Opinion Model for Estimating Video Quality of Videophone Services

Kazuhisa YAMAGISHI and Takanori HAYASHI
NTT Service Integration Laboratories, NTT Corporation,
3–9–11 Midori–Cho, Musashino–shi,
Tokyo 180–8585, Japan
E-mail: yamagishi.kazuhisa@lab.ntt.co.jp
Telephone: +81–422–59–4397
Fax: +81–422–59–5671

***Abstract*— We propose a computational opinion model for estimating video quality of videophone services. Opinion models for speech such as the E-model have been established and widely used; however, little attention has been given to opinion models for video quality estimation. Our proposed opinion model is useful as a network-planning tool for assessing several video parameters that affect the quality of videophone services. First, we established a function for estimating video quality affected by coding degradation, which expresses the quality of a video affected by coding bit rate and frame rate. Second, we established a packet loss degradation index that estimates the degree of video quality degradation due to packet loss. Finally, we integrated these two functions into the opinion model for estimating video quality. We applied this model for video quality estimation of a videophone service with various video formats and displayed video sizes. The results indicated that the estimation errors of our model were less than the mean of the 99% confidence intervals for the subjective scores. Therefore, our model could be applied to effective design, implementation, and management of videophone applications and communication networks.**

I. Introduction

Advances in broadband IP networks, applications, and terminal technologies have enabled multimedia communication services such as videophones, instant messaging, and teleconferencing that use speech and video to become popular as promising high-quality multimedia applications. Methodologies for evaluating the quality of such services are indispensable because quality is generally not guaranteed in an IP network. Establishing an objective quality assessment method is important for network quality planning and management.

Objective quality assessment [1] can be categorized into media-layer objective models, packet-layer objective models, and the opinion model from the viewpoint of the input information. For estimating the perceptual quality of service (QoS) for users, media-layer objective models use media signals [2], [3], [4], packet-layer objective models use information about IP packet transmission quality [5], and opinion models use network and/or application parameters that affect the quality of media [6], [7], [8], where network parameters mean packet loss rate, delay, and so on and application parameters mean coding bit rate, frame rate, and so on.

Media-layer objective models are highly correlated with subjective quality. However, this approach is inconvenient for network planning purposes because relationships among media quality and quality parameters are not directly considered. Packet-layer objective models are mainly used for in-service quality management and cannot be used for network planning purposes because this model cannot estimate the quality before a service is offered. The opinion model is convenient for network planning purposes because it formulates the relationships among subjective quality, network, and/or application parameters. For example, this model can be used by transmission planners to help ensure that users will be satisfied with end-to-end transmission performance and to avoid building networks that have parameters beyond the required specifications. Network, application, and terminal equipment parameters of high importance to network planners are incorporated into this model.

The opinion model, called the E-model, which has been standardized as ITU-T Recommendation G.107 [6], has been established and widely used for speech services including IP telephony. The E-model can estimate the overall communication quality using a combination of quality factors. The E-model takes 21 parameters as inputs that represent terminal, network, and environmental quality factors. Its output is called R-value, which is a psychological scale that is an index of overall quality. On the other hand, few opinion models for video quality have been studied [9]. In [9], an opinion model for multimedia services was proposed, considering application parameters, but network parameters were not taken into account. In [10], the concept of an opinion model for multimedia services was proposed. This model takes into account individual speech and/or video qualities, absolute delay, and media synchronization. In this paper, we establish a video quality opinion model that estimates video quality affected by coding distortion and packet loss based on the concept presented in [10] because the conventional model cannot estimate video quality degraded by frame reduction and packet loss.

