

# MULTIMEDIA QUALITY INTEGRATION USING PIECEWISE FUNCTION

Shuai Wan, Kaifang Yang and Haiyong Zhou

School of Electronics and Information, Northwestern Polytechnical University, China

Emails: swan@nwpu.edu.cn, yangkaifang5@126.com

**Keywords:** Multimedia Quality, Quality Integration, Piecewise Function, G.1070

**Abstract.** In this paper the important issue of multimedia quality evaluation is concerned, given the unimodal quality of audio and video. Firstly, the quality integration model recommended in G.1070 is evaluated using experimental results. Theoretical analyses aide empirical observations suggest that the constant coefficients used in the G.1070 model should actually be piecewise adjusted for different levels of audio and visual quality. Then a piecewise function is proposed to perform multimedia quality integration under different levels of the audio and visual quality. Performance gain observed from experimental results substantiates the effectiveness of the proposed model.

## 1. Introduction

The next-generation network (NGN) is expected to enable service providers to build high-performance networks for converged services. Rapid development of technologies in networks and multimedia has led to a proliferation of various multimedia services such as videophones, videoconferencing, video on demand (VoD), and IPTV. Since for most multimedia applications over the NGN, the end users would be human beings, the services should be provided with sufficient quality in terms of human perception. It is therefore important to properly evaluate users' perceptual quality of multimedia applications, i.e., quality of experience (QoE), in order to provide users with economical and comfortable services.

Research on understanding human perception of the quality of services has been traditionally focused on individual modalities. Given that both audio and visual quality models are relatively mature [1]-[3], it seems logical to use them as primary inputs if the quality of multimedia services is concerned. The research on multimedia quality modeling then addresses how to integrate information of two modalities of video and audio together, to describe the multimedia quality combination rules, and to apply these rules to the unimodal inputs entering the multimedia quality model. A general framework of the multimedia quality model is shown in Fig. 1. In the framework, information about audio and visual quality is employed as the input of the multimedia combination function model where specific rules apply to predict subjective quality of the multimedia.

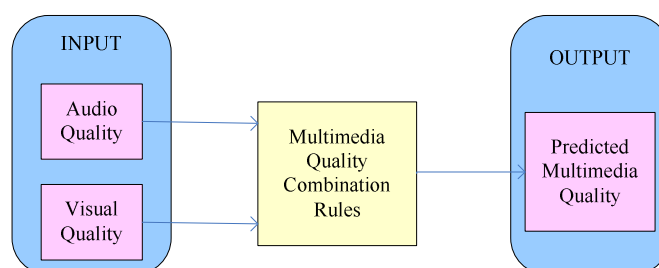


Fig. 1. General framework of multimedia quality model

Research found that a human integrates multimodal information together mainly following three algebraic rules [4]: adding, averaging and multiplying. D. S. Hands further suggests that subjects integrate errors in audio and video together following the multiplicative rule [5]. Using regression analysis method, objective multimedia quality models have been proposed for multimedia quality assessment [5]-[10], whereas little attention has been paid to the variability of the model under different levels of the quality of the input audio and video.

In this paper, we propose a model for multimedia quality integration using a piecewise function, which estimates the overall multimedia quality under different level intervals of the audio quality and the visual quality. Firstly, the multimedia quality model reported in G.1070 [9] is studied and analyzed. The analyses and observations show that coefficients in G.1070 should be piecewise changed instead of being constant as its present form. Therefore we propose to perform multimedia quality integration using a piecewise function accommodating different quality levels of the input audio and video.

The paper is organized as follows. Section 2 introduces the source materials, test conditions and the subjective assessment method used for evaluation in this paper. In section 3, we employ the data obtained in the subjective experiments to analyze the multimedia quality integration function reported in G.1070, where the problem of the related coefficients being constant under different levels of the audio and visual quality is described in section 4. Based on the analysis, a piecewise function is proposed to formulate multimedia quality integration in the same section, and its performance is verified in terms of correlation coefficient and evaluation error between subjective and estimated qualities. This paper closes with conclusions given in section 5.

## 2. Subjective assessment

### 2.1. Source Materials

Three video clips from the consumer digital video library [11] were used for all experiments, where the content is shown in Fig. 2. Specifically, clip A shows a man giving an overview of National Telecommunications and Information Administration (NTIA). Clip B is about a woman describing her granddaughter's art work, and clip C shows a man introducing the U.S. spectrum chart. The source materials are about 15 seconds each, 16-bit PCM mono sampled at 8kHz, and at 30fps, QCIF in the frame size.



