frequency *PLF*

Experiment 1	Experiment 2	Experiment 3
0	0	0
1	1	1
2	2	2
3	3	3
5	4	5
10	5	10

(e) Averaged burst packet-loss length *ABL* [Packets]

Experiment 1	Experiment 2	Experiment 3
1	1	1 (1/1) (Note 2)
2	2	2 (1/3)
3	3	8 (1/15)
5	5	-
10	10	-

Note 1: p indicates conditions with packet loss.
Note 2: Range of each *BL* is indicated in parenthesis. Left value is minimum *BL* and right value is maximum *BL*.

†*PLF* indicates the number of counted packet-loss events. For example, when a packet-loss event occurred once with five continuous lost packets, the *PLF* is 1.

††*ABL* indicates the number of lost packets divided by the number of packet-loss events. *BL* indicates the number of consecutive lost packets. For example, when two IP packets were lost consecutively, *BL* is 2. When two IP packets were lost non-consecutively, each *BL* is 1.

Table 3 Five-grade impairment scale.

Score	Impairment scale (in Japanese)
5	Imperceptible
4	Perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

In the subjective-quality assessment, the video quality was evaluated using a degradation category rating (DCR) method [19]. Subjects watched test sequences presented in pairs: the first sequence was always the source reference, while the second sequence was the same source with various levels of quality. Subjects were asked to rate the impairment of the second sequence relative to the reference by using the five-grade impairment scale listed in Table 3. The quality descriptions on the rating scale were given in Japanese. The order of presentation of video sequences and test conditions were randomized in this test.

Twenty-four subjects aged 20–39 participated in each experiment. They were non-experts who were not directly concerned with video quality as part of their work and, therefore, not experienced assessors. The subjects viewed each video sequence at a distance of 3H (about 110 cm), where H indicates the ratio of viewing distance to picture height.

Subjective video quality (Vqs) was represented as a degradation mean opinion score (DMOS) averaged over the eight types of video content in Table 1.

4. Experimental Results

We obtained the video-quality characteristics affected by coding and packet loss for each experiment.

4.1 Video-Quality Characteristics Affected by Coding

When *PLR* was 0, the subjective video qualities, which were averaged over the eight types of video content, affected by coding for each experiment are shown in Fig. 3. The 95% confidence intervals (CIs) are also shown in Fig. 3. The *BR*s were calculated from transmitted packets. Although these video-quality curves depend on video-codec implementation and group of video content, the qualitative tendency of video-quality degradations does not depend on video-codec implementation and group of video content (i.e., the Vqs increased and saturated as *BR* increased), as shown in Fig. 3. This result suggested that *BR* reduction leads to spatial quality degradation. Therefore, subjective video qualities with respect to *BR* for each video codec and group of video content can be expressed by a logistic function, where the logistic function's coefficients are different for each assumed service condition.

4.2 Video-Quality Characteristics Affected by Packet Loss

Subjective video qualities, which were averaged over the eight types of video content, affected by packet loss for each experiment are shown in Figs. 4, 5, and 6. The 95% CIs are also shown in Figs. 4, 5, and 6. *PLR*s and *PLF*s were calculated from transmitted packets. Although the degrees of video-quality degradation with respect to *PLR* were different from each other, the qualitative tendency of video-quality degradations does not depend on video-codec implementation and group of video content (i.e., the Vqs decreased as packet-loss ratio increased, as shown in Fig. 4). These results indicate that the video quality should be estimated by taking into account the burst packet-loss length. However, the degrees of video-quality degradation with respect to *PLF* were almost the same regardless of the *ABL*, and the Vqs degraded as *PLF* increased, as shown in Fig. 5. Moreover, when $BR \geq 6$ Mbps, the degrees of video-quality degradation with respect to *PLF* were almost the same regardless of *BR*, as shown in Fig. 6. Therefore, subjective video qualities with respect to *PLF* for each video codec and group of video content can be expressed by an exponential function, where the exponential function's coefficients are different for each assumed service condition.

