MULTIMEDIA OPINION MODEL BASED ON MEDIA INTERACTION OF AUDIO-VISUAL COMMUNICATIONS

Nobuhiko Kitawaki, Yusuke Arayama, and Takeshi Yamada

University of Tsukuba,
1-1-1, Tennoudai, Tsukuba-shi, 305-8573 Japan.
e-mail: Kitawaki@cs.tsukuba.ac.jp

URL: http://www.mmlab.cs.tsukuba.ac.jp/~kitawaki

Abstract

Recently, a number of PC-based applications have been developed for videoconferencing and videophone as well as mobile videophone services. In order to provide a means for designing, evaluating, and monitoring multimedia services such as teleconferencing and videophone applications, it is quite important to develop an opinion model which is applicable to the evaluation of such multimedia communication services. This paper describes a multimedia opinion model based on an objective quality assessment for audio-visual communications, taking into account the mutual interaction of audio and video information. This paper also describes an application of the proposed method to estimate a multimedia quality of videophone and PDA as well as mobile videophone services.

Keywords

opinion model, audio-visual communication, media interaction, objective quality assessment

1. INTRODUCTION

A number of PC-based applications have been developed for videoconferencing and videophone as well as mobile videophone services. It is likely that such applications will become the major telecommunications services in the near future since the access network is evolving quite rapidly.

For speech communications, ITU-T Recommendation G.107 "The E-model, a computational model for use in transmission planning" has been widely used as a network planning tool for IP-telephony services, as well as for conventional PSTN services [1]. Its extension is now being studied in ITU-T from the viewpoints of wideband-speech, and multimedia.

In order to provide a means for designing, evaluating, and monitoring multimedia services such as teleconferencing, videophone, and CSCW (Computer Supported Cooperative Work) applications, it is quite important to develop an opinion model which is applicable to the evaluation of such multimedia communication services.

This paper describes an objective perceptual multimedia quality model for audio-visual communications, taking account of the mutual interaction of audio and video information.

This paper also describes an application of the proposed method to estimate a multimedia quality of videophone and PDA as well as mobile videophone services.

2. CONCEPT OF OPINION MODEL

Multimedia is defined as the combination of multiple forms of media such as audio, video, text, graphics, fax, and telephony in the communication of information. The initial goal of the work described in this paper is to produce an objective measurement of quality for audio-visual communications. The primarily use of the model is to measure the quality of limited bandwidth services.

It is considered that limited bandwidth represents a more critical level of service from the viewpoints of multimedia quality degradations.

Second, the proposed multimedia opinion model is applied to estimate multimedia quality of videophone services with limited bandwidth.

2.1 Previous Multimedia Opinion Model

(Method 1)

Perceptual models for the objective measurement of audio quality and video quality have been studied [2]-[4]. In these studies, multimedia quality (MOS_{AV}) is estimated from separate objective measurements of audio quality (Aq) and of video quality (Vq), as shown in Fig.1. Here, Aq is defined as independent audio quality without video interaction (namely, audio alone), and Vq is vice versa (namely, video alone). In this paper, those are called “media independent quality.”

The multimedia integration function is shown in Eq.(1). The multimedia integration function aims to accommodate human perceptual and cognitive processes active in the formation of quality judgments of audio-visual services.

Multimedia Integration Function (Method 1)

$$MOSAV= \alpha Vq+ \beta Aq+ \gamma Vq\, Aq+ \delta \quad (1)$$

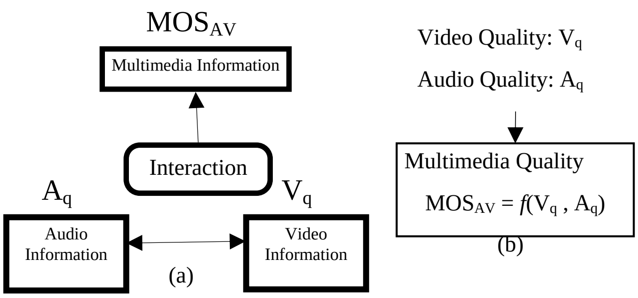

Fig.1 Previous Multimedia Opinion Model (Method 1).

