PAPER

No Reference Video-Quality-Assessment Model for Monitoring Video Quality of IPTV Services

Kazuhisa YAMAGISHI[†a)], **Jun OKAMOTO**[†], **Takanori HAYASHI**[†], ***and*** **Akira TAKAHASHI**[†], ***Members***

SUMMARY Service providers should monitor the quality of experience of a communication service in real time to confirm its status. To do this, we previously proposed a packet-layer model that can be used for monitoring the average video quality of typical Internet protocol television content using parameters derived from transmitted packet headers. However, it is difficult to monitor the video quality per user using the average video quality because video quality depends on the video content. To accurately monitor the video quality per user, a model that can be used for estimating the video quality per video content rather than the average video quality should be developed. Therefore, to take into account the impact of video content on video quality, we propose a model that calculates the difference in video quality between the video quality of the estimation-target video and the average video quality estimated using a packet-layer model. We first conducted extensive subjective quality assessments for different codecs and video sequences. We then model their characteristics based on parameters related to compression and packet loss. Finally, we verify the performance of the proposed model by applying it to unknown data sets different from the training data sets used for developing the model.

key words: *QoE, IPTV, monitoring, compression, packet loss*

1. Introduction

Significant progress has recently been made in the development of technologies such as encoders and decoders (codecs) [1], [2] and networks. In addition, there has been domestic and international standardization [3]–[8] regarding Internet protocol television (IPTV). As a result, content, network, and Internet service providers can deliver high-definition television (HDTV) content over IP networks.

In general, IPTV services are provided using user datagram protocol (UDP) or transmission control protocol (TCP) packets. To avoid transmission delay, a UDP-based stream is widely used for broadcasting and for video on demand (VoD). On the other hand, TCP and automatic repeat request (ARQ) are the retransmission schemes used to recover lost packets in VoD. In this paper we focus on UDP-based streams because they are widely used for IPTV services. Retransmission by TCP- and ARQ-based streams is outside the scope of this paper.

The quality of experience (QoE) [9] of IPTV services is affected by many factors, such as media quality, service fees, and customer support. This work considers the video quality (VQ) of IPTV services as a part of QoE because VQ is a dominant factor in IPTV services.

Manuscript received April 6, 2011.
Manuscript revised October 5, 2011.

[†]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/transcom.E95.B.435

Content, network, and Internet service providers need to conduct quality planning and monitoring to provide a high-quality IPTV service. First, on the basis of subjective video quality characteristics, an appropriate codec (e.g., H.264 and MPEG-2) and coding parameters (e.g., video format, group-of-picture (GoP), bit rate (BR), and frame rate) should be determined and an appropriate network performance should be designed. This is called QoE planning. Next, while this designed service is provided to end-users, the QoE should be monitored in real time to check that an end-user is watching high-quality IPTV content. This is called in-service QoE monitoring. Subjective quality test results can be applied to QoE planning even if an objective quality assessment model has not been developed. However, when providers monitor the QoE, an objective quality assessment model is indispensable because QoE monitoring should be carried out in real time.

Service management is conducted in three domains: the head-end (H/E), network, and end-user premises.

In H/E, H/E QoE monitoring is conducted to check encoding quality. If source video signals (i.e., pixel signals) are available, full reference (FR) media-layer models [10]–[15], which take source and encoded video signals as input, are suitable. In contrast, if source video signals cannot be obtained (e.g., source video signals do not exist because the video was encoded at a different site), no reference (NR) media-layer models [16]–[18] are suitable.

In a network, network monitoring is conducted to confirm network performance. Monitoring quality of service (QoS) parameters, such as throughput, packet-loss ratio, and delay, is suitable from the viewpoint of computational power because many streams pass through IP networks.

In end-user premises, end-user QoE monitoring is carried out to check the QoE affected by the source video, encoder, network performance, decoder, and display. Reduced reference (RR) media-layer [19], [20], NR media-layer, NR packet-layer [21]–[24], NR bitstream-layer [25]–[29], and NR hybrid models [30]–[33] can be applied to end-user QoE monitoring. To estimate the QoE, RR media-layer models take degraded video signals and features derived from the source as input, and NR media-layer models take degraded video signals as input. Packet-layer models take transmitted packet headers (e.g., IP, UDP, real-time transport protocol (RTP), transport stream (TS) [8], and packetized elementary stream (PES) [8] headers) without bitstream information as input. Bitstream-layer models take bitstream information (e.g., quantization parameters and motion vectors)

as input, and hybrid models take a combination of video signals, transmitted packet headers, and bitstream information as input.

Although FR and RR media-layer models [10], [11], [15], [19] have been standardized and studied thoroughly, NR media-layer, packet-layer, bitstream-layer, and hybrid models that can be applied to end-user QoE monitoring have not been standardized and are currently being studied. Therefore, we focus on end-user QoE monitoring in this work.

End-user QoE monitoring can be classified as temporary or continuous monitoring. Temporary monitoring is carried out in cases where, for example, a repairman needs to fix a device. This is called on-site end-user QoE monitoring. In contrast, continuous monitoring is carried out to monitor the QoE of all users or a certain group of users in real time. This is called real-time end-user QoE monitoring. In on-site end-user QoE monitoring, the issue of computational power is not serious because a repairman can take a device, such as a laptop computer, to an end-user's home. Therefore, in general, although media-layer or hybrid models require high computational power to scan video signals and to estimate QoE, they can be applied to on-site end-user QoE monitoring. However, in real-time end-user QoE monitoring, as Gustafsson et al. described [22], high computational power is fatal from the viewpoint of the cost of a client terminal, such as a set-top box or home gateway. Therefore, incorporating an NR objective-quality-assessment model with low computational power into such a client terminal is desirable.

When transmitted packets are scrambled (encrypted), a bitstream model cannot be used for monitoring the VQ because bitstream models take bitstreams as input. Therefore, a packet-layer model is more suitable for real-time end-user QoE monitoring.

We previously developed a packet-layer model [21] for estimating the average VQ of typical video content assumed by an IPTV service provider using the BR and packet-loss information. With this model, assumptions about video content are made because it has access to only transmitted packet headers. That is, there is a difference in VQ between the VQ perceived by the user and that estimated using the model.

In this paper, we focus on the development of an NR video-quality-assessment model that can be used for estimating the VQ per video content (Q). As described above, the conventional packet-layer models cannot calculate the impact of video content on the VQ. Therefore, we propose a model that calculates the difference in VQ between the VQ of the estimation-target video and the average VQ estimated using our packet-layer model [21]. The proposed model uses the average bits over I-frame, ABI (BI), in addition to the BR (B) and the number of video frames damaged by packet loss, DF (D), which conventional models [21]–[24], [29] also use. In general, the video-frame type (i.e., I-, P-, or B-frame) is not indicated in the transmitted packet headers. Therefore, we incorporate the video-frame-type-estimation model [34], which estimates the video-frame type, into our model because the estimation errors for I-, B-, and P-frames are low.

We first describe conventional packet-layer models in Sect. 2. We then present the concept of the proposed model used for estimating the VQ of IPTV in Sect. 3. Subjective quality assessments and subjective video quality characteristics affected by compression and packet loss derived from the subjective quality assessments are explained in Sect. 4. Then, mathematical equations of the proposed model that are unchanged for various codecs are shown in Sect. 5. We verify in Sect. 6 that the proposed model has sufficient quality-estimation accuracy for unknown data sets that are different from the training data sets and also show that it is superior to conventional models in terms of quality-estimation accuracy. Finally, we conclude with a summary and mention further studies in Sect. 7. Abbreviations used in this paper are listed in Table A· 1 in the Appendix.

2. Related Work

Conventional packet-layer models [21]–[24] estimate the average VQ of typical video content by using transmitted packet headers that exclude video-related bitstream information. These models do not have access to video signals and bitstream and codec information (e.g., codec type and implementation), so assumptions must be made with respect to video content and codec characteristics.

Several models for estimating the average VQ affected by compression by using the BR have been proposed [21]–[23]. In general, as described in Ref. [31], VQ affected by compression depends on video content and indicates that compression affecting VQ cannot be taken into account using only the BR. That is, the model requires more information to take into account the impact of video content on the VQ.