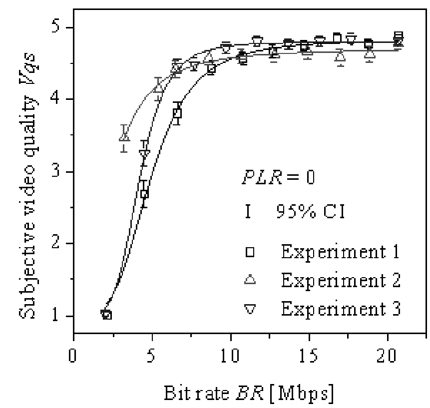


Fig. 3 Perceptual video-quality characteristics affected by coding (Experiments 1, 2, and 3).

5. Proposed Model

5.1 Quality-Estimation Module (Coding)

We developed a quality-estimation module for estimating the average video quality, which is averaged over an assumed set of video content, affected by coding (Vqs). As mentioned in Sect. 4.1, Vqs could be approximated by using a logistic function:

$$Vqs \approx Vqe\Big|_{PLF=0} = 1 + Ic = 1 + a - \frac{a}{1 + (\frac{BR}{b})^c}, \qquad (1)$$

where Vqe represents the estimated average video quality, $1 + Ic$ indicates the average video quality when *PLF* is 0, and a, b, and c are constants selected for the assumed service condition. Moreover, these coefficients are optimized for training data sets by nonlinear regression analysis.

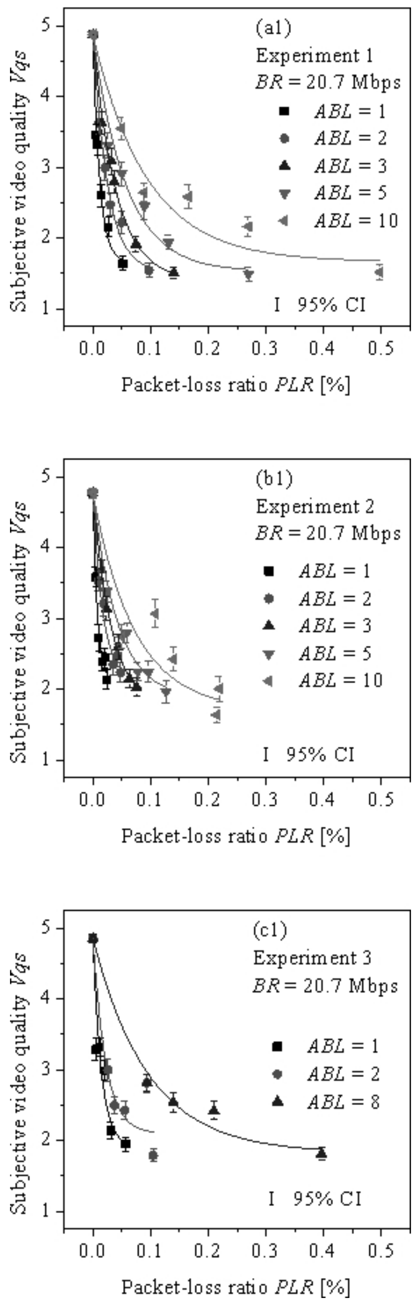


Fig. 4 Subjective video quality vs. packet-loss ratio at each averaged burst packet-loss length.

