The Third International Workshop on Image Media Quality and its Applications, IMQA2008
September 5-6, 2008, Kyoto, Japan

MULTIMEDIA QUALITY ASSESSMENT METHOD FOR IPTV SERVICES

Toshiko Tominaga, Kazuhisa Yamagishi, and Takanori Hayashi

NTT Service Integration Laboratories, NTT Corporation

ABSTRACT

This paper proposes a method for constructing a multimedia quality estimation model for IPTV services. The model is expressed by audio quality, video quality, and the quality of their multiplicative interaction term. To make the model, we have to obtain those three qualities. A subjective assessment method for multimedia quality is described in the ITU-T recommendation P.920 appendix. However, an adequate subjective multimedia quality assessment method based on the model's structure has not been discussed thoroughly. Therefore, we verified the appropriateness of such methods, including the method described in P.920, for use in our proposed model by performing two assessment tests. We then performed multiple regression analysis using the more appropriate method's data, constructed the multimedia quality estimation model, and showed its accuracy.

1. INTRODUCTION

IPTV services are being developed and provided with the advance of broadband IP technology. To provide these services with appropriate quality, the in-service quality management of IPTV services based on quality of experience (QoE) is important. Therefore, QoE assessment methods that exhibit good accuracy and efficiency are needed.

Subjective quality assessment is the most fundamental method of evaluating QoE. In this method, subjects evaluate multimedia quality based on psychological evaluation criteria. However, because subjective testing is time-consuming and expensive, objective methods to predict QoE are necessary.

At present, perceptual QoE evaluation models for the objective measurement of speech/audio quality are at a reasonably advanced stage. ITU-T Rec. P.862 [1] describes PESQ (perceptual evaluation of speech quality), which is an objective speech quality assessment method for narrowband telephones. ITU-R Rec. BS.1387 [2] describes PEAQ (perceptual evaluation of audio quality), which is an objective audio quality assessment method. On the other hand, for video quality evaluation, ITU-T Rec. J.144 [3] describes objective full-reference video quality metrics for digital TV applications without error impairments such as bit errors and dropped packets. On the other hand, draft ITU-T Rec. J.247 describes perceptual video quality measurements for multimedia applications with error impairments.

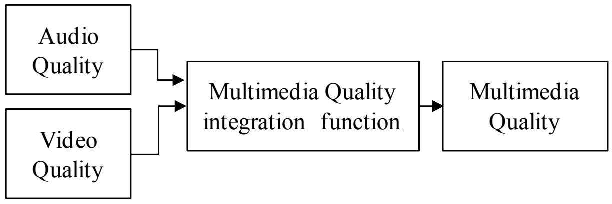


Fig. 1 Multimedia quality estimation model

QoE evaluation model can be applied to quality design and management of multimedia services in conjunction with proper objective quality assessment methods for audio and video.

Multimedia quality estimation model's framework is described in ITU-T Rec. J.148 [4] and discussed in Refs. [5]–[9]. The multimedia quality estimation model we propose is shown in Fig. 1. The model considers audio quality, video quality, and the quality of the multiplicative interaction term between video and audio. In the ITU-T Rec. P.920 appendix [10], subjects assess the three qualities at the same time. In our proposed model, however, individual audio and video quality is needed. It is not clear that the subjective assessment method is appropriate to our proposed model.

Therefore, we performed two subjective assessment tests to verify the subjective assessment methods for evaluating the three qualities at the same time and separately. After selecting the more appropriate subjective test method, we applied multiple regression analysis and constructed the multimedia quality estimation model.

The remainder of this paper is structured as follows. In Section 2, we present two subjective assessment tests. In each test, a different assessment method was performed on HDTV video streaming services. The test result shows the most appropriate method for the multimedia quality estimation model. In Section 3, we construct the model from the selected method's data. We then use regression analysis to clarify the multimedia quality estimation model and showed its accuracy. In Section 4, we conclude this paper.

2. SUBJECTIVE ASSESSMENT TEST

In this section, we clarify the influence of the three qualities using two subjective assessment methods and verify an appropriate subjective assessment method for the proposed multimedia quality estimation model.