Several models for estimating the average VQ by using the packet-loss ratio have been proposed [22], [23]. Consecutive IP packets are often lost by a network. In such a case, the VQ degraded by packet loss cannot be estimated based only on random packet-loss characteristics [35]–[37]. Therefore, to improve quality-estimation accuracy, several packet-layer models [21], [24] have been proposed for estimating the average VQ based on packet-loss-event frequency (PLEF), which indicates the number of counted packet-loss events (e.g., if a packet-loss event occurs once with five continuous lost packets, the PLEF is 1). However, the models [21], [24] using PLEF are not sufficient in terms of estimating the VQ per video content because, according to Masuda et al. [29], the VQ affected by packet loss depends on the positions of lost video-frame types (i.e., I-, P-, and B-frames). In addition, one lost packet may lead to serious quality degradation in multiple video frames in some cases, as shown in Fig. 1. The DF[†] depends on the video-frame type that has the lost packet and on the GoP structure.

[†]In the H.264 codec, the multiple-reference-frames mode (MRFM) can be used for motion compensation. In such a case,

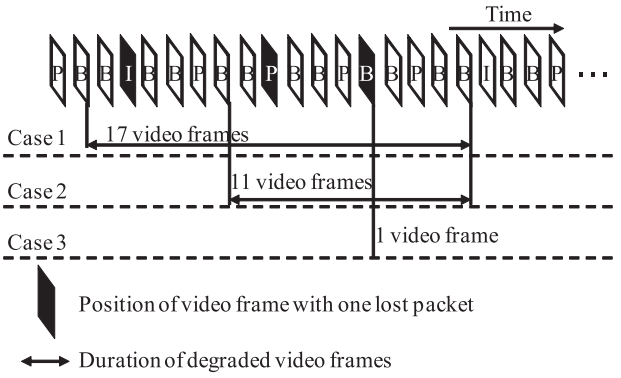
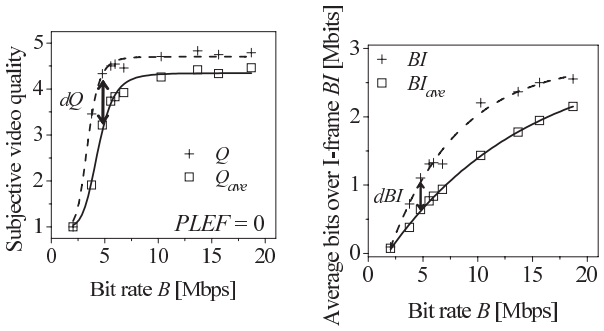


Fig. 1 Various numbers of damaged video frames.



(a) BR vs. subjective video quality. (b) BR vs. average bits over I-frame.

Fig. 2 Impact of video content on subjective video quality and average bits over I-frame.

For example, in Case 1 in Fig. 1, 17 video frames are degraded by 1 lost packet in an I-frame. In a P-frame, 5–14 video frames (11 frames in Case 2 in Fig. 1) are degraded by one lost packet, where the DF depends on the position of the P-frame with the lost packet. In Case 3 in Fig. 1, one video frame is degraded by one lost packet in a B-frame. That is, it is essential to take into account the DF to estimate the VQ degraded by packet loss.

ITU-T SG12 has started investigating packet-layer models for IPTV services. This project is provisionally called the non-intrusive parametric model for the assessment of the performance of multimedia streaming (P.NAMS) and is expected to be standardized in 2012.

3. Concept of Proposed Model

As described in Sect. 2, conventional packet-layer models that take the BR and packet-loss information as input cannot be used for estimating the VQ per video content. That is, there is much room for improvement in quality-estimation accuracy. To address these issues, we propose a model that calculates the difference in VQ (dQ) between the VQ of the estimation-target video and the average VQ (Q_{ave}) estimated using the packet-layer model [21]. The dQ comes from the bit allocation per video frame. It especially depends on the bit amount of the I-frame. When $B = 10$ Mbits/sec (Mbps), the difference between BI and BI_{ave}, which represents BI averaged over typical video sequences, affects the dQ, as shown in Fig. 2[†]. In addition, dQ depends on B. Therefore, the dQ is mainly estimated using the difference between the BI and BI_{ave} (dBI) and B. A diagram of the proposed model is shown in Fig. 3. The mathematical equations of the proposed model are shown in Sect. 5 after characteristics derived from subjective quality assessments are presented in Sect. 4.

the approach Masuda et al. proposed [29] cannot be used for QoE estimation because the duration of degradation is not determined by only video-frame type. However, in broadcasting, service providers need to broadcast TV content with short delay, so MRFM is often not used. Even if MRFM is used in broadcasting, the approach can be used for QoE estimation because only 2–4 frames for MRFM are used for motion compensation to avoid long delay [4].

3.1 Definition of Proposed Model

Our proposed model outputs the estimated VQ per video content (Q) based on transmitted packet headers.

As described in Sect. 1, our proposed model handles UDP-based streams that consist of PES/TS/UDP/IP or PES/TS/RTP/UDP/IP as input. However, the proposed model cannot be applied to cases where it does not have access to information about the boundary of a video frame. The proposed model uses the following information to derive video-frame type. Information about the boundary of a video frame in a packet stream is often contained by the payload-unit-start indicator (PUSI) in the TS header [4], [5], [8] or by the presentation time stamp (PTS) in the PES header [4], [5], [8]. Whether the PUSI or PTS contains the information about the boundary of a video frame is described in the domestic and/or international standards. In such cases, our proposed model uses the model that Ushiki et al. proposed [34] to estimate the video-frame type (i.e., I-, P-, or B-frame), as described in Sect. 1. Video-frame-type estimation is out of this paper's scope because Ushiki et al. have already proposed a model that appropriately estimates the video-frame type based on the number of packets per video frame.

We apply the framework in Ref. [21] to the proposed model. To accurately estimate the QoE, it is ideal to use all usable information. However, this model does not have systematic access to information about the codec type (e.g., MPEG-2, MPEG-4, and H.264) and codec implementation (e.g., motion-detection algorithm, rate-distortion algorithm, and packet-loss concealment (PLC)). In addition, coding parameters (e.g., profile and level, video format, frame rate, and GoP) cannot be used for QoE estimation when they are scrambled. Therefore, our proposed model requires a priori information, as shown in Fig. 3. A priori information (i.e., video codec and coding parameters) can be provided, for ex-

[†]The characteristics are derived from the experiment described in Sect. 4. The experimental conditions and characteristics are described in Sect. 4. Subjective video quality is represented as a mean opinion score (MOS).

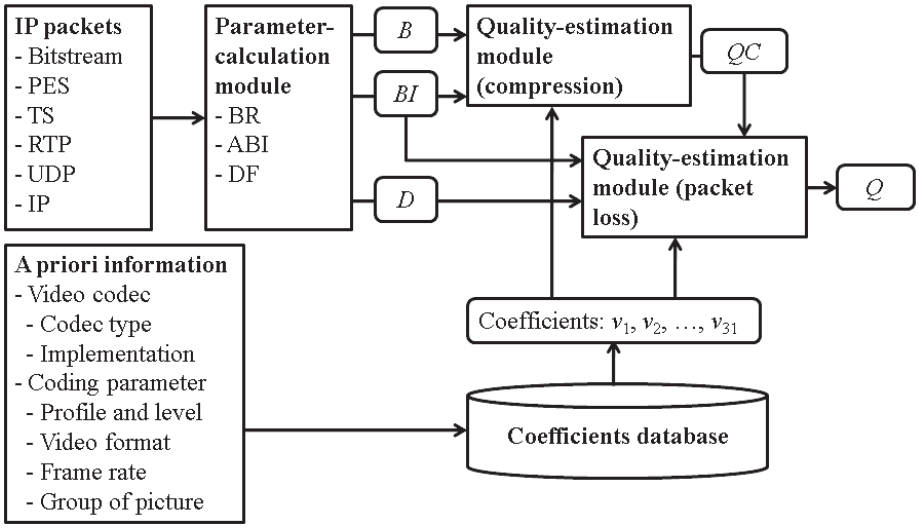


Fig. 3 NR video-quality-assessment model for IPTV services.

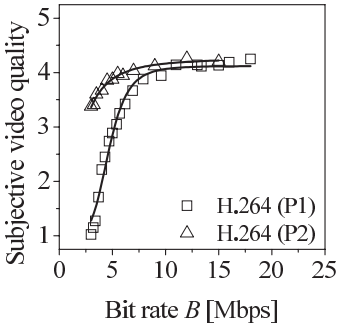


Fig. 4 Relationship between bit rate and subjective video quality for two different codecs.

ample, by an IPTV service provider because it must know such information to encode TV content.