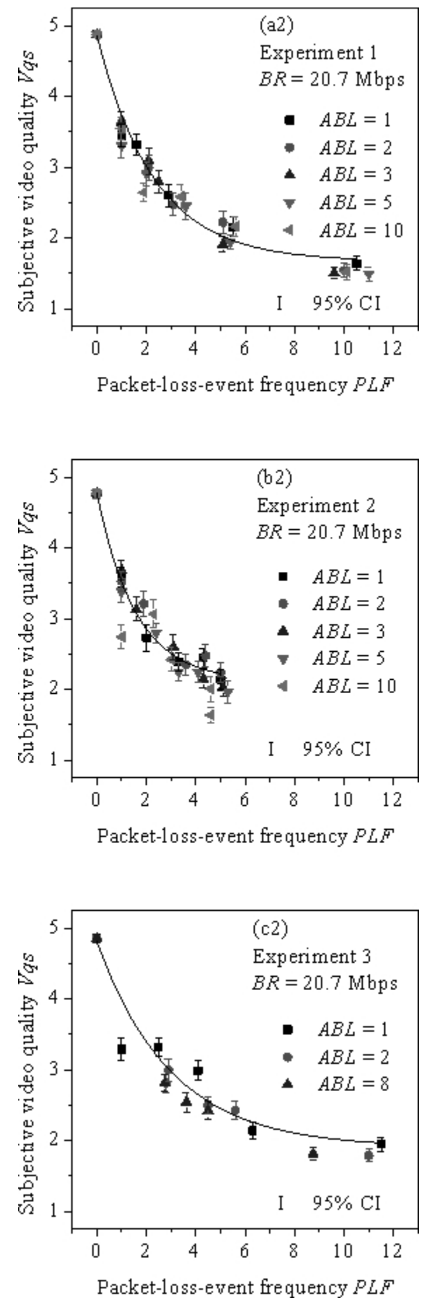


Fig. 5 Subjective video quality vs. packet-loss-event frequency at each averaged burst packet-loss length.

5.2 Quality-Estimation Module (Packet Loss)

We developed a quality-estimation module for estimating the average video quality, which is averaged over an assumed set of video content, affected by packet loss (Vqs). As mentioned in Sect. 4.2, Vqs was approximated by an exponential function. In addition, when $BR \geq 6$ Mbps, the degree of video-quality degradation for PLF is almost the same regardless of BR and ABL. From these results, Vqs was expressed as follows:

$$Vqs \approx Vqe = 1 + IcIp, \tag{2}$$

$$Ip = (1-d)\exp\left(-\frac{PLF}{e}\right) + d\exp\left(-\frac{PLF}{f}\right), \tag{3}$$

$$e \overset{def}{\Longleftrightarrow} e\Big|_{BR=BR_{Max}, ABL=1}, \tag{4}$$

and

$$f \overset{def}{\Longleftrightarrow} f\Big|_{BR=BR_{Max}, ABL=1}, \tag{5}$$

where Ip indicates the degree of video-quality degradation

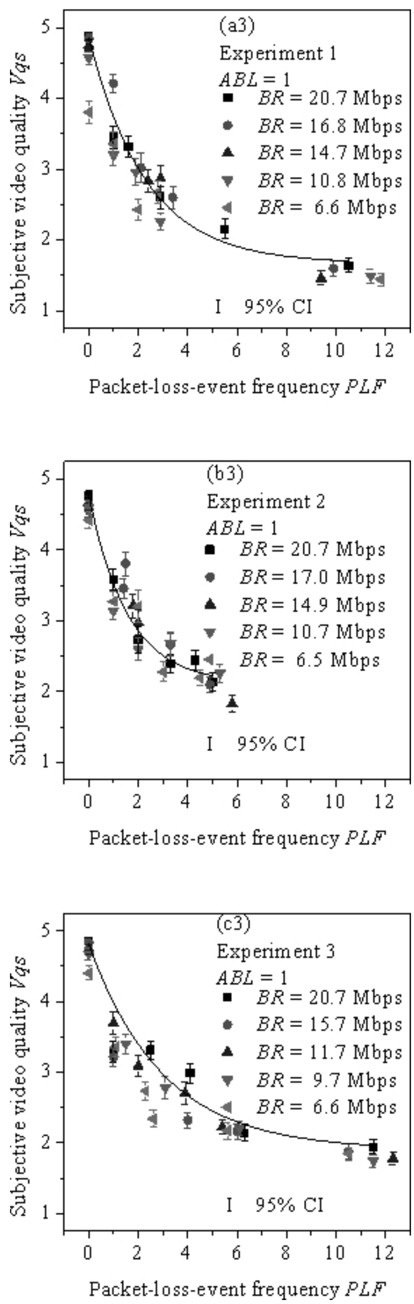


Fig. 6 Subjective video quality vs. packet-loss-event frequency at each bit rate.

for PLF, BR_{Max} denotes the maximum BR used in each experiment, and d, e, and f are constants selected for the assumed service condition. Moreover, these coefficients are optimized for training data sets by nonlinear regression analysis.