We establish a function for estimating video quality affected by coding distortion and a packet loss degradation index for estimating the degree of video quality degradation due to packet loss from the relationships among video quality and quality parameters. We show that our model accurately

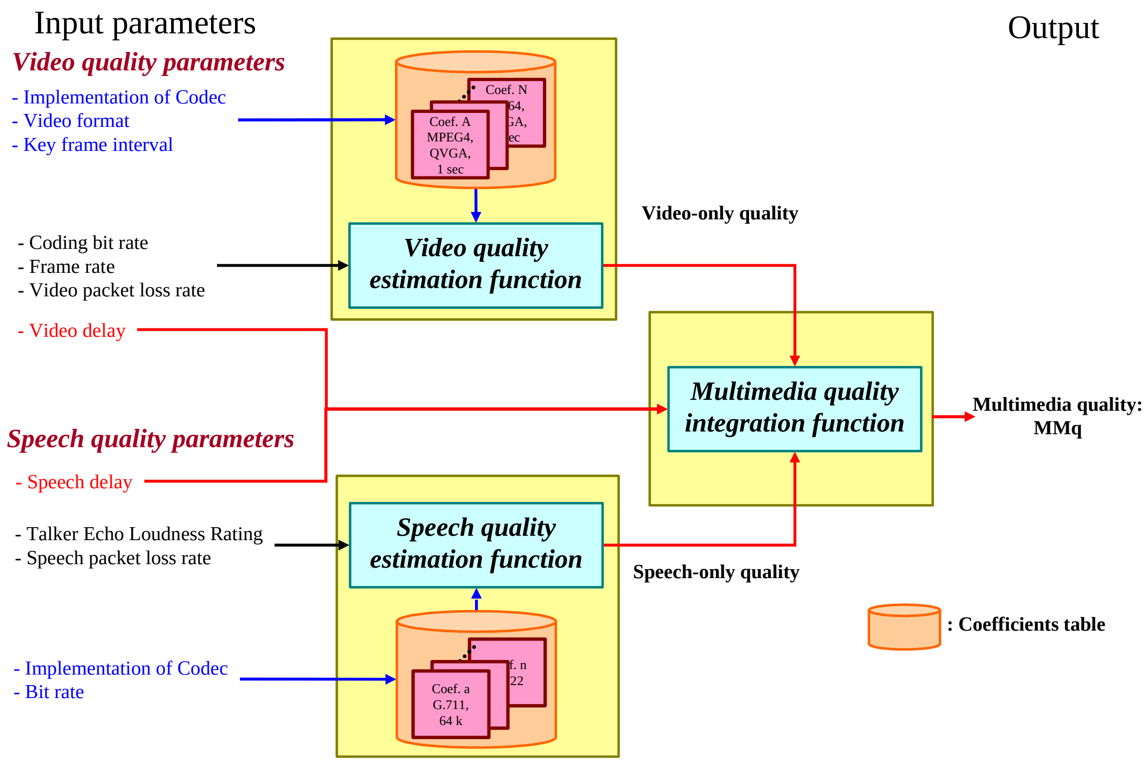

Fig. 1. Opinion model for videophone services [10].

estimates the quality of several applications by changing its coefficients, which are optimized for each application. We can integrate our model with speech quality estimation functions such as the E-model [6] and multimedia quality estimation functions [9], [10], [11], [12], [13].

The remainder of this paper is structured as follows. An opinion model for videophone services is described in Sect. II. A method of subjective quality experiments is described in Sect. III. We propose a video quality estimation model using quality parameters in Sect. IV. Guidelines for quality design are described in Sect. V. Finally, in Sect. VI, we summarize our findings and suggest possible directions for future studies.

II. Framework of Opinion Model for videophone Services

We previously proposed the concept of an opinion model for videophone services [10]. This model can be used as a network-planning tool that estimates videophone quality based on combined effects of variations in several video and speech quality parameters.

The opinion model for videophone services is shown in Fig. 1. Its input parameters are video quality parameters and speech quality parameters. This model contains three functions: a video quality estimation function, a speech quality estimation function, and a multimedia quality integration function. The output of this model is multimedia quality, which is denoted as MMq in Fig. 1. Our initial target is to establish a video quality estimation model, as described in Sect. I.