We use the framework [21] that optimizes a model's coefficients for a priori information because it is impossible to develop a universal model for various codecs without using video signals. VQ degradations depend on a priori information, even if the BR (e.g., $B = 5$ Mbps) of each video content is the same [21], as shown in Fig. 4[†]. However, the tendency of VQ degradations with respect to the BR does not depend on a priori information. For example, the VQ increases with BR [21], and the VQ degrades with increasing DF [29]. That is, the mathematical equation forms (e.g., logistic and exponential functions) of the proposed model can be unified, while the model's coefficients are optimized for a priori information (i.e., video codec and coding parameters). To create a coefficient database, service providers need to conduct subjective tests for their own service systems and then optimize the coefficients by using nonlinear regression analysis.

3.2 Function of Proposed Model

The coefficient database stores coefficients of the proposed model, and those corresponding to a priori information (e.g., video encoder: H.264, profile and level: high profile level 4, video format: HD, frame rate: 30 fps, GoP: M = 3 N = 15, and video decoder without PLC) that an IPTV service provider gives are output to quality-estimation modules.

The parameter calculation module calculates the parameters B, BI, and D based on transmitted packets (e.g., PES/TS/RTP/UDP/IP). Video packets are first detected by a packet identifier (PID) [8] that a program map table (PMT) indicates. Second, B is calculated based on the number of detected video TS packets and the arrival time of transmitted packets. Third, BI is calculated based on the number of detected video TS packets of an I-frame. Next, the positions of lost video frames are detected based on the RTP sequence number in an RTP header and/or the continuity counter [8] in the TS header. Finally, D is calculated based on the position of the lost video frame and the video-frame type.

The quality-estimation module (compression) outputs the compression affecting VQ (QC), using the parameters B and BI and the equation's coefficients (i.e., v_1–v_{20}).

The quality-estimation module (packet loss) outputs Q using the parameters QC, BI, and D and the equation's coefficients (i.e., v_{21}–v_{31}).

4. Subjective Quality Assessment

We conducted subjective quality assessments to derive the subjective quality characteristics necessary for developing the proposed model and to verify the validity of the model.

4.1 Experimental Conditions

We conducted 4 different types of subjective quality assess-

[†]This characteristic is derived from the experiment described in Sect. 4. The experimental conditions can be found in Sect. 4. Subjective video quality is represented as a mean opinion score (MOS).

Table 1 Experimental settings (Experiments 1–4).

(a) Codecs.

Experiments 1 and 2	H.264 (Product P1*)
Experiments 3 and 4	H.264 (Product P2*)

*: Video decoder does not have PLC.

(b) Coding parameters.

Profile and level	High profile level 4
Video format	HD (1440 × 1080i)
Frame rate	30 fps
GoP	M=3, N=15

(c) Bit rate B [Mbps].

Experiment 1	18 p**, 15, 13, 9.6, 6.2, 5.4, 5.0, 4.3 p, 3.4, 2.0
Experiment 2	16, 13.4, 11, 7.9, 6.9 p 5.7 p, 4.7, 4.0, 3.7, 3.0
Experiment 3	15 p, 9 p, 7, 6, 5 4.5 p, 3.5 p, 3.0 p
Experiment 4	12 p, 5.5, 4.0, 3.25 p

**: p indicates conditions with packet loss.

(d) Packet-loss-event frequency $PLEF$.

Experiment 1	1, 2, 4, 12
Experiment 2	1, 2, 4, 12
Experiment 3	1***, 2, 4, 7
Experiment 4	1, 2, 4, 7

***: We emulate three times for $PLEF = 1$, where lost positions differ from each other.

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

Experiment 1	1 (1/1)****, 2 (1/3), 4 (1/7), 8 (1/15)
Experiment 2	1 (1/1), 2 (1/3), 4 (1/7), 8 (1/15)
Experiment 3	1 (1/1)
Experiment 4	8 (1/15)

****: Range of each BL is indicated in parentheses. Left value is minimum BL, and right value is maximum BL.

ments (Experiments 1–4) for 16 source video sequences (SRCs) to train our proposed model and to verify whether the model can appropriately estimate the VQ per video sequence for unknown data. As described in Sect. 3, our proposed model requires a priori information for optimization, so we used two different H.264 encoders to verify the framework proposed in Ref. [21]. The two H.264 encoders (Products P1 and P2) have different implementations. Product P1 was used for Experiments 1 and 2, and Product P2 was used for Experiments 3 and 4, as listed in Table 1(a). In these experiments, we used a decoder without PLC because such decoders are widely used in existing IPTV services. The decoder used in all the experiments generated block noise when packet loss occurred. The same coding parameters (i.e., profile and level, video format, frame rate, and GoP) listed in Table 1(b) were used in these experiments because their values are widely used in Japan and some other countries. Although the encoded video format was 1440 × 1080i, the encoded video was displayed at 1920 × 1080 (Full HD), the native resolution of a 42-inch LCD monitor. We used different combinations of BR and packet loss to verify whether the model can appropriately estimate the VQ for unknown experimental conditions. We used a random and burst packet loss because they occur in networks. As listed in Tables 1(c)–(e), the experimental parameters were B, $PLEF$, and averaged burst packet-loss length (ABL), where 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 are lost consecutively, BL is 2. When two IP packets are lost non-consecutively, each BL is 1.). To generate packet loss, we used a network emulator. BLs were varied by the uniform distribution, and packet losses were generated at specified Bs, as listed in Table 1. One IP packet was composed of seven TS packets, where one TS packet was 188 bytes.

Table 2 Video sequence for each group.

(a) Group A.

No.	Title	SI	TI
1	Soccer action	139	44
2	Green leaves	115	39
3	Baseball	65	34
4	Weather report	81	11
5	Streetcar	77	25
6	Buildings along the canal	92	26
7	Summertime tanning	85	97
8	Flamingos	39	18
Average		87	37

(b) Group B.

No.	Title	SI	TI
9	European market	93	68
10	Harbour scene	112	30
11	Whale show	112	45
12	Japanese room	66	28
13	Opening ceremony	129	15
14	Crowded crosswalk	79	31
15	Boy and toys	53	27
16	Ice hockey	66	26
Average		89	34

As described in ITU-T Rec. P.910, a video sequence that has various spatial-temporal characteristics (e.g., spatial detail and motion) needs to be selected for subjective tests. We used 16 different types of SRCs [38] that each lasted 10 seconds (300 frames) to verify whether our proposed model can estimate the VQ per video sequence. The SRCs were classified into two groups (A and B) so that the ranges of SI and TI would be almost the same, as listed in Table 2. The maximum values of spatial information (SI) and temporal information (TI) [39] in 300 frames are also listed.

For each video group (A or B), Experiments 1 and 2 had 336 (42 test conditions × 8 SRCs) processed video sequences (PVSs), Experiment 3 had 304 (38 test conditions × 8 SRCs), and Experiment 4 had 128 (16 test conditions × 8 SRCs). The number of PVSs corresponded to the number of combinations of the above-mentioned three experimental parameters and the number of SRCs, respectively. For compression, Experiments 1 and 2 had 10 test conditions, Experiment 3 had 8, and Experiment 4 had 4. For packet loss, Experiments 1 and 2 had 32 test conditions, Experiment 3 had 30, and Experiment 4 had 12.

Table 3 Five-grade quality scale.

Score	Quality scale (in Japanese)
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

Here, we summarize the purpose of each experiment. Experiment 1 or 3 for video group A (training data) was used for training the model that requires a priori information for optimization. Experiment 1 or 3 for video group B (unknown data) was used for verifying whether the model can estimate the VQ for an unknown video sequence with the same experimental conditions as the training data. Experiment 2 or 4 for video group A (unknown data) was used for verifying whether the model can estimate the VQ for an unknown combination of BR and packet loss with the same video sequence as the training data. Experiment 2 or 4 for video group B (unknown data) was used for verifying whether the model can estimate the VQ for an unknown video sequence and combination of BR and packet loss.

In the subjective-quality assessment, the subjective video quality was evaluated using an absolute category rating (ACR) method [39] with the five-grade quality scale shown in Table 3. The quality descriptions on the rating scale were given in Japanese. The presentation order of the PVSs was randomized in these tests.

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

Subjective video quality was represented as a mean opinion score (MOS) per PVS.

4.2 Characteristics Derived from Experiments

In this section, we present how the *BI* affects the VQ using only the results from Experiment 1 for video group A because the results from the other experiments were similar to those from this experiment.