2.2 Proposed Multimedia Opinion Model

(Method 2)

We focus on the mutual interaction between audio and video information. That is, perceived multimedia quality depends on the mutual interaction of audio and video information, and can not be adequately assessed from independent audio and video information.

This paper proposes a new multimedia opinion model, shown in Fig.2, in which multimedia quality (MOS_{AV}) is estimated from Aq(Vq) and Vq(Aq). Here, Aq(Vq) implies an objective measure of audio quality taking account of the influence of video quality, and Vq(Aq) implies a similar measure of video quality taking account of the influence of audio quality. Namely, Aq(Vq) and Vq(Aq) are called as “mutual interaction quality” in this paper.

The multimedia quality is estimated by taking the following two steps:

Step 1;

each objective measurement of mutual interaction quality Aq(Vq) and Vq(Aq) is estimated from subjectively evaluated media independent quality Aq and Vq.

Step 2;

objective measurement of the multimedia integration function (MOS_{AV}) is calculated from objectively estimated mutual interaction quality Aq(Vq) and Vq(Aq).

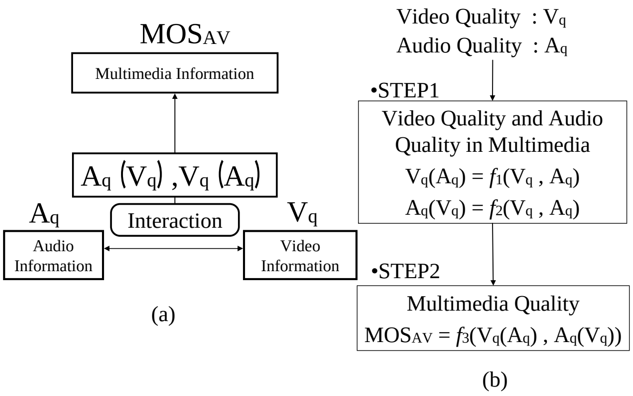

Fig.2 Proposed Multimedia Opinion Model (Method 2)

3. SUBJECTIVE EXPERIMENT

The proposed multimedia opinion model (Method 2) has been verified from the viewpoints of the consistency between the subjectively evaluated MOS and the objectively estimated MOS for audio-visual communications, and a comparison of the estimation accuracy with that of the previous multimedia opinion model (Method 1).

3.1 Experimental Condition in Subjective Test

Table 1 shows the experimental conditions for the subjective assessment of audio-visual communication which was performed in accordance with ITU-T Rec. P.911. Headphones and a PC-display were used for the listening and viewing tests. The Absolute Category Rating (ACR) method is a category judgment where the test sequences are presented one at a time and are rated independently on a five-level category quality-scale. The presentation time of the audio-visual sample was 20s which is longer than the nominal value of the recommendation. The number of subjects in each viewing and listening test was 21. Source signals and codecs were used to provide streaming services of TV programs and Video programs in the limited bandwidth network services.

Table 1 Experimental condition for the subjective tests.

Assessment regulation	ITU-T Rec. P.911
Test method	Five-level ACR
Number of subjects	21
Source signals	#1 report, #2 news, #3 scene + BGM
Video codec	MPEG-4 128, 320, 512kb/s, reference Size 320x240 pixels, 15fps
Audio codec	AAC (MPEG-4 Audio) 48kHz-Stereo 32, 48, 64 kb/s, reference

Audio-visual quality assessment tests were performed in four sessions as shown in Table 2. These sessions were divided in accordance with four media quality factor categories: multimedia quality (namely, comprehensive overall quality), audio quality and video quality in multimedia (namely, mutual interaction quality), video quality, and audio quality (namely, media independent quality), and the related scores were obtained as MOS_{AV}, Aq(Vq), Vq(Aq), Vq, and Aq.

Table 2 Content categories in subjective assessment.