We consider that the effects of a codec and terminal on subjective video quality are heavily dependent on their implementation. In particular, the quality of a video codec and terminal cannot be estimated simply based on the information about the coding technology and terminal technology. For example, there are a number of different implementations for MPEG-4 codecs due to variation of coding-parameter settings, variation of decoder characteristics, and the number of different types of terminals for PCs, PDAs, and mobile phones. For this reason, we have to conduct subjective quality assessment experiments to calculate the coefficients of the opinion model.

III. Subjective Quality Assessment Experiments

We built a viewing system for deriving subjective quality characteristics that are necessary for constructing the video quality estimation model. We used four video sequences that lasted 10 seconds each. In each sequence, one person explains a completed figure made from multicolor, interlocking construction blocks, as shown in Fig. 2. Video movements in these sequences are larger than those of the head and shoulders that are typical in video sequences of videophones. Therefore, the test sequences were severe in terms of quality assessment. An image whose diagonal measurement was about 4.2 or 8.5 inches was displayed on a 17-inch LCD monitor whose resolution was 1280×1024. The displayed video formats were video graphics array (VGA: 640×480) or quarter video graphics array (QVGA: 320×240).

Fig. 2. Four video sequences.

The experimental parameters were coding bit rate, frame rate, and packet loss rate, as shown in Tables I and II. The video codec was MPEG-4. There were 48 and 49 test conditions, which are the numbers of combinations of parameters in Tables I and II, respectively. We conducted experiments under three video-size conditions, as shown in Table III.

TABLE I
EXPERIMENTAL SETTINGS FOR VGA

Coding bit rate (kbps)	1024, 1280, 1536, 2048, 3072, 4096
Frame rate (fps)	2, 5, 10, 15, 30
Packet loss rate (%)	0.0, 0.1, 0.2, 0.5, 1.0, 2.0

TABLE II
EXPERIMENTAL SETTINGS FOR QVGA

Coding bit rate (kbps)	512, 768, 1024, 1280, 1536, 2048
Frame rate (fps)	2, 5, 10, 15, 30
Packet loss rate (%)	0.0, 0.1, 0.2, 0.5, 1.0, 2.0

TABLE III
RELATIONSHIP BETWEEN VIDEO FORMAT AND DISPLAYED VIDEO SIZE

	Video format	Displayed video size (diagonal)
#1	QVGA	4.2 inches
#2	QVGA	8.5 inches
#3	VGA	8.5 inches

In subjective quality assessment, video quality was evaluated using an ACR (absolute category rating) method [14]. The quality descriptions on the rating scale were given in Japanese.

Thirty two subjects aged 20 - 39 participated in the experiments. They were nonexperts who were not directly concerned with multimedia quality as part of their work, and, therefore, not experienced assessors. The subjects viewed each video sequence at a distance of about 80 cm.

Video quality was represented as an MOS (mean opinion score) averaged over four video sequences.

IV. OPINION MODEL FOR VIDEO QUALITY

First, we analyzed the relationships among video quality, frame rate, coding bit rate, and packet loss rate. Second, we established a function for estimating video quality affected by coding distortion, which estimates the video quality affected by frame rate and coding bit rate. Finally, we determined the packet loss degradation index, which indicates the degree of the video quality degradation due to packet loss.

A. Function for Estimating Video Quality Affected by Coding Distortion

We tried to establish a function for estimating video quality affected by coding distortion from the relationships among video quality, coding bit rate, br, and frame rate, fr. There is an optimal frame rate, ofr, that maximizes the video quality, which is $MOS(fr, br)$, at each coding bit rate. Since $MOS(fr, br)$ has a maximum video quality value, when $fr = ofr$, the function for estimating video quality affected by coding distortion was approximated by a convex function such as the following Gaussian function $G(fr, br)$:

$$MOS(fr, br) = 1 + G(fr, br) \quad (1)$$

and

$$G(fr, br) = \alpha(br) \exp(-\frac{(\ln(fr) - \ln(ofr(br)))^2}{2\omega^2(br)}). \quad (2)$$

When $fr = ofr$, $G(fr, br) = \alpha(br)$; then, $1 + \alpha(ofr, br)$ indicates the optimal video quality $MOS(ofr, br)$ for each br, and $\omega(br)$ indicates the degree of video quality degradation due to the frame rate.