4.2.1 Bits Characteristics of I-Frame

The relationship between the B and BI for video group A in Experiment 1 is shown in Fig. 5. The definitions of BI_{ave}, BI_{max}, and BI_{min} are as follows: BI_{ave} is ABI averaged over eight video sequences for each experimental condition, BI_{max} represents the maximum ABI in eight video sequences for each experimental condition, and BI_{min} represents the minimum ABI in eight video sequences for each experimental condition. Bs were calculated from transmitted packets. As shown in Fig. 5, the curves of BI_{ave}, BI_{max}, and BI_{min} with respect to B can be expressed by an exponential function because, in general, the increase in B results in an increase in BI. However, the coefficients of their exponential functions are different for each codec because the increase ratio of B depends on the video sequence and codec. When B is constant, BI is high in the video sequence without large motion (i.e., the weather report, which had the lowest TI value in video group A) because P- and B-frames do not need more bits than those of a video sequence with average motion, and vice versa.

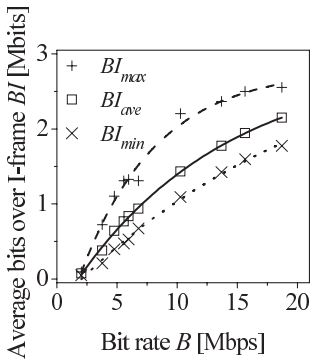


Fig. 5 Bit rate vs. average bits over I-frame.

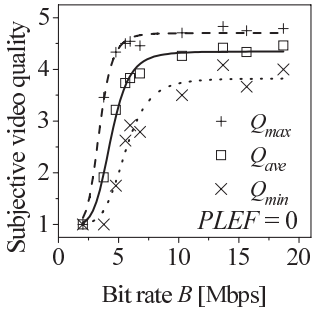


Fig. 6 Perceptual quality characteristics affected by compression.

4.2.2 Quality Characteristics Affected by Compression

For $PLEF = 0$, the Q_{ave}, maximum video quality (Q_{max}), and minimum video quality (Q_{min}) affected by compression for video group A in Experiment 1 are shown in Fig. 6. The definitions of Q_{ave}, Q_{max}, and Q_{min} are as follows: Q_{ave} is VQ averaged over eight video sequences for each experimental condition, Q_{max} represents the maximum VQ in eight video sequences for each experimental condition, and Q_{min} represents the minimum VQ in eight video sequences for each experimental condition. The Bs were calculated from transmitted packets. These results suggested that BR reduction led to spatial quality degradation and that these curves depended on the video sequence. Although these VQ curves depended on the video sequence, the qualitative tendency of VQ degradations did not (i.e., the subjective video quality increased and saturated as the BR increased), as shown in Fig. 6. Therefore, Q_{ave}, Q_{max}, and Q_{min} with respect to the B could be expressed by a logistic function, while their logistic function's coefficients were different for each codec.

We investigated the relationships between the B and ei-

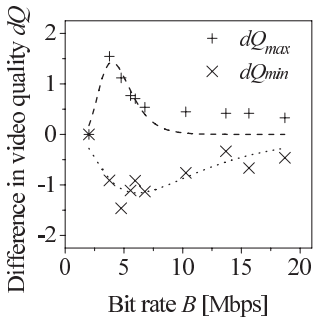


Fig. 7 Bit rate vs. difference in video quality.

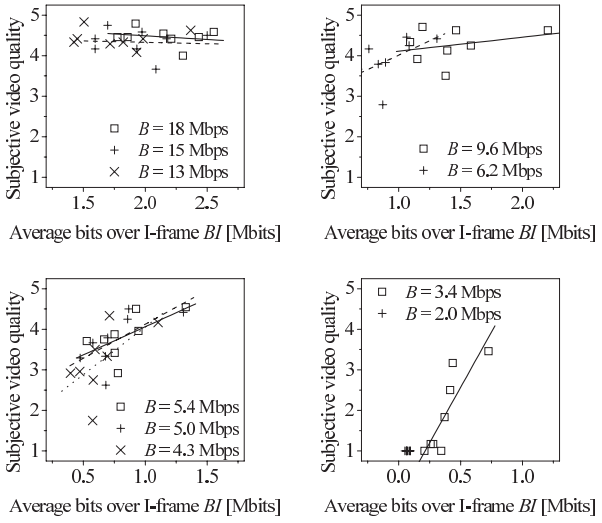


Fig. 8 Average bits over I-frame vs. subjective video quality.

ther dQ_{max} or dQ_{min}, as shown in Fig. 7. The definitions of dQ_{max} and dQ_{min} are as follows: dQ_{max} represents the difference between Q_{max} and Q_{ave}, and dQ_{min} represents the difference between Q_{min} and Q_{ave}. The results showed that dQ_{max} and dQ_{min} with respect to the B were expressed by a convex function. That is, the dQ ($dQ = Q - Q_{ave}$) depended on the B.

As described in Sect. 4.2.1, even if the B is the same, the BI depends on the video sequence. Therefore, we investigated the relationship between the BI and subjective video quality. When the BI of a video sequence was low, subjective video quality was low, and vice versa, as shown in Fig. 8†. That is, this result indicated that variation of BI affected subjective video quality and that each line is expressed by a linear function.

We also studied the relationship between dBI ($dBI = BI - BI_{ave}$) and dQ. The line in Fig. 9 is one of the lines shown in Fig. 8. The points X, Y, and Z in Fig. 9 are defined as X = (BI_{max}, Q_{max}), Y = (BI, Q), and Z = (BI_{ave}, Q_{ave}). As shown in this figure, when $BI > BI_{ave}$, dQ is proportional to the product of dQ_{max} and $(BI - BI_{ave})/(BI_{max} - BI_{ave})$. On the other hand, when $BI \le BI_{ave}$, dQ is proportional to the product of dQ_{min} and $(BI - BI_{ave})/(BI_{min} - BI_{ave})$††. In addition, the characteristics between the dBI and dQ shown in Fig. 8

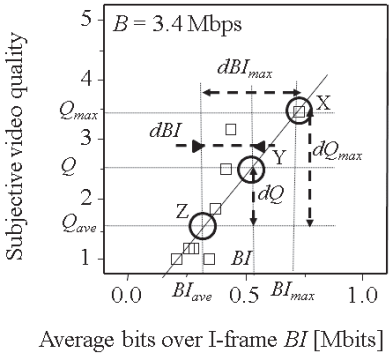


Fig. 9 Relationship between dBI and dQ.

are the same because all the lines are linear. Therefore, dQ is expressed by a linear function as

$$dQ = \begin{cases} a + b \cdot dQ_{max} \cdot F_{max}, & BI > BI_{ave} \\ a + b \cdot dQ_{min} \cdot F_{min}, & BI \le BI_{ave}, \end{cases} \quad (1)$$

$$F_{max} = \frac{BI - BI_{ave}}{BI_{max} - BI_{ave}}, \quad (2)$$

$$F_{min} = \frac{BI - BI_{ave}}{BI_{min} - BI_{ave}}, \quad (3)$$

where a and b are constant and do not depend on the above conditional.

4.2.3 Quality Characteristics Affected by Packet Loss

Next, we studied the dQ affected by packet loss. The VQ affected by packet loss is also affected by compression. Therefore, we introduce normalized video quality, NVQ (N), to remove the impact of the compression on Q. N is the degree of packet-loss degradation. For each B, it is defined as

$$N = \frac{Q(B) - 1}{Q(B)\big|_{PLEF=0} - 1}.$$

Moreover, N_{ave}, N_{max}, and N_{min} are defined as follows. N_{ave} is NVQ averaged over eight video sequences for each experimental condition, N_{max} represents the maximum NVQ in eight video sequences for each experimental condition, and Nmin represents the minimum NVQ in eight video sequences for each experimental condition. The N_{ave}, N_{max}, and N_{min} affected by the D for video group A in Experiment 1 are shown in Fig. 10.

Although the curves of N_{ave}, N_{max}, and N_{min} with respect to D differed, the qualitative tendencies of their curves did not depend on the video sequence (i.e., N_{ave}, N_{max}, and N_{min} decreased as D increased, as shown in Fig. 10). N_{ave}, N_{max}, and N_{min} with respect to D could almost be expressed by an exponential function, while their exponential function's coefficients were different for each codec. There were

†Four figures are used because distinguishing the lines is difficult if there are ten in one figure.

††A figure for $BI \le BI_{ave}$ is omitted because such a figure would be similar to Fig. 9.

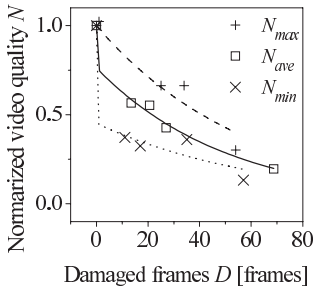


Fig. 10 Damaged frames vs. normalized video quality.