6. Performance Evaluation of Proposed Model

6.1 Performance Requirements of Quality-Estimation Module

We used Pearson-correlation coefficients (r), root mean square errors (RMSE), and outlier ratios (OR) as performance requirements to determine that the quality-estimation accuracy of the model was sufficient:

1. $r \geq 0.94$,
2. RMSE ≤ 0.27, and
3. OR ≤ 0.66.

Because the above values of r, RMSE, and OR are the same as those in ITU-T Rec. J.247 models [5] in the secondary analysis, which are averaged over all experiments, it is reasonable to use these values as criteria.

6.2 Quality-Estimation Accuracy of Quality Estimation Module (coding)

To verify the validity of the quality-estimation module (coding), we calculated the coefficients (a, b, and c) of the packet-layer model for Experiments 1, 2, and 3 based on the nonlinear least-squares approximation (NLSA), respectively. Then, by changing the coefficients of each experiment, we estimated the subjective video qualities. The quality-estimation accuracy of our model are shown in Fig. 7. The rs, the RMSEs, and the ORs are also shown.

In these experiments, rs, RMSEs, and ORs satisfy the performance requirements, as described in Sect. 6.1. Therefore, we concluded that the average video quality, which is averaged over an assumed set of video content, can be expressed by Eq. (1) with sufficient quality-estimation accuracy.

6.3 Quality-Estimation Accuracy of Quality-Estimation Module (packet loss)

To verify the validity of the quality-estimation module (packet loss), we calculated the coefficients (d, e, and f) of the packet-layer model for Experiments 1, 2, and 3 based on the NLSA, respectively. The coefficients for Experiments 1, 2, and 3 are listed in Table 4. Then, by changing the coefficients of each experiment, we estimated the subjective video qualities. The relationships between subjective video quality (Vqs) and estimated video quality (Vqe) are shown in Fig. 8. The values of r, RMSE, and OR are also shown.

In these experiments, rs, RMSEs, and ORs satisfy the performance requirements, as described in Sect. 6.1. Therefore, we concluded that the model can be applied to the quality estimation of the average video quality degraded by coding and/or packet loss, that our model can be applied to a codec with PLC that generates freeze frames when packet loss occurs, and that our model can be used to estimate the average video quality by changing the model's coefficients, which were optimized for assumed service conditions.

6.4 Validity of Proposed Model's Coefficients

To verify the validity of the optimized coefficients (a, ..., f), we conducted an experiment with a group of video content (group A) different from the group of video content (group