	Source signal	Quality factor	Score
1	Audio+Video	Multimedia quality	MOS_{AV}
2	Audio+Video	Mutual interaction quality: audio quality and video quality in multimedia	Aq(Vq) Vq(Aq)
3	Video	Media independent quality: video quality	Vq
4	Audio	Media independent quality: audio quality	Aq

3.2 Subjective Assessment Results

Fig.3 shows an example of Vq(Aq) and Vq, and Fig.4 shows an example of Aq(Vq) and Aq, for source signal #1 report. For each bit-rate of video codec, Vq(Aq) has five scores, depending on the different bit-rate of audio codec, with the fifth score showing the score for Vq alone, with no audio signal. Vq(Aq) does not show the same quality as Vq, while similarly, Aq(Vq) does not show the same quality as Aq. Therefore, it is concluded that there is an interaction between the video and audio information.

Fig.5 shows an example of $MOS_{AV,}$ for source signal #1. Figs. 3 to 5 showed the results of source signal #1 as an example, and for the other source signal #2 and #3, the similar results were obtained.

In the next section, the assessed MOS_{AV} will be estimated by the proposed method.

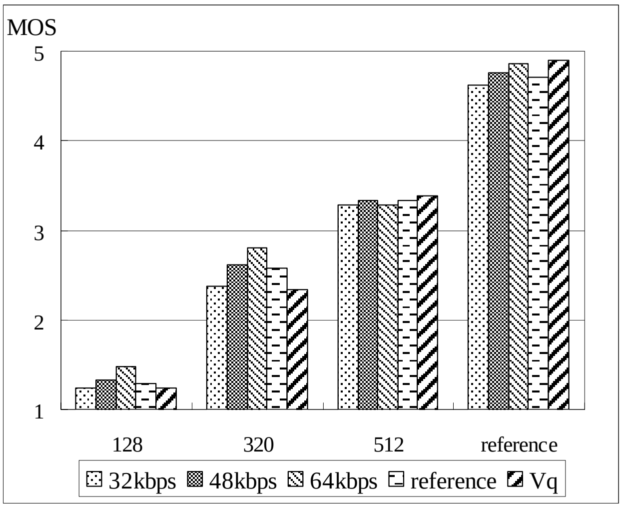

Fig.3 An example of Vq(Aq), Vq for source signal #1.

Step 1 of the proposed multimedia opinion model (Method 2 in Section 2.2) may be expressed as Eqs. (2) and (3) by modeling the relation of Vq(Aq), Aq(Vq) and Vq, Aq obtained by the subjective assessment test.

$$Vq(Aq) = 0.931\ Vq + 0.039\ Aq + 0.099 \qquad (2)$$

$$Aq(Vq) = 0.043\ Vq + 0.890\ Aq + 0.387 \qquad (3)$$

Figs. 6 and 7 show the relation between the objectively estimated value from subjective media independent quality Aq and Vq, and that evaluated in the subjective assessment test, for Vq(Aq) and Aq(Vq), respectively. For Vq(Aq), R^2 (Decision Coefficient) is 0.981, and RMSE (Root Mean Square Error) is 0.159. For Aq(Vq), R^2 is 0.966, and RMSE is 0.216. These experimental results show that the correlation between estimated value and evaluated value is very strong, and the RMSE is very small.

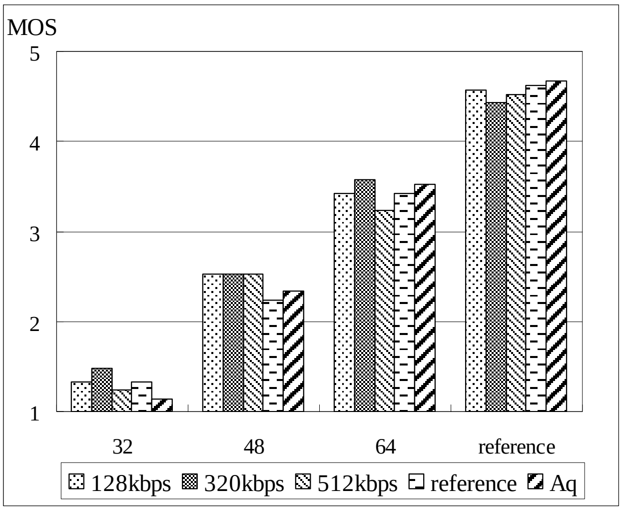

Fig.4 An example of Aq(Vq), Aq for source signal #1.