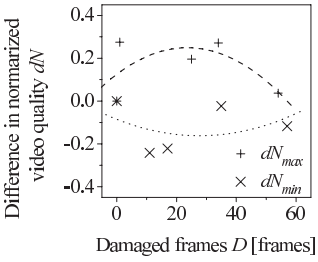


Fig. 11 Damaged frames vs. difference in normalized video quality.

some characteristic variations for each curve as well as the relationship between the B and either Q_{ave}, Q_{max}, or Q_{min}, as shown in Fig. 10.

We investigated the relationship between the D and dN (the difference between N and N_{ave}), where dN_{max} represents the difference between N_{max} and N_{ave}, and dN_{min} represents the difference between N_{min} and N_{ave}. Figure 11 shows the relationship between the D and either dN_{max} or dN_{min}. In Fig. 11, dN_{max} is higher than dN_{min}. This characteristic is the same as that between the B and dQC.

As variation in the BI affected dQC, this variation is also assumed to affect dN. That is, when $BI > BI_{ave}$, dN could be proportional to the product of dN_{max} and $(BI - BI_{ave})/(BI_{max} - BI_{ave})$. On the other hand, when $BI \leq BI_{ave}$, dN is proportional to the product of dN_{min} and $(BI - BI_{ave})/(BI_{min} - BI_{ave})$. As with Eq. (1), dN was expressed by a linear function as

$$dN = \begin{cases} c + d \cdot dN_{max} \cdot F_{max}, & BI > BI_{ave} \\ c + d \cdot dN_{min} \cdot F_{min}, & BI \leq BI_{ave}, \end{cases} \tag{4}$$

where c and d are constant and do not depend on the above conditional.

5. Proposed Model

Here we show mathematical equations of the proposed model derived from the characteristics described in Sect. 4.2. In this section, v_i (i = 1, ...,31) represents the coefficient of the proposed model that are constants optimized for a priori information and that are optimized for training data sets by nonlinear regression analysis.

5.1 Quality-Estimation Module (Compression)

We developed a quality-estimation module for estimating the video quality affected by compression (QC). When $PLEF = 0$, the parameters Q, Q_{ave}, Q_{max}, Q_{min}, and dQ represent QC, QC_{ave}, QC_{max}, QC_{min}, and dQC, respectively.

As mentioned in Sect. 4.2.1, the relationship between B and either BI_{ave}, BI_{max}, or BI_{min} can be modeled using an exponential function:

$$BI_{ave} = v_1 + v_2 \exp\left(-\frac{B}{v_3}\right), \tag{5}$$

$$BI_{max} = v_4 + v_5 \exp\left(-\frac{B}{v_6}\right), \tag{6}$$

$$BI_{min} = v_7 + v_8 \exp\left(-\frac{B}{v_9}\right). \tag{7}$$

As mentioned in Sect. 4.2.2, when $PLEF = 0$, the relationship between B and either QC_{ave}, QC_{max}, or QC_{min} can be modeled using a logistic function:

$$QC_{ave} = 1 + v_{10} - \frac{v_{10}}{1 + (B/v_{11})^{v_{12}}}, \tag{8}$$

$$QC_{max} = 1 + v_{13} - \frac{v_{13}}{1 + (B/v_{14})^{v_{15}}}, \tag{9}$$

$$QC_{min} = 1 + v_{16} - \frac{v_{16}}{1 + (B/v_{17})^{v_{18}}}. \tag{10}$$

As mentioned in Sect. 4.2.2, dQC can be modeled using a linear function:

$$dQC = \begin{cases} v_{19} + v_{20} \cdot dQ_{max} \cdot F_{max}, & BI > BI_{ave} \\ v_{19} + v_{20} \cdot dQ_{min} \cdot F_{min}, & BI \leq BI_{ave}. \end{cases} \tag{11}$$

The video quality affected by compression (QC) is expressed as

$$QC = QC_{ave} + dQC. \tag{12}$$

5.2 Quality-Estimation Module (Packet Loss)

We developed a quality-estimation module for estimating the VQ affected by compression and packet loss.

As mentioned in Sect. 4.2.3, the relationship between D and either N_{ave}, N_{max}, or N_{min} can be modeled using an exponential function:

$$N_{ave} = (1 - v_{21}) \exp\left(-\frac{D}{v_{22}}\right) + v_{21} \exp\left(-\frac{D}{v_{23}}\right), \tag{13}$$

$$N_{max} = (1 - v_{24}) \exp\left(-\frac{D}{v_{25}}\right) + v_{24} \exp\left(-\frac{D}{v_{26}}\right), \tag{14}$$

$$N_{min} = (1 - v_{27}) \exp\left(-\frac{D}{v_{28}}\right) + v_{27} \exp\left(-\frac{D}{v_{29}}\right). \tag{15}$$

The VQ (Q) affected by compression and packet loss can be modeled using QC and N:

$$\begin{aligned} Q &= 1 + (QC - 1) \cdot N \\ &= 1 + (QC_{ave} + dQC - 1) \cdot (N_{ave} + dN) \\ &= 1 + (QC_{ave} - 1) \cdot N_{ave} + dQC \cdot N_{ave} \\ &\quad + (QC_{ave} + dQC - 1) \cdot dN, \end{aligned} \tag{16}$$

where N is expressed as

$$N = \begin{cases} N_{ave} + dN, & D > 0, \\ 1, & D = 0, \end{cases} \tag{17}$$

and dN is expressed as

$$dN = \begin{cases} v_{30} + v_{31} \cdot dN_{max} \cdot F_{max}, & BI > BI_{ave} \\ v_{30} + v_{31} \cdot dN_{min} \cdot F_{min}, & BI \le BI_{ave}. \end{cases} \tag{18}$$

Q can be transformed to

$$Q = Q_{ave} + dQ \tag{19}$$

because Q_{ave} and dQ are expressed as

$$Q_{ave} = 1 + (QC_{ave} - 1) \cdot N_{ave}, \tag{20}$$

$$dQ = dQC \cdot N_{ave} + (QC - 1) \cdot dN. \tag{21}$$

6. Performance Evaluation of Proposed Model

To verify the validity of the proposed model, we first calculated the coefficients (v_1–v_{31}) of the model then applied the model to unknown data. At the end of this section, we note some considerations.

6.1 Performance Requirements

This section describes our target quality-estimation accuracy. It is very important to bring the quality-estimation accuracy of a packet-layer model close to that of a full-reference media-layer model because the end-user QoE is most important for service providers. However, as described in Sect. 3, the packet-layer model does not have systematic access to information about the codec type and codec implementation. In addition, coding parameters cannot be used for QoE estimation due to the encryption. Therefore, to achieve higher quality-estimation accuracy, our proposed model must be trained using a-priori information. According to the HDTV test plan [40], the Video Quality Experts Group (VQEG) decided to adopt the root mean square error (RMSE) as a criterion for verifying the performance of a model. According to the test results of ITU-T Rec. J.341 [15], [41] and ITU-T Rec. J.247 [11], which are for FR media-layer models for HD, VGA, CIF, and QCIF, respectively, the minimum RMSE is about 0.5[†]. The RMSEs were almost the same regardless of the different video formats, video content, and test conditions. Although the peak-signal-to-noise ratio (PSNR) was used as a minimum requirement in VQEG, PSNR is not suitable because it generally does not correlate with subjective video quality. Therefore, we used "RMSE $\le$ 0.5" as the performance requirement to verify that the quality-estimation accuracy of the model was sufficient.

To verify the validity of dQ, we used a comparative model, which is the same as the proposed model without dQ and a combination of conventional models from Refs. [21]–[23], and [29]. The quality-estimation accuracy of the comparative model is greater than that of each conventional packet-layer model in Ref. [21], [22], or [23]. The comparative model is expressed as

$$\begin{aligned} &Q = 1 + (QC_{ave} - 1) \cdot N_{ave}, \\ &QC_{ave} = 1 + v_{10} - \frac{v_{10}}{1 + (B/v_{11})^{v_{12}}}, \\ &N_{ave} = (1 - v_{21}) \exp\left(-\frac{D}{v_{22}}\right) + v_{21} \exp\left(-\frac{D}{v_{23}}\right), \end{aligned} \tag{22}$$

where v_{10}, v_{11}, v_{12}, v_{21}, v_{22}, and v_{23} are the same as the coefficients of the proposed model.