Fig. 2. Content of test scenes

### 2.2. Test Conditions

The audio codec selection was EVRC [12], and video codec selection was MPEG-4 in ffmpeg [13]. For each of the test clips, the video quality was set variable by using different quantization parameters ( $QP$ ). That is, in the experiments,  $QP = 3$ ,  $QP = 6$ ,  $QP = 9$ ,  $QP = 12$ ,  $QP = 15$ ,  $QP = 18$ ,  $QP = 21$ ,  $QP = 25$ ,  $QP = 27$ , and  $QP = 31$ , were respectively used to encode the source clips to produce sequences in different visual quality for tests. Audio conditions were changed by bitrates and packet loss rates as illustrated in Table 1.

### 2.3. Assessment Method

The absolute category rating (ACR) method [14] was used in the subjective tests. The viewers have evaluated the video quality using a slider device and a continuous grading scale marked with "Excellent", "Good", "Fair", "Poor" and "Bad" respectively. The subjective scores are therefore quantized on a scale of 1 to 5.

20 subjects (10 female, 10 male) with normal vision and hearing participated in the tests. The monitor used in the subjective assessments was a 19" LCD screen, which was operated at its native resolution of  $1440 \times 900$  pixels. The viewing distance was not fixed, and was about 5 times the height of the video frame. The subjective tests consisted of one session of about 30 minutes,

including 5 minutes for rest. The tests were composed of three parts: the audiovisual quality test, the audio-only quality test and the video-only quality test, respectively. In the audio-only quality test, the screen was set gray. In the video-only test, there was no headphone provided. Subjects were asked to rate the quality of the presentation in each case. The order of the clips was randomized for each individual subject. The subjective quality of each clip was represented as a mean opinion score (MOS) calculated by averaging the scores of 20 subjects.

Table I AUDIO TEST CONDITIONS

Audio condition	1	2	3	4	5
Bit rate (kb/s)	4.00	8.55	8.55	8.55	source
Packet-loss rate (%)	0	0	10	20	0

### 3. Multimedia Quality Integration Function In G.1070

#### 3.1. Multimedia Quality Integration Function in G.1070

To evaluate the integration mechanism in multimedia quality perception, the perceived audio, visual and audiovisual qualities were obtained by extensive subjective evaluations, where the test conditions are described in section 2. In this paper, the audiovisual or multimedia quality is denoted as  $MOS_{AV}$ , while the corresponding perceived audio quality and video quality are denoted as  $MOS_A$  and  $MOS_V$ , respectively. It has been accepted that subjects integrate audio and video together following the multiplicative rule [5], which can be described as:

$$\rho = MOS_A \times MOS_V, \quad (1)$$

$$MOS_{AV} = c_0 + c_1 \rho, \quad (2)$$

where  $\rho$ ,  $c_0$  and  $c_1$  are parameters constantly determined for a specific task.

Based on the above multiplicative rule, a more suitable function for multimedia quality integration is further proposed through regression analysis [5], used in G.1070, as:

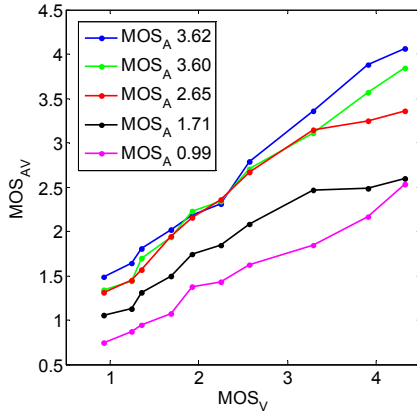
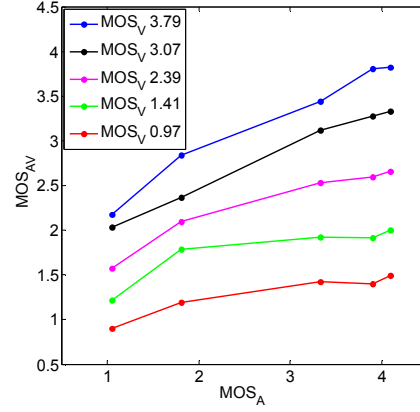
$$MOS_{AV} = a_1 \times MOS_A + a_2 \times MOS_V + a_3 \times MOS_A \times MOS_V + a_4, \quad (3)$$

where coefficients  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  are dependent on the display size of the related video and characteristics of the conversational task.

#### 3.2. Empirical Analysis

In this subsection, the multiplicative rule and the model provided by G.1070 are evaluated using MOSs. Given the model formulated by (1) and (2), the multiplicative rule shows that if  $MOS_A$  is a constant,  $MOS_{AV}$  is a linear function of  $MOS_V$  where the intercept of the linear function, i.e.  $c_0$ , is unchanged for various values of the given  $MOS_A$ . However, the subjective results do not support this argument. As shown in Fig. 3,  $MOS_{AV}$  values do change linearly with  $MOS_V$  for a given constant  $MOS_A$ . However, for different values of  $MOS_A$ , the intercepts are clearly not the same. It comes to the same conclusion when the visual quality is treated as a constant, as shown in Fig. 4.