Then, the optimal frame rate increases monotonically as coding bit rate increases and saturates at the maximum frame rate, as follows:

$$ofr(br) = a + b\ br, \quad (3)$$

where $ofr(br)$ clipped at 30, and a and b are constants.

After that, when $fr = ofr$, we found that $MOS(ofr, br)$ increases as br increases and saturates at the maximum MOS. Using these results, $\alpha(br)$ was approximated by a logistic function, as follows:

$$\alpha(br) = c - \frac{c}{1 + (\frac{br}{d})^e}, \quad (4)$$

where c, d, and e are constants.

Finally, $\omega(br)$ increases as br increases, as shown in Eq. 5:

$$\omega(br) = f + g\ br, \quad (5)$$

where f and g are constants.

Coefficients a, b, c, d, e, f, and g are determined from the subjective data obtained for each codec implementation.

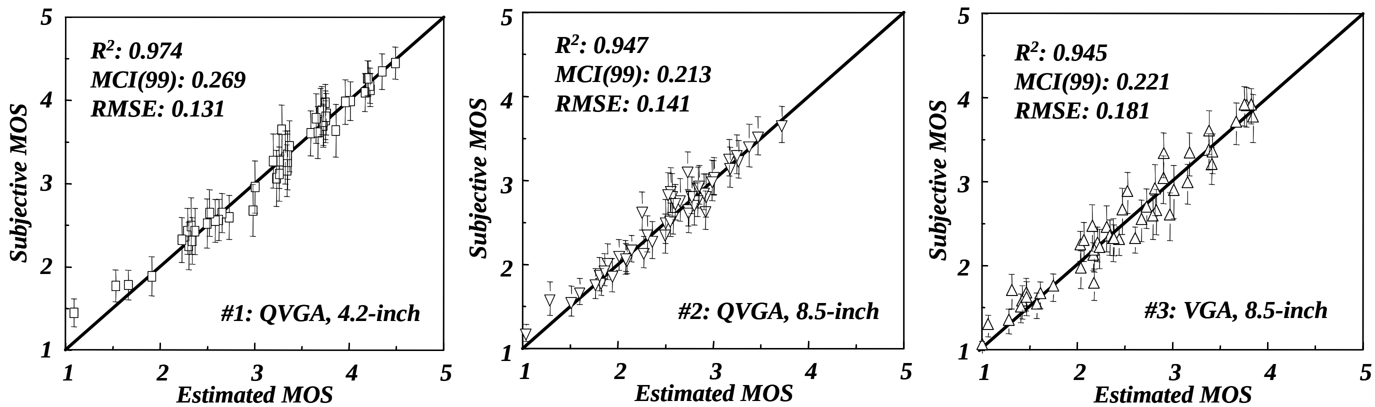

Fig. 3. Relationship between subjective MOS and estimated MOS.

B. Packet Loss Degradation Index

We tried to establish the packet loss degradation index, which indicates the degree of video quality degradation due to packet loss. $MOS(fr, br, pl)$ decreased exponentially as the packet loss rate, pl, increased. That is, $MOS(fr, br, pl)$ was approximated by an exponential function as follows:

$$MOS(fr, br, pl) = 1 + G(fr, br)\exp(-\frac{pl}{\tau(fr, br)}), \quad (6)$$

where $1 + G(fr, br)$ indicates the video quality when the packet loss rate is 0, and $\tau(fr, br)$ indicates the degree of video quality degradation due to the packet loss rate.