6.2 Quality-Estimation Accuracy of Proposed Model

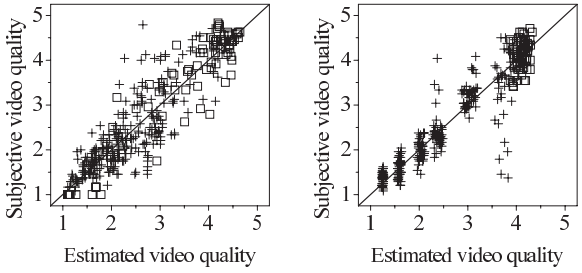
As described in Sect. 3, the proposed model must be optimized for a priori information because, by definition, the model does not have access to the codec implementation. Therefore, we calculated the coefficients (v_1–v_{31}) of the proposed model for video group A in Experiments 1 (Product P1) and 3 (Product P2) based on nonlinear least-squares approximation (NLSA). The comparative model was also optimized using the same training data. The coefficients for video group A in Experiments 1 and 3 are listed in Table 4. We then estimated the subjective video qualities. Since the video-frame-type-estimation model [34] is out of this paper's scope (i.e., the video-frame-type-estimation model is not incorporated into our proposed model), we used the true value of the video-frame type in this work. When information about the boundary of a video frame is lost though packet loss, false-positive video-frame-type detection occurs (i.e., an incorrect DF is calculated)[††]. Therefore, the impact of the false-positive video-frame type on the VQ is included in the results. The relationships between estimated VQ and subjective video quality for training and unknown data are shown in Figs. 12 and 13, respectively. The RMSEs are listed in Table 5, where PM denotes the proposed model and CM denotes the comparative model expressed by equation 22. The improvement rate (IR) of RMSE is also listed. In addition, to analyze the quality-estimation accuracy in detail, the correlation (R) and outlier ratio (OR) are listed in Tables 6 and 7. The RMSE of the unknown data (Experiments 1 and 3 for video group B and Experiments 2 and 4 for video groups A and B) was smaller than that of the training data (Experiments 1 and 3 for video group A) in most of the experiments. It is conceivable that the estimation of the unknown data is slightly easier than that of the

[†]The minimum RMSE listed in Table 7 of Ref. [41] and Tables 1–3 of ITU-T Rec. J.247 [11] was 0.550 for HD, 0.571 for VGA, 0.526 for CIF, and 0.516 for QCIF, respectively.

[††]When information about the boundary of a video frame is lost though packet loss, the false-positive video-frame-type estimation also occurs in the model [34].

Table 4 Coefficients of proposed model in video Group A.

	Experiment 1 (Product P1)	Experiment 3 (Product P2)
v_1	2.921	3.024
v_2	−3.357	−3.021
v_3	12.693	12.323
v_4	2.799	2.669
v_5	−3.730	−3.643
v_6	6.345	3.769
v_7	3.400	2.566
v_8	−3.734	−2.698
v_9	21.894	12.439
v_{10}	3.346	3.327
v_{11}	4.372	0.585
v_{12}	5.817	1.188
v_{13}	3.704	5.336
v_{14}	3.417	0.013
v_{15}	6.414	0.111
v_{16}	2.825	2.779
v_{17}	5.571	1.096
v_{18}	5.726	1.795
v_{19}	0.065	0.015
v_{20}	0.540	0.144
v_{21}	0.804	0.587
v_{22}	2.960	4.163
v_{23}	52.053	63.376
v_{24}	0.760	0.721
v_{25}	3.979	0.018
v_{26}	71.838	58.996
v_{27}	0.750	0.462
v_{28}	0.995	7.031
v_{29}	37.740	51.452
v_{30}	−0.027	−0.009
v_{31}	0.362	−0.029



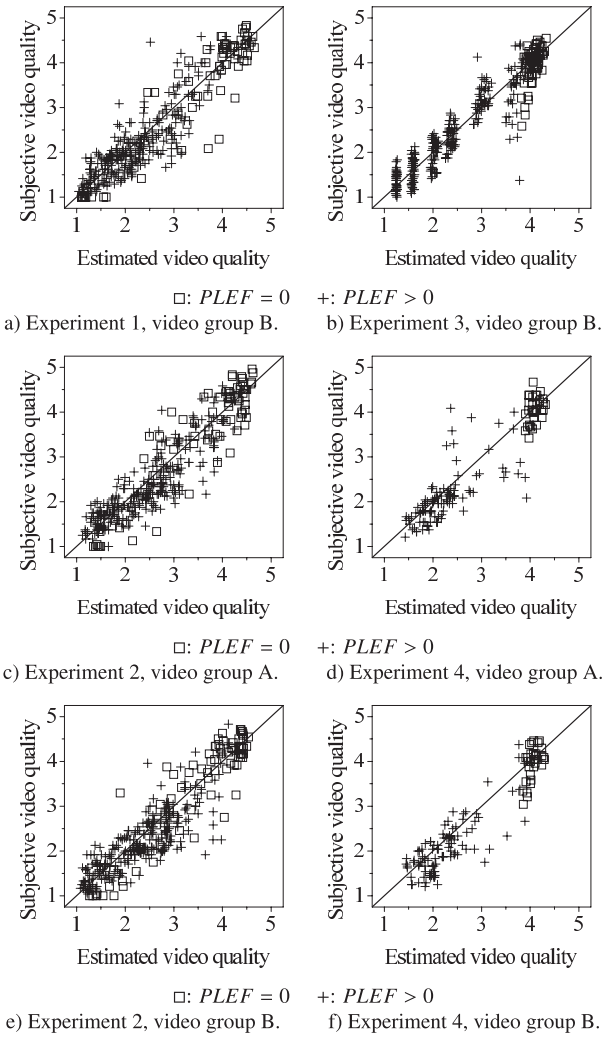
□: $PLEF = 0$ +: $PLEF > 0$

a) Experiment 1, video group A. b) Experiment 3, video group A.

Fig. 12 Quality-estimation accuracy of VQ for training data.

training data.

For the training (Experiments 1 and 3 for video group A) and unknown data (Experiments 1 and 3 for video group B and Experiments 2 and 4 for video groups A and B), the quality-estimation accuracy of our proposed model was equivalent to that of the full reference media-layer models because the RMSEs of the proposed model satisfied the performance requirement described in Sect. 6.1. In addition, compared with the RMSEs, Rs, and ORs of the comparative model, those of the proposed model were almost the same or better. We concluded that the proposed model was able to be optimized for each codec (i.e., Product P1 or



□: $PLEF = 0$ +: $PLEF > 0$

a) Experiment 1, video group B. b) Experiment 3, video group B.

□: $PLEF = 0$ +: $PLEF > 0$

c) Experiment 2, video group A. d) Experiment 4, video group A.

□: $PLEF = 0$ +: $PLEF > 0$

e) Experiment 2, video group B. f) Experiment 4, video group B.

Fig. 13 Quality-estimation accuracy of VQ for unknown data.

Table 5 RMSEs for training and unknown data.

a) Training data.

Exp	VC	VG	CM	PM	IR
1	Product P1	A	0.53	0.49	8%
3	Product P2	A	0.41	0.41	1%
Average			0.47	0.45	4%

b) Unknown data.

Exp	VC	VG	CM	PM	IR
1	Product P1	B	0.56	0.43	22%
3	Product P2	B	0.38	0.36	5%
2	Product P1	A	0.52	0.44	16%
4	Product P2	A	0.47	0.45	3%
2	Product P1	B	0.57	0.45	21%
4	Product P2	B	0.43	0.40	7%
Average			0.49	0.42	12%

Note: Exp denotes Experiment, VC denotes video codec, VG denotes video group, CM denotes comparative model, PM denotes proposed model, and IR denotes improvement rate.

Table 6 Rs for training and unknown data.

a) Training data.

Exp	VC	VG	CM	PM
1	Product P1	A	0.87	0.89
3	Product P2	A	0.93	0.93
Average			0.90	0.91

b) Unknown data.

Exp	VC	VG	CM	PM
1	Product P1	B	0.87	0.91
3	Product P2	B	0.94	0.94
2	Product P1	A	0.88	0.91
4	Product P2	A	0.90	0.90
2	Product P1	B	0.86	0.90
4	Product P2	B	0.92	0.93
Average			0.89	0.91

Note: Abbreviations are the same as Table 5.

Table 7 ORs for training and unknown data.

a) Training data.

Exp	VC	VG	CM	PM
1	Product P1	A	0.59	0.54
3	Product P2	A	0.28	0.29
Average			0.44	0.42

b) Unknown data.