2.1. Test Conditions

2.1.1. Test Methodology

The subjective assessment tests were executed by using evaluation methods A and B. In method A, subjects evaluate three qualities, which are video, audio, and overall multimedia quality, at the same time in one evaluation session (evaluation method described in ITU-T Rec. P.920 Appendix). In method B, subjects evaluate only one quality in each session. Characteristics of the two evaluation methods are as follows.

- Evaluation method A evaluates the three qualities at the same time. Collecting three qualities' evaluation values is a merit from the viewpoint of efficient examination. However, subjects need to memorize each quality accurately, and that causes a load on the subjects.
- Evaluation method B evaluates each quality individually in each session. In the audio (video) session, subjects assess audio (video) quality in the environment of the media except multimedia quality assessment session. A subject's load is lighter than that of method A because only one quality is evaluated. However, more time is needed because three individual quality evaluation sessions are needed.

2.1.2. Test Environment

We constructed an experimental HDTV-based IPTV service system. The video codec was H.264 and the audio codec was MPEG-2 AAC. A 32-inch CRT monitor and an audio speaker were used as multimedia quality evaluation equipment. Subjective assessment conditions and codec conditions are shown in Tables 1 and 2.

2.1.3. Test Material

The test material was eight different video sequences (10 seconds each) consisting of 6 sequences chosen by ITU-R Rec. BT.1210 and 2 TV conference video sequences. Audio sequences of background music were attached to 4 of the ITU-R Rec. BT.1210 video sequences, sound quality assessment material by European Broadcasting Union (EBU) was attached to 2, and the TV conference video sequences had voice only [11],[12].

The experimental quality parameters were video bit rate and packet loss rate. From the matrix of the above conditions and 8 video sequences, we chose 24 test conditions for each quality evaluation. Quality parameter conditions are shown in Table 3.

In the subjective quality assessment, audio quality, video quality, and overall multimedia quality were evaluated using a degradation category rating (DCR) method based on ITU-T Rec. P.911 [13]. The DCR method stipulates that the test sequences be presented in pairs: the first stimulus presented in each pair is always the source reference, while the second stimulus is the same source presented through one of the systems being tested. The subjects are asked to rate the impairment of the second stimulus in relation to the reference by using a five-level impairment scale. The quality descriptions for individual rating categories were given in Japanese.

Table 1 Subjective assessment conditions

Subjects	24 (12 males and 12 females)
Viewing distance	3H (H: picture height)
Video monitor	SONY BVM-D32E1WJ
Illumination intensity	About 30 lx
Listening equipment	ONKYO GX-77M stereo speaker
Ambient noise at receiving side	Hoth noise at 35 dB (A)
Evaluation rating	Degradation category rating (DCR)

Table 2 Codec conditions

Video codec	H.264 GOP: M=3, N=15 Framerate: 30 fps Resolution: 1440*1080
Audio codec	MPEG-2 audio AAC-LC

Table 3 Quality parameter conditions

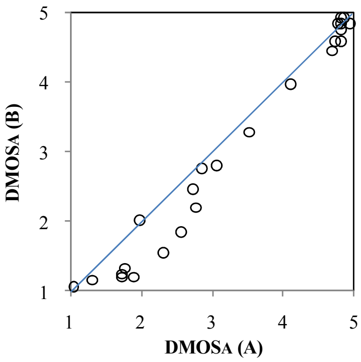
Number of sequences		8
Sequence length		10 s
Video	Bandwidth	14, 10, 6 Mbps
	Random packet loss ratio (%)	14 Mbps : 0, 0.01, 0.03, 0.07 10 Mbps:0, 0.015, 0.04,0 0.100 6 Mbps:0, 0.02, 0.08, 0.16
Audio	Bandwidth	192 kbps
	Random packet loss ratio (%)	0, 0.1, 0.2, 0.6

2.1.4. Subjects

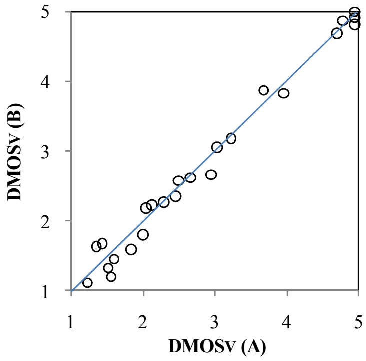
Twenty-four subjects aged 20–29 participated in the experiments. They were non-experts who were not concerned with multimedia quality as part of their work and, therefore, not experienced assessors.