When the coding bit rate is a constant, we define $\tau_1(fr)$ as follows:

$$\tau_1(fr) \overset{def}{\Longleftrightarrow} \tau(fr, br)\Big|_{br=const}.$$

When the frame rate is a constant, we define $\tau_2(br)$ as follows:

$$\tau_2(br) \overset{def}{\Longleftrightarrow} \tau(fr, br)\Big|_{fr=const}.$$

$\tau_1(fr)$ and $\tau_2(br)$ increase monotonically as the frame rate or coding bit rate decreases, as follows:

$$\tau_1(fr) = w + x \ \exp(-fr/j) and \quad (7)$$

$$\tau_2(br) = y + z \ \exp(-br/l), \quad (8)$$

where $\tau_1(fr)$ indicates the degree of video quality degradation due to packet loss when br is constant, $\tau_2(br)$ indicates that when fr is constant, and w, x, j, y, z, and l are constants.

We integrated Eqs. 7 and 8 using regression analysis, and obtained:

$$\tau(fr, br) = h + i \ \exp(-fr/j) + k \ \exp(-br/l), \quad (9)$$

where $\exp(-fr/j)$ and $\exp(-br/l)$ are independent variables, and $\tau(fr, br)$ is a dependent variable.

Coefficients h, i, j, k, and l are determined from the subjective data obtained for each codec implementation.

C. Performance of Opinion Model

First, we calculated the coefficient tables of the opinion model for applications [#1, #2, and #3], as shown in Table III. Then, using these coefficients, we estimated the respective subjective video qualities. These results are shown in Fig. 3. Multiple correlation coefficients (R^2), the mean of the 99% confidence interval (MCI) for the subjective MOS, and the root mean square error (RMSE) are also shown in Fig. 3.

We used $R^2 \geq 0.9$ and RMSE $\leq$ MCI as criteria to determine whether the accuracy of the model is sufficient. In every experiment, the R^2 value was larger than 0.9, and the RMSE was less than the MCI. Therefore, we concluded that video coding quality can be formulated by Eqs. 1 - 9. Moreover, we verified that our model is applicable not only to the MPEG-4 codec but also to the H.264 codec [15]. Therefore, we concluded that our proposed model could be applied to effective design, implementation, and management of videophone services using these codecs.

V. Quality design using our proposed model

Using our proposed opinion model, we can give some guidelines for quality design and management of videophone services. Designing an application and/or network to improve video quality for various video sequences is extremely important to avoid creating an over-engineered network.

First, we show an example of application planning. In general, a network designer can use only a limited bandwidth. For example, when bandwidth is restricted to 0.5 or 2.0 Mbps, our proposed opinion model enables us to find the optimal frame rate, which is about 15 or 30 fps, respectively, as shown in Fig. 4. When the packet loss rate is 0.5%, we need to find the optimal coding bit rate because the degree of degradation due to packet loss at each coding bit rate is different. That is, we can find the optimal coding bit rate, which is about 0.5 Mbps, as shown in Fig. 4.

Next, we explain an example of a network guideline. To satisfy the requirement that $MOS \geq 3.0$ when the bandwidth is 0.5 or 2.0 Mbps, the thresholds for the packet loss rate must be less than 0.50 or 0.27%, respectively, as shown in Fig. 4.