Exp	VC	VG	CM	PM
1	Product P1	B	0.57	0.52
3	Product P2	B	0.32	0.30
2	Product P1	A	0.58	0.54
4	Product P2	A	0.40	0.38
2	Product P1	B	0.60	0.51
4	Product P2	B	0.41	0.41
Average			0.48	0.44

Note: Abbreviations are the same as Table 5.

P2) with high quality-estimation accuracy in Experiments 1 and 3 for video group A, and that the validity of the proposed model for an unknown video sequence was verified in Experiments 1 and 3 for video group B. We also concluded that the proposed model was able to appropriately estimate the VQ for an unknown combination of the BR and random and/or burst packet loss in Experiments 2 and 4 for video group A, and that the validity of the proposed model for an unknown video sequence and combination of the BR and packet loss was verified in Experiments 2 and 4 for video group B. In addition, although the proposed model has 31 coefficients, we concluded that the proposed model is of practical use because there was no deterioration in quality-estimation accuracy for the training and unknown data. Therefore, our model can be applied to the end-user QoE monitoring because the quality-estimation accuracy of our proposed model is equivalent to that of the full reference media-layer models.

6.3 Considerations for Performance of Proposed Model

Some considerations should be noted for the training (Experiments 1 and 3 for video group A) and unknown data (Experiments 1–4 for video group B and Experiments 2 and 4 for video group A).

In the training phase, although there are some scattered plots with quality-estimation accuracy in the lower mid-range in Fig. 12 a), our model was well-trained because even when we used the test conditions described in Experiments 1 and 2 and 16 video sequences (i.e., both video groups A and B) as the training data, the quality-estimation accuracy for the training data did not improve drastically. Therefore, it is conceivable that the numbers of SRCs and PVSs for the training data were adequate and that the quality-estimation accuracy of our proposed model for the training data was sufficient.

For the training and unknown data, the quality-estimation accuracy of the model was low for some scattered plots of the packet-loss condition. The low quality-estimation accuracy can be classified into two main cases. One is the false-positive video-frame-type detection when the boundary of the video frame (i.e., PUSI and PTS) is lost. The other is the significant impacts of the packet loss on the VQ when the packet of a B-frame is lost. Outliers below the line with the 45-degree angle in Figs. 12 and 13 are mainly caused by the lost boundary (i.e., PUSI and PTS) of an I- or P-frame or the significant impact of a lost B-frame on the VQ. On the other hand, outliers above the line with the 45-degree angle in Figs. 12 and 13 are mainly caused by the lost boundary (i.e., PUSI and PTS) of the B-frame.

We first discuss the impact of the false-positive video-frame-type detection on the VQ. When the boundary of a B-frame is lost, this B-frame is detected as an I- or P-frame in most situations because there is an I- or P-frame before a B-frame in most situations. When the boundary of an I- or P-frame is lost, this I- or P-frame is detected as a B-frame in most cases. As a result, the model derives an incorrect DF. In such a case, the proposed model cannot appropriately estimate the VQ because there is a difference in DF between the true and false values. However, when information about the boundary of the video frame is lost through packet loss, by definition, miscalculation of DF is inevitable because the proposed model cannot use bitstream information.

Next, we discuss the significant impacts of the lost B-frame on the VQ. Even when $D = 1$, the packet loss significantly impacted the VQ. In such a case, the proposed model cannot appropriately estimate the VQ. To take into account such degradations, the model needs to take pixel information as input. Note that the occurrence frequency of DF miscalculation was much higher than that of the significant impact of $D = 1$ on the VQ. In these two cases, the comparative model also cannot appropriately estimate the VQ for reasons the same as described above.

Figures 12b) and 13b) show some plots that are lined up vertically. Some reasons for this behavior are as follows. The burst packet-loss length was 1 in Experiment 3, whereas many types of burst packet-loss length were used in Experiments 1, 2, and 4. As a result, one packet was lost for a GoP in Experiment 3, while one or several packets were lost for GoPs in Experiments 1, 2, and 4. In addition, in general, the possibility of a lost I-frame occurring is very high because

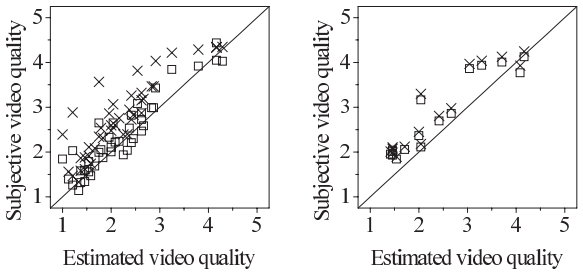


Fig. 14 Quality-estimation accuracy of VQ for SRC12.

the I-frame bits are high. In such a case, for example, when PLEF is 1, 2, 4, or 7, then 1, 2, 4, or 7 I-frames is lost (i.e., when the GoP size is 15, DF is 17, 34, 68, or 119). Therefore, these plots are stripped because the stripped plots have the same number of DF.

We next discuss the improvement rate of the proposed model for two codecs in Experiments 1–4. As described in Sect. 4, we used two different H.264 encoders (Products P1 and P2), which have different implementations. The improvement rate of the proposed model for the H.264 codec (Product P2) in Experiments 3 and 4 was lower than that for the H.264 codec (Product P1) in Experiments 1 and 2. This is because although making use of the BR and DF is sufficient in the quality estimation for the H.264 codec (Product P2), it is more efficient to improve the quality estimation by making use of the ABI in the H.264 codec (Product P1).

Next, we investigate whether the proposed model accurately estimated the VQ per video sequence. Figure 14 shows that the quality estimation accuracy in the proposed model was improved for both Products P1 and P2. Specifically, the RMSEs of our proposed model for SRC 12 in Experiments 2 and 3 were 0.56 and 0.53, whereas the comparative model were 0.64 and 0.55. This trend did not change in the cases of the other video sequences, the miscalculation of DF, or when the lost B-frame had a significant on the video quality. Therefore, we conclude that the proposed model can appropriately estimate the VQ per video sequence using the dQ.

7. Conclusion

To take into account the impact of video content on the VQ, we proposed a new model that calculates the difference in VQ (dQ) between the VQ of the estimation-target video and the average VQ. First, we pointed out several problems of conventional models with regard to the impact of video content on the VQ. To investigate the impact of compression and packet loss on the VQ, we conducted subjective quality assessments. The results showed that the Q_{ave} can be expressed by the B and D, and the VQ per video content (Q) depends on BI. To take these characteristics into account, we introduced the concept of dQ between VQ of the estimation target and Q_{ave} and modeled the dQ using the BI.

We compared the performances of our proposed model and that of the conventional model, which estimates the Q_{ave}, to verify the validity of our model for unknown video sequences and/or combinations of the BR and packet loss. The performance results showed that our proposed model can be used for estimating the VQ per content rather than the Q_{ave} and for estimating the VQ for different video codec implementations by changing the model's coefficients. The quality-estimation accuracy of our model was also shown to be better than that of the conventional model. However, when the boundary of a video frame (i.e., PUSI and PTS) was lost or when there was significant impact on the VQ by the packet loss of a B-frame, the quality-estimation accuracy was low because the proposed model cannot take bitstream and/or pixel information as input.

The following issue calls for further study. The video-frame-type estimation model, which was not used in this work, needs to be incorporated into the proposed model and the validity of the resultant model verified.