2.2. Test results and discussion

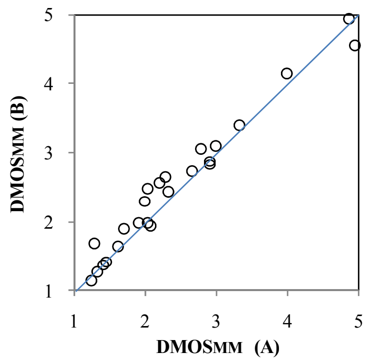
To see the difference between the quality characteristics evaluated by each method, the degradation mean opinion score (DMOS) characteristics of the two methods for each quality are shown in Fig. 2. If the same characteristics are found for the three qualities, the values for DMOS (A) and DMOS (B) would be almost



(a) Audio quality



(b) Video quality



(c) Multimedia quality

Fig. 2 DMOS characteristics for methods A and B

the same. However, there is a difference between audio quality and the other two qualities. For audio quality, the DMOSs of method A tended to have higher scores than that of method B (Fig2 (a)).

Significant difference analysis of the three qualities that had been obtained by the experiments was conducted. The results showed no significant differences in both multimedia and video quality evaluations by the two methods. However, there was a significant difference in evaluation of audio quality between the two methods with a 5% level of significance. Method A obtained higher audio quality.

This result suggested that when the subjects evaluate audio quality in an environment of audio and video, audio quality is influenced by video quality. This indicates that evaluation method A cannot be applied to obtain appropriate individual media quality values. We conclude that method B is a more appropriate subjective assessment method for the proposed multimedia quality estimation model.

3. APPLICATION OF EVALUATION METHOD TO MULTIMEDIA QUALITY ESTIMATION MODEL

We performed subjective assessment method B, then, conducted multiple regression analysis on the resultant data to obtain a multimedia quality estimation model for HDTV-based IPTV services.

3.1. Experimental conditions for constructing model

To construct the multimedia quality estimation model, a subjective evaluation experiment using method B was executed. The same experimental environment as that of the above experiments was used (Tables 1-3). From the matrix of the quality parameter conditions, we chose each 40 test conditions for audio and video quality, and 136 test conditions for multimedia quality.

3.2. Multimedia quality estimation model

Multimedia quality consists of audio quality and video quality (Fig. 1). A multimedia quality integration function also considers the interaction term between audio and video. A basic multimedia quality integration function is given by the following equation.

$$DMOS_{MM} = \alpha\, DMOS_A + \beta\, DMOS_V + \gamma\, DMOS_A * DMOS_V + \delta, \quad (1)$$

where $DMOS_A$ means audio quality by mean opinion score of degradation rating scale, $DMOS_V$ means video quality, $DMOS_A*DMOS_V$ is the product of audio quality and video quality, and α–δ are constants.

3.3. Regression analysis result

Multiple regression analysis is performed for combinations of the parameters of Eq. (1). We checked the effectiveness of multiple regression analysis, that is, the p-values for the equation and each parameter are less than 0.01, and found no multicollinearity among the parameters. Regression analysis results summary are shown in Table 4.

From these results, the multimedia quality estimation model was expressed by using two parameters, video quality and the multiplicative interaction term between video and audio. The correlation coefficient (r) between subjective quality and its objective estimation by this model was 0.96, the root mean square error (RMSE) was 0.26, and the mean of the 95% confidence interval (MCI) for the subjective DMOS was 0.28. Because the evaluation error of the multimedia quality estimation model was less than the statistical ambiguity of the subjective score (i.e., RMSE < MCI), we concluded that multimedia quality can be evaluated by this model.