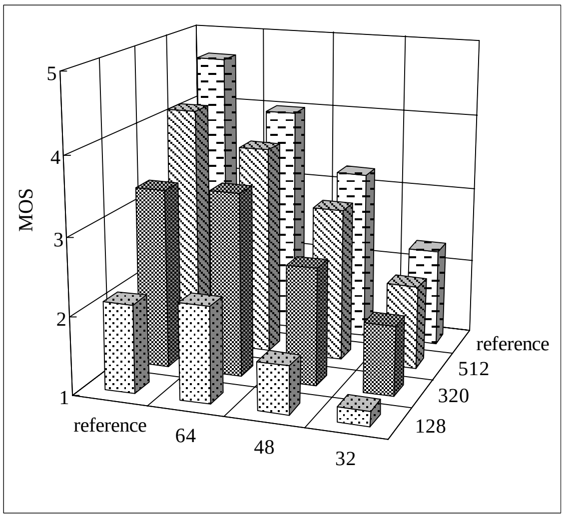

Fig.5 An example of MOS_{AV} for #1 source signal.

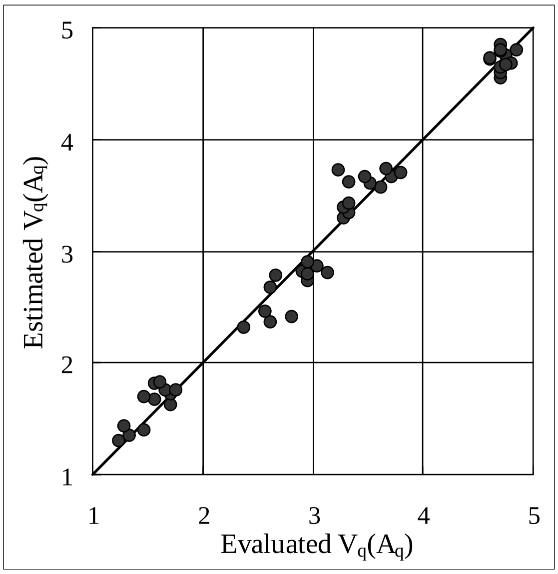

Fig.6 Relationship between evaluated Vq(Aq) in subjective test and estimated Vq(Aq) from model.

4. OBJECTIVE ESTIMATION OF MULTIMEDIA QUALITY

4.1 Objective Estimation of Mutual Interaction Quality Vq(Aq) and Aq(Vq) from Media Independent Quality Vq and Aq

4.2 Objective Estimation of MOS_{AV}

This section describes the estimation of MOS_{AV} from Vq(Aq) and Aq(Vq). Here, the scores were obtained in two ways:

Case 1: from subjectively evaluated Vq(Aq) and Aq(Vq), Case 2: from objectively estimated Vq(Aq) and Aq(Vq) from subjective Vq and Aq.

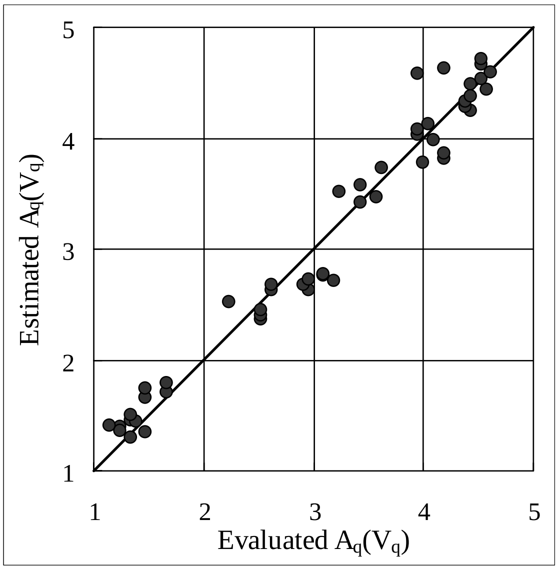

Fig.7 Relationship between evaluated Aq(Vq) in subjective test and estimated Aq(Vq) from model.