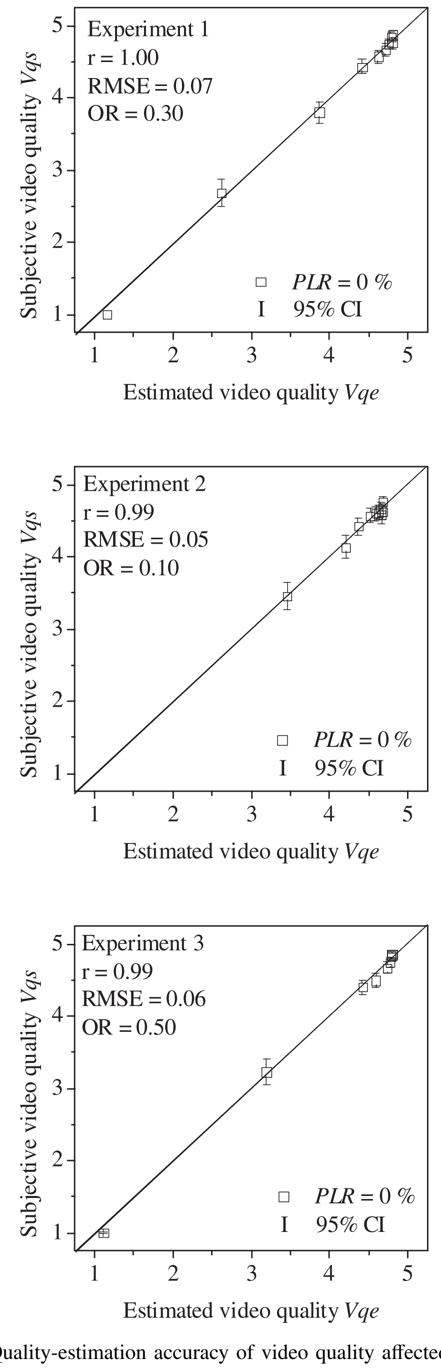


Fig. 7 Quality-estimation accuracy of video quality affected by coding for Experiments 1, 2, and 3.

B) used in Experiment 3, as listed in Table 5. In this experiment, Experiment 4, we used the same video codec and coding parameters that were used in Experiment 3. The *BL*s were variable and based on the uniform distribution in Experiment 4. Packet loss was generated at specified *BR*s, as shown in Table 5. The experiment had 480 video sequences (60 test conditions × 8 video sequences), which corresponds to the number of combinations of 3 experimental parameters (*BR*, *PLF*, and *ABL*) and the number of video sequences.

The coefficients (a, ..., f) of the fourth column in Ta-

Table 4 Coefficients of packet-layer model.

Coefficient	Experiment 1	Experiment 2	Experiment 3
a	3.82	3.70	3.80
b	4.91	2.40	4.19
c	3.65	2.31	4.80
d	0.599	0.512	0.816
e	0.948	1.14	0.0305
f	8.04	10.0	6.56

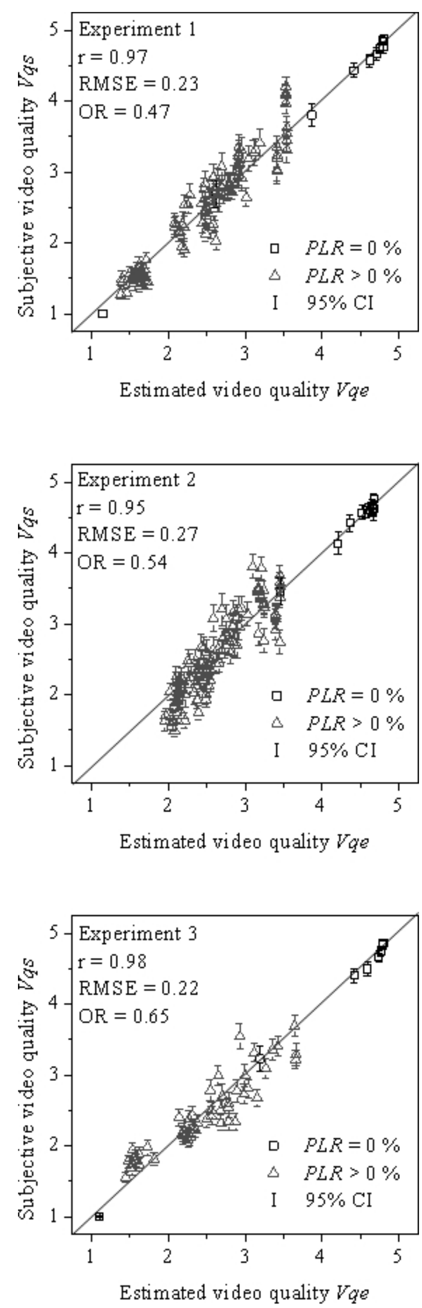


Fig. 8 Quality-estimation accuracy of video quality affected by coding and packet loss for Experiments 1, 2, and 3.

Table 5 Experimental settings (Experiment 4).