Table 4 Regression analysis results

Objective parameters	Correlation coefficient (r)	Effectiveness of multiple regression analysis	RMSE
① DMOSA	0.508	-	0.846
② DMOSV	0.752	-	0.647
③ DMOSA*DMOSV	0.951	-	0.302
④ DMOSA, DMOSV	0.915	×	0.392
⑤ DMOSA, DMOSA*DMOSV	0.963	×	0.262
⑥ DMOSV, DMOSA*DMOSV	0.963	○	0.261
⑦ DMOSA, DMOSV, DMOSA*DMOSV	0.964	×	0.257

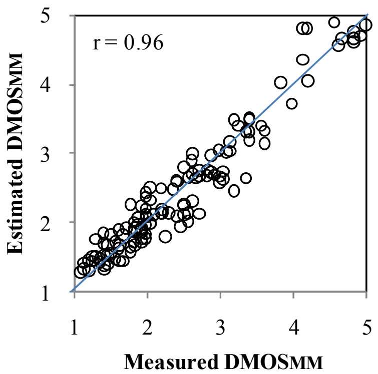


Fig. 3 Relationship between measured and estimated multimedia quality

The multimedia quality estimation model is given by the following equation.

$$DMOS_{MM} = 0.16\, DMOS_V + 0.13\, DMOS_A * DMOS_V + 0.91. \quad (2)$$

The model's form indicates that video quality is given significantly higher weighting than audio quality in the case of HDTV-based IPTV services. The relationship between measured and estimated multimedia quality is shown in Fig. 3. Good correlation coefficient (r=0.96) was achieved between measured and estimated multimedia quality. Hence, we concluded that the proposed method works for estimating multimedia quality for IPTV services.

4. CONCLUSION

We proposed a method for constructing a multimedia quality estimation model for IPTV services. This model consists of audio quality, video quality, and their multiplicative interaction term. In subjective quality assessment tests, we found that the method of evaluating the individual quality separately was adequate to measure audio and video quality. We constructed the multimedia quality estimation model by multiple regression analysis and also showed the model's estimation accuracy.

This model can estimate overall multimedia quality based on individual audio and video qualities, so it can be applied to quality design and management of multimedia services in conjunction with a proper objective quality assessment method for audio and video (such as ITU-T Rec. J.144 and ITU-R Rec. BS.1387).

ACKNOWLEDGMENT

This research was partly supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications, Japan under project 073103002.

REFERENCES

[1] ITU-T Rec. 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," (2001).

[2] ITU-R Rec. BS.1387, "Method for objective measurements of perceived audio quality," (2001).

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

[4] ITU-T Rec. J.148, "Requirements for an objective perceptual multimedia quality model," (2003).

[5] C. Jones and D. J. Atkinson, "Development of opinion based audiovisual quality model for desktop video teleconferencing," IWQoS '98, pp. 196–203, May 1998.

[6] A. Rimell and A. Owen, "The effect of focused attention on audio-visual quality perception with applications in multi-modal codec design," IEEE ICASSP 2000, vol. 6, 5–9, pp. 2377–2380, June 2000.

[7] D. S. Hands, "A basic multimedia quality model," IEEE Trans. Multimed., vol. 6, no. 6, pp. 806–816, Dec. 2004.

[8] N. Kitawaki, Y. Arayama, and T. Yamada, "Multimedia opinion model based on media interaction of audio-visual communications," MESAQIN 2005, pp. 5–10, June 2005.

[9] S. Winker and C. Faller, "Perceived audiovisual quality of low-bitrate multimedia content," IEEE Trans. Multimed., vol. 8, no. 5, pp. 973–980, Oct. 2006.

[10] ITU-T Rec. P.920, "Interactive test methods for audiovisual communications," May 2000.

[11] ITU-R Rec. BT.1210, "Test materials to be used in subjective assessment," Oct. 1995.

[12] European Broadcasting Union, "SQAM (sound quality assessment material) – recording for subjective tests," April 1988.

[13] ITU-T Rec. P.911, "Subjective audiovisual quality assessment methods for multimedia applications," Dec. 1998.