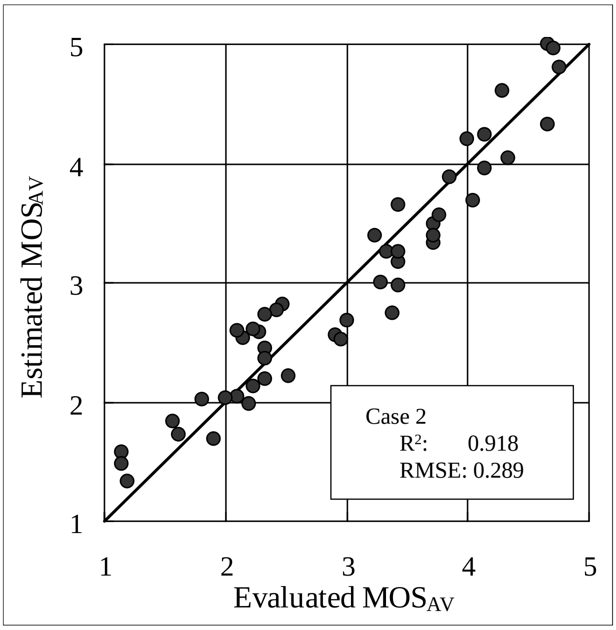

Fig.8 Relationship between evaluated MOS_{AV} and estimated MOS_{AV} (Case 2).

Case 1 uses the subjectively evaluated Vq(Aq) and Aq(Vq). In case 2, MOS_{AV} is estimated as shown in Eq.(4) from the estimated Vq(Aq) and Aq(Vq) described in Section 4.1.

$$MOS_{AV} = 0.188\ Vq(Aq) + 0.211\ Aq(Vq) + 0.112\ Vq(Aq)Aq(Vq) + 0.618 \quad (4)$$

Fig.8 shows an example of the relation between estimated MOS_{AV} from estimated Vq(Aq), Aq(Vq) (Case 2), and evaluated MOS_{AV}. Table 3 shows R^2 and RMSE for each case and also for the previous opinion model (Method 1).

It is shown that the proposed multimedia opinion model (Case 1) has best performance index, and the performance index of the Case 2 is almost equivalent to that of Method 1. Therefore, it is concluded that for construction of multimedia opinion model, it is important to take into account of mutual interaction of audio and video information. Further improvement is needed in the performance of Case 2.

Table 3 R^2 (Decision Coefficient) and RMSE (Root Mean Square Error) between evaluated and estimated scores.

Multimedia opinion model		R^2	RMSE
Method 1	Subjectively evaluated Vq, Aq	0.922	0.287
Method 2	Case 1: subjectively evaluated Vq(Aq), Aq(Vq)	0.951	0.233
	Case 2: Objectively estimated Vq(Aq), Aq(Vq)	0.918	0.289

5. Multimedia Quality Estimation from Physical System Parameters

The final goal of the work is multimedia quality estimation from physical system parameters such as bit-rate, packet loss rate, and display size. As a first step, this paper describes multimedia quality estimation from bit-rate of the coding system for videophone and PDA as well as mobile videophone systems.

5.1 Subjective Experiment

To obtain the modeling equation of media independent quality Vq and Aq (Case 2 of proposed Method 2) from bit-rate as physical parameter, subjective experiment was performed. Table 4 shows the experimental conditions. Assessment regulation and test method were same as Table.1. Number of subjects was 30, of which each half was used for modeling and verification.

Content categories and obtained scores in the subjective assessment were same as Table.2.

Table 4 Experimental conditions for the subjective tests.