(a) Codecs

Experiment 4	Product A without PLC

(b) Coding parameters

Video format	HD (1440 × 1080)
Frame rate	30 fps
GoP	M=3, N=15

(c) Experimental parameters

BR [Mbps]	*PLF*	*ABL* [Packets]
20 p (Note 1)	0	1 (1/1) (Note 2)
17	1	2 (1/3)
15 p	2	8 (1/15)
13	3	
11 p	5	
9 p	10	
7		
6 p		
4		
2		

Note 1: p indicates conditions with packet loss.
Note 2: Range of each *BL* is indicated in parenthesis. Left value is minimum *BL* and right value is maximum *BL*.

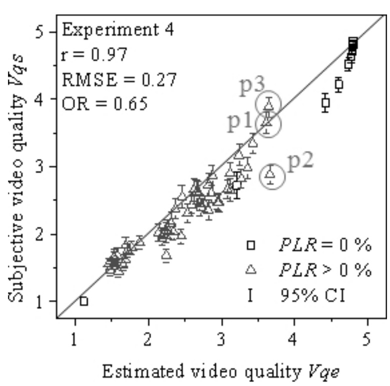


Fig. 9 Quality-estimation accuracy of video quality affected by coding and packet loss for Experiment 4.

ble 4 were trained by Experiment 3. We verified their validity using the unknown data sets shown in Experiment 4. The quality-estimation accuracy of our model is shown in Fig. 9. The values of r, RMSE, and OR are also shown. The r, RMSE, and OR satisfy the performance requirements, as described in Sect. 6.1. Therefore, if the characteristics of video content for the training and unknown data sets are almost the same, we concluded that our model with coefficients optimized for the assumed service condition is also valid for different video content.

6.5 Effects of Lost Video Frame Types

We showed that our model satisfied the performance requirements. However, some errors, which are the differences between the subjective video qualities and estimated video qualities, were greater than the 95% CIs, as shown in Figs. 7–9. In this section, we discuss the impacts of lost video frame types on the performance of the proposed model.

In general, video quality depends on lost video frame types (i.e., I-, B-, and P-frame) [20]. According to Hayashi et al. [20], there are some cases in which one lost packet leads to serious quality degradation in multiple video frames. The duration of the degradation depends on the type of video frame that has the lost packet and on the structure of a GoP. For example, when $PLF = 1$ and $ABL = 1$ in Experiment 4, the Vqs of $BR = 20.7$ Mbps was 3.7 (p1), the Vqs of $BR = 15.7$ Mbps was 2.9 (p2), and the Vqs of $BR = 11.7$ Mbps was 3.9 (p3), as shown in Fig. 9. Next, frames damaged by one lost packet were 5.8, 14.0, 5.6, respectively. That is, these results suggest that some estimation errors depend on the duration of damage caused by lost frames. However, by definition, the proposed model cannot be used to estimate the video quality by using video frame type because video frame type is included in bitstream information. Therefore, we concluded that these estimation errors were inevitable in the packet-layer model.

7. Conclusion

We proposed a packet-layer model that can be used to accurately estimate the average video quality, which is averaged over an assumed set of video content, affected by coding and packet loss. First, we conducted three subjective quality assessments to obtain two video-quality characteristics affected by coding and packet loss for two codecs and two groups of video content. We then developed a packet-layer model for a codec without PLC. After that, by changing the model's coefficients, we verified that our model can be used to estimate the average video quality for a different codec with PLC and a different group of video content. Finally, we verified that our model was valid for unknown data sets. The video quality affected by coding and packet loss can be taken into account using our model, so it is a useful QoE monitoring tool for estimating the average video quality of IPTV services.

The following issues call for further study. Although, in the case of using STBs for providing IPTV services, the client terminal is assumed to have a significant buffer capacity, for mobile terminals providing IPTV services using ARQ, re-buffering events often occurred [12]. Therefore, incorporating these quality features into the model based on quality characteristics obtained from subjective experiments is important in such a case. The proposed packet-layer model cannot be used to estimate the video quality per content because it does not use video-related information (e.g., video signals and bitstream information). Therefore, this model needs to be expanded for estimating the video quality per content by using video-related information.