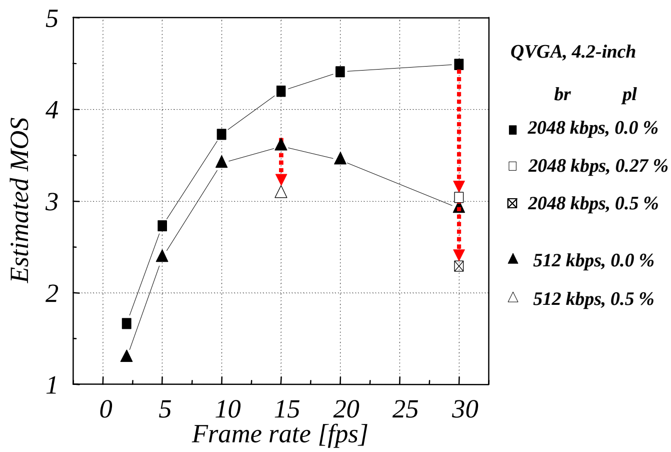

Fig. 4. Estimated MOS vs. frame rate used for quality design.

Based on these guidelines, we can design the application and/or network appropriately.

VI. Conclusion

We defined the video quality estimation function for codecs and the packet loss degradation index from the relationships among perceptual video quality and quality parameters. The former function can be used to estimate the optimal video quality at each coding bit rate. The latter can be used to estimate the degree of video quality degradation due to packet loss. By using these functions, we estimated video quality based on video quality parameters because the RMSEs of the estimated quality [#1, #2, and #3] obtained by our proposed model were equivalent to the statistical accuracy of the subjective quality evaluation.

We described an example of quality design guidelines used for videophone services by using the proposed model. We believe that our study provides a basis for an opinion model for estimating video quality of videophone services.

The following issues call for further study. We need to apply our model to the multimedia quality integration model [10] and verify that the model will have sufficient accuracy. Moreover, we will extend our model to estimate the quality of various video sequences for video streaming services including IPTV.

References

[1] A. Takahashi, H. Yoshino, and N. Kitawaki, “Perceptual QoS assessment technologies for VoIP,” IEEE Communications Magazine, pp. 28-34, July 2004.

[2] ITU-T Recommendation P.862, “Perceptual Evaluation of Speech Quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs,” Feb. 2001.

[3] VQEG. Video Quality Experts Group website, http://www.vqeg.org/.

[4] ITU-T Recommendation J.148, “Requirements for an objective perceptual multimedia quality model,” 2003.

[5] A. Clark, “Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality,” IP Telephony Workshop. 2001.

[6] ITU-T Recommendation G.107, “The E-model, a computational model for use in transmission planning,” Mar. 2005.

[7] K. Yamagishi and T. Hayashi, “Analysis of psychological factors for quality assessment of interactive multimodal service,” Electronic Imaging 2005, pp. 130-138, Jan. 2005.

[8] K. Yamagishi and T. Hayashi, “Opinion Model using Psychological Factors for Interactive Multimodal Services,” IEICE Trans. Commun. Vol. E89-B, No. 2, pp. 281-288, Feb. 2006.

[9] N. Kitawaki, Y. Arayama, and T. Yamada, “Multimedia opinion model based on media interaction of audio-visual communications,” MESAQIN2005, pp. 5-10, June 2005.

[10] T. Hayashi, K. Yamagishi, and H. Yoshino, “Perceptual QoS Evaluation Model for Audiovisual Communication Services,” World Telecommunications Congress 2006 (WTC’ 06), May 2006.

[11] C. Jones and D.J. Atkinson, “Development of Opinion-Based Audiovisual Quality Models for Desktop Video-Teleconferencing,” 6th IEEE International Workshop on Quality of Service, May 1998.

[12] T. Hayashi, “State of the Art of Multimedia Quality Assessment Methods,” ITU-T Workshop on Video and Image Coding and Applications (VICA), July 2005.

[13] D. S. Hands, “A Basic Multimedia Quality Model,” IEEE TRANS. ON MULTIMEDIA, VOL. 6, NO. 6, pp. 806-812, Dec. 2004.

[14] ITU-T Recommendation P.910, “Subjective video quality assessment methods for multimedia applications,” Sep. 1999.

[15] K. Yamagishi and T. Hayashi, “Verification of Video Quality Opinion Model for Videophone Services,” ISCA workshop on Perceptual Quality of Systems, Sep. 2006.