Assessment regulation	ITU-T Rec.P.911
Test method	Five-level ACR
Number of subjects	30
Source signals	#1 report, #2 news, #3 scene + BGM
Video codec	MPEG-4 (QVGA) 128-1024 kb/s Size 320x240 pixels, 15fps
Audio codec	AAC (MPEG-4 Audio) 48kHz-Stereo 48-96 kb/s

5.2 Objective Estimation of Multimedia Quality from Bit-Rate of Coding System

First, independent media quality Vq and Aq were estimated respectively from bit-rate of the coding system, using the relation between subjectively evaluated Vq and Aq, and bit-rate of coding. The estimation equations were Eqs. (5) and (6).

$$Vq = \alpha x^2 + \beta x + \gamma \text{ (x: video bit-rate of kb/s)} \quad (5)$$

$$Aq = \alpha y^2 + \beta y + \gamma \text{ (y: audio bit-rate of kb/s)} \quad (6)$$

Then, Vq(Aq) and Aq(Vq) in the Case 2 of proposed Method 2 are:

$$Vq(Aq) = \alpha Vq + \beta Aq + \gamma \quad (7)$$

$$Aq(Vq) = \alpha Vq + \beta Aq + \gamma \quad (8)$$

Finally, MOS_{AV} is:

$$MOS_{AV} = \alpha Vq(Aq) + \beta Aq(Vq) + \gamma Vq(Aq)Aq(Vq) + \delta \quad (9)$$

Fig.9 shows the relation between the objectively estimated MOS_{AV} computed by Eqs.(5)-(9) using bit-rate of coding system as an input, and that directly evaluated in the subjective assessment test for MOS_{AV}.
In Fig.9, R^2 is 0.98, and RMSE is 0.154. These experimental indexes show that the correlation between estimated value and evaluated value is very strong, and the RMSE is very small.
It is concluded that the proposed multimedia opinion model is promising methodology for estimation multimedia quality. However, other physical parameters than bit-rate should be studied further.

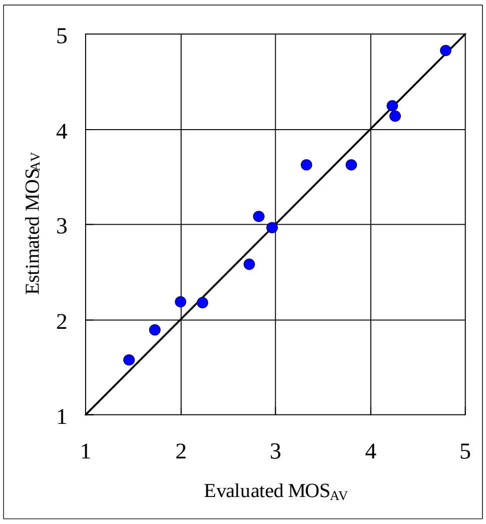

Fig.9 Relationship between evaluated MOS_{AV} and estimated MOS_{AV} using bit-rate of coding system as input parameters.

6. CONCLUSION

This paper describes an objective perceptual multimedia quality model for audio-visual communications, taking into account the mutual interaction of audio and video information. This paper also described an application of the proposed method to estimate a multimedia quality of videophone and PDA as well as mobile videophone services.
It is concluded that the construction of a multimedia opinion model, it is important to take account of the mutual interaction of audio and video information. The proposed method shows good performance for commercial use. The performance of Case 2 of the proposed method needs further improvement. Other physical parameters than bit-rate should also be studied further.

7. REFERENCES

[1] ITU-T Rec. G.107, "The E-model, a computational model for use in transmission planning," March 2003.
[2] ITU-T Rec. J.148, "Requirements for an objective perceptual multimedia quality model," May 2003.
[3] J.G. Beerends and F.E.De Caluwe, "The Influence of Video Quality on Perceived Audio Quality and Vice Versa," *J. Audio Eng. Soc.*, Vol.47, No.5, 1999.
[4] T. Hayashi, and K. Yamagishi, "Multi-Modal Quality Estimation Model based on Interaction of Audio and Video Quality (in Japanese), Proc. The 2004 IEICE General Conference, B-11-12, p.505, March 2004.