References

[1] ITU-T Recommendation G.100/P.10 Amendment 1, "New appendix I — Definition of quality of experience (QoE)," Jan. 2007.
[2] ITU-T Recommendation G.1081, "Performance monitoring points for IPTV," Oct. 2008.
[3] A. Takahashi, D. Hands, and V. Barriac, "Standardization activities

in ITU on QoE assessment of IPTV," IEEE Commun. Mag., vol.46, no.2, pp.78–84, Feb. 2008.

[4] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," March 2004.

[5] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," Aug. 2008.

[6] A. Reibman, V. Vaishampayan, and Y. Sermadevi, "Quality monitoring of video quality over a packet network," IEEE Trans. Multimed., vol.6, no.2, pp.327–334, April 2004.

[7] H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol.50, no.3, pp.312–322, Sept. 2004.

[8] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol.15, no.11, pp.3441–3452, Nov. 2006.

[9] F. Yang, S. Wan, Y. Chang, and H.R. Wu, "A novel objective No-Reference metric for digital video quality assessment," IEEE Trans. Signal Process., vol.12, no.10, pp.685–688, Oct. 2005.

[10] P.L. Callet, C. Viard-Gaudin, and D. Barba, "A convolutional neural network approach for objective video quality assessment," IEEE Trans. Neural Netw., vol.17, no.5, pp.1316–1327, Sept. 2006.

[11] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," IEEE ICC 2008, pp.110–114, May 2008.

[12] 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.

[13] 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, March 2008.

[14] O. Verscheurei and X. Garcia, "User-oriented QoS in packet video delivery," IEEE Netw., vol.12, no.6, pp.12–21, Nov. 1998.

[15] S. Mohamed and G. Rubino, "A study of real-time packet video quality using random neural networks," IEEE Trans. Circuits Syst. Video Technol., vol.12, no.12, pp.1071–1083, Dec. 2002.

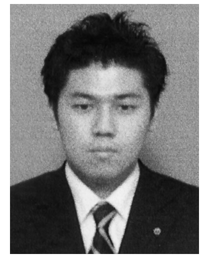
[16] D. Hands, "Quality assurance for IPTV," ITU-T Workshop on "End-to-end QoE/QoS," June 2006.

[17] H. Yang and K. Rose, "Advances in recursive per-pixel end-to-end distortion estimation for robust video coding in H.264/AVC," IEEE Trans. Circuits Syst. Video Technol., vol.17, no.7, pp.845–856, July 2007.

[18] ITU-R Recommendation BT.1210.3, "Test materials to be used in subjective assessment," Feb. 2004.

[19] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," Sept. 1999.

[20] T. Hayashi, M. Masuda, T. Tominaga, and K. Yamagishi, "Non-intrusive QoS monitoring method for realtime telecommunication services," NTT Technical Review, vol.4, no.4, pp.35–40, April 2006.



Kazuhisa Yamagishi received a B.E. degree in Electrical Engineering from Tokyo University of Science and a M.E. degree in Electronics, Information, and Communication Engineering from Waseda University in Japan. He joined NTT Laboratories in 2003. He has been engaged in subjective quality assessment of multimedia telecommunications and image coding. Currently, he is working on the quality assessment of multimedia services over IP networks. He has been contributing to ITU-T SG12 since 2006. He received the Young Investigators' Award (IEICE) in Japan in 2007 and the Telecommunication Advancement Foundation Award in Japan in 2008.

Takanori Hayashi received B.E., M.E., and Ph.D. degrees in Engineering from the University of Tsukuba, Ibaraki, in 1988, 1990, and 2007. Since joining NTT Laboratories in 1990, he has been engaged in the subjective quality assessment of multimedia telecommunication and network performance measurement methods. He is currently working on a multi-modal quality assessment method over IP networks. He received the Telecommunication Advancement Foundation Award in Japan in 2008.