References

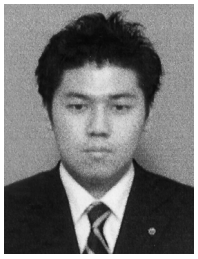
[1] ITU-T Recommendation H.262, "Information technology — Generic coding of moving pictures and associated audio information: Video," Feb. 2000.
[2] ITU-T Recommendation H.264, "Advanced video coding for generic audiovisual services," March 2010.
[3] IPTV Forum Japan IPTVFJ STD0001, "Overview," Sept. 2009.
[4] IPTV Forum Japan IPTVFJ STD0004, "IP broadcasting specifications," July 2010.
[5] DVB Document A001, "Specification for the use of video and audio coding in broadcasting applications based on the MPEG-2 transport stream," Feb. 2007.
[6] ITU-T Recommendation H.720, "Overview of IPTV terminal devices and end systems," Oct. 2008.
[7] ITU-T Recommendation H.721, "IPTV terminal devices: Basic model," March 2009.
[8] ITU-T Recommendation H.222.0, "Information technology — Generic coding of moving pictures and associated audio information: Systems," May 2006.
[9] ITU-T Recommendation G.100/P.10 Amendment 1, "New appendix i — Definition of quality of experience (QoE)," Jan. 2007.
[10] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," March 2004.
[11] ITU-T Recommendation J.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," Aug. 2008.
[12] 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.
[13] 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.
[14] J. Okamoto, K. Watanabe, A. Honda, M. Uchida, and S. Hangai, "HDTV objective video quality assessment method applying fuzzy measure," International Workshop on Quality of Multimedia Experience (QoMEX 2009), pp.168–173, July 2009.
[15] ITU-T Recommendation J.341, "Objective perceptual multimedia video quality measurement of HDTV for digital cable television in the presence of a full reference," Jan. 2011.
[16] P.L. Callet, C. Viard-Gaudin, S. Pechard, and E. Caillault, "No reference and reduced reference video quality metrics for end to end

QoS monitoring," IEICE Trans. Commun., vol.89, no.2, pp.289–296, Feb. 2006.
[17] M. Farias and S. Mitra, "No-reference video quality metric based on artifact measurements," IEEE ICIP 2005, pp.141–144, Sept. 2005.
[18] 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. Lett., vol.12, no.10, pp.685–688, Oct. 2005.
[19] ITU-T Recommendation J.246, "Perceptual audiovisual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference," Aug. 2008.
[20] T. Yamada, Y. Miyamoto, Y. Senda, and M. Serizawa, "Video-quality estimation based on reduced-reference model employing activity-difference," IEICE Trans. Fundamentals, vol.E92-A, no.12, pp.3284–3290, Dec. 2009.
[21] 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, Dec. 2009.
[22] 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.
[23] 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.
[24] F. You, W. Zhang, and J. Xiao, "Packet loss pattern and parametric video quality model for IPTV," IEEE ICIS 2009, pp.824–828, June 2009.
[25] O. Verscheurei and X. Garcia, "User-oriented QoS in packet video delivery," IEEE Netw., vol.12, no.6, pp.12–21, Nov. 1998.
[26] K. Watanabe, K. Yamagishi, J. Okamoto, and A. Takahashi, "Proposal of new QoE assessment approach for quality management of IPTV services," IEEE ICIP 2008, pp.2060–2063, Oct. 2008.
[27] K. Jung, S. Lee, and D. Sim, "Perceptual quality assessment method based on the parameter extraction from H.264/AVC bitstream," Proc. International Workshop on Image Media Quality and its Applications, pp.61–66, Sept. 2008.
[28] A. Silva, P. Rodriguez-Bocca, and G. Rubino, "Optimal quality-of-experience design for a P2P multi-source video streaming," IEEE ICC 2008, pp.22–26, May 2008.
[29] M. Masuda, K. Ushiki, T. Tominaga, T. Hayashi, A. Takahashi, and K. Kawashima, "End-user QoE estimation for video communication services by packet-layer," IEICE Trans. Commun. (Japanese Edition), vol.J94-B, no.1, pp.24–35, Jan. 2011.
[30] O. Sugimoto, S. Naito, S. Sakazawa, and A. Koike, "Objective perceptual video quality measurement method based on hybrid no reference framework," IEEE ICIP 2009, pp.2237–2240, Oct. 2009.
[31] K. Yamagishi, T. Kawano, and T. Hayashi, "Hybrid video-quality-estimation model for IPTV services," IEEE Globecom 2009, Nov. 2009.
[32] A. Khah, L. Sun, and E. Ifeachor, "Content clustering based video quality prediction model for MPEG4 video streaming over wireless networks," IEEE ICC 2009, pp.1–5, June 2009.
[33] M. Ries, C. Crespi, O. Nemethovaand, and M. Rupp, "Content based video quality estimation for H.264/AVC video streaming," IEEE WCNC 2007, pp.2668–2673, March 2007.
[34] K. Ushiki, M. Masuda, T. Hayashi, and A. Takahashi, "Packet-layer video-quality-estimation method by using frame-type estimation," IEICE Trans. Commun. (Japanese Edition), vol.J94-B, no.1, pp.36–48, Jan. 2011.
[35] D. Hands and M. Wilkins, "A study of the impact of network loss and burst size on video streaming quality and acceptability," Proc. 6th International Workshop on Interactive Distributed Multimedia Systems and Telecommunication Services, vol.1718, pp.45–57, Oct. 1999.
[36] 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.
[37] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," IEEE ICC 2008, pp.110–114, May 2008.
[38] ITU-R Recommendation BT.1210.3, "Test materials to be used in subjective assessment," Feb. 2004.
[39] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," April 2008.
[40] VQEG, "Test plan for evaluation of video quality models for use with high definition tv content v3.0," 2009.
[41] VQEG, "Report on the validation of video quality models for high definition video content v2.0," 2010.

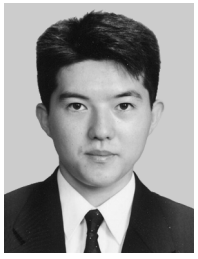
Appendix

Table A·1 Abbreviations.

Abbreviation	Description
ABI	Average bits over I-frame
ACR	Absolute category rating
ARQ	Automatic repeat request
BR	Bit rate
CM	Comparative model
Codec	Encoder and decoder
DF	Number of video frames damaged by packet loss
FR	Full reference
GoP	Group-of-picture
H/E	Head-end
HDTV	High-definition television
IPTV	Internet protocol television
IR	Improvement rate
Mbps	Mbits/sec
MOS	Mean opinion score
MRFM	Multiple-reference-frames mode
NAMS	Non-intrusive parametric model for the assessment of the performance of multimedia streaming
NLSA	Nonlinear least-squares approximation
NR	No reference
NVQ	Normalized video quality
OR	Outlier ratio
PES	Packetized elementary stream
PID	Packet identifier
PLC	Packet-loss concealment
PLEF	Packet-loss-event frequency
PM	Proposed model
PMT	Program map table
PSNR	Peak-signal-to-noise ratio
PTS	Presentation time stamp
PUSI	Payload-unit-start indicator
PVS	Processed video sequence
QoE	Quality of experience
QoS	Quality of service
R	Correlation
RMSE	Root mean square error
RR	Reduced reference
RTP	Real-time transport protocol
SI	Spatial information
SRC	Source video sequence
TCP	Transmission control protocol
TI	Temporal information
TS	Transport stream
UDP	User datagram protocol
VC	Video codec
VG	Video group
VQ	Video quality
VQEG	Video Quality Experts Group



Kazuhisa Yamagishi received his B.E. degree in Electrical Engineering from Tokyo University of Science and M.E. degree in Electronics, Information, and Communication Engineering from Waseda University in Japan in 2001 and 2003. 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. From 2010 to 2011, he was a Visiting Researcher at Arizona State University. He received the Young Investigators' Award (IEICE) in Japan in 2007 and the Telecommunication Advancement Foundation Award in Japan in 2008.



Jun Okamoto received his B.S. and M.S. degrees in Electrical Engineering from the Science University of Tokyo in Japan in 1994 and 1996. He joined NTT Laboratories in 1996 and has been engaged in the quality assessment of visual communication services. Currently, he is studying objective video assessment methods and is leading its standardization in ITU-T SG9 and Video Quality Experts Group (VQEG). He received the Telecommunication Advancement Foundation Award in Japan in 2009.

Takanori Hayashi received his B.E., M.E., and Ph.D. degrees in Engineering from the University of Tsukuba, Ibaraki, in 1988, 1990, and 2007. He joined NTT Laboratories in 1990 and has been engaged in the quality assessment of multimedia telecommunication and network performance measurement methods. Currently, he is the Manager of the Service Assessment Group in NTT Laboratories. He received the Telecommunication Advancement Foundation Award in Japan in 2008.

Akira Takahashi received his B.S. degree in Mathematics from Hokkaido University in Japan in 1988, M.S. degree in Electrical Engineering from the California Institute of Technology in the U.S. in 1993, and Ph.D. degree in Engineering from the University of Tsukuba in Japan in 2007. He joined NTT Laboratories in 1988 and has been engaged in the quality assessment of audio and visual communications. Currently, he is the Manager of the IP Service Network Engineering Group in NTT Laboratories. He is a Vice-chairman of ITU-T SG12. He is a co-Rapporteur of ITU-T Question 13 on Multimedia QoE and its assessment in SG12. He received the Telecommunication Technology Committee Award in Japan in 2004, the ITU-AJ Award in Japan in 2005, the Best Tutorial Paper Award (IEICE Com. Soc.) in 2006, and the Telecommunication Advancement Foundation Award in Japan in 2008 and 2009.