The above mentioned problem can be avoided using the model given by G.1070. Without loss of generality, let us consider a constant audio quality of  $MOS_A$ . Following (3), the multimedia quality  $MOS_{AV}$  would change linearly with  $MOS_V$  for a constant  $MOS_A$ , whereas the related intercept varies for different values of  $MOS_A$ . Through further investigations on the coefficients of G.1070 model, we found that piecewise characteristics exist and the coefficients in (3) should be modified according to different levels of audio and visual quality. The detailed analyses are provided in the following section.

Fig. 3.  $MOS_{AV}$  vs  $MOS_V$  given  $MOS_A$ Fig. 4.  $MOS_{AV}$  vs  $MOS_A$  given  $MOS_V$ 

#### 4. Multimedia Quality Integration Using Piecewise Function

##### 4.1. Piecewise Characteristics of Coefficients in G.1070

For clarity of analyses, the model in (3) is reformulated as:

$$MOS_{AV} = a_5 \times MOS_V + a_6, \quad (4)$$

where  $a_5 = a_3 \times MOS_A + a_2$ , and  $a_6 = a_1 \times MOS_A + a_4$ .

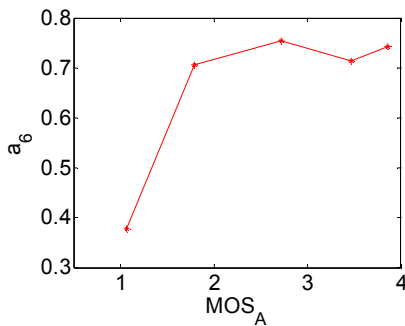
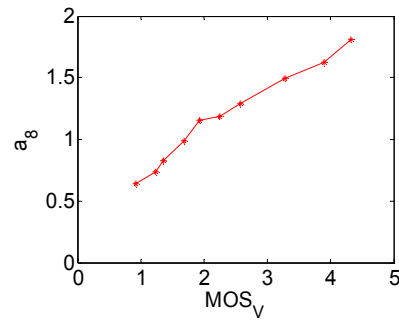
$$MOS_{AV} = a_7 \times MOS_A + a_8, \quad (5)$$

where  $a_7 = a_3 \times MOS_V + a_1$ , and  $a_8 = a_2 \times MOS_V + a_4$ .

Given a constant  $MOS_A$ ,  $MOS_{AV}$  changes linearly with  $MOS_V$  where the slope and intercept equal to  $a_5$  and  $a_6$ , respectively.

Similarly, if  $MOS_V$  is constant,  $MOS_{AV}$  values are shown as a linear function of  $MOS_A$  in (5). The slope of the function in (5) equals to  $a_7$ , and the intercept equals to  $a_8$ .

In the original form of G.1070 model, the coefficients of  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  are given for a certain task when the display size of video is determined. In this case,  $a_5$  and  $a_6$  should change linearly with  $MOS_A$ , while  $a_7$ , and  $a_8$  should change linearly with  $MOS_V$ . However, experimental results show that the linearity in  $a_6$  and  $a_8$  is quite questionable, as shown in Fig. 5 and 6.

Fig. 5.  $a_6$  vs  $MOS_A$ Fig. 6.  $a_8$  vs  $MOS_V$ 

Generally speaking, when  $MOS_A < 2$ , the slope of  $a_6$  is greater than that at  $MOS_A > 2$ , and the intercept is comparatively smaller in Fig. 5. Accordingly,  $a_1$  should be greater when  $MOS_A < 2$  than  $MOS_A > 2$ , and  $a_4$  should be smaller in the same case. As shown in Fig. 6, the linearity of  $a_8$  twists

when  $MOS_V$  reaches about 2 as well. Similarly, when  $MOS_V < 2$ , the slope of the curve is greater than  $MOS_V > 2$ , and the intercept is smaller. Therefore,  $a_2$  should be greater when  $MOS_V < 2$  than at  $MOS_V > 2$ , and  $a_4$ , on the other hand, should be smaller for  $MOS_V < 2$ .

Based on the above analyses, coefficients  $a_1$ ,  $a_2$ ,  $a_3$ , and  $a_4$  should be different when the audio or the visual quality lie in the interval of (0,2) and when any of the quality is in the interval of (2,5). Considering possible combinations, the coefficients in G.1070 have piecewise characteristics and should not be set the same for different levels of audio or visual quality.

#### 4.2. Multimedia Quality Integration Using Piecewise Function

To incorporate the piecewise characteristics of the coefficients into quality integration, a piecewise function is proposed to obtain the multimedia quality where the coefficients are empirical obtained for different levels of the audio and the visual quality. That is,

If  $MOS_A < 2, MOS_V < 2$ ,

$$MOS_{AV} = 0.0925647 \times MOS_A + 0.125142 \times MOS_V + 0.340851 \times MOS_A \cdot MOS_V + 0.246395 ,$$

If  $MOS_A < 2, MOS_V \geq 2$ ,

$$MOS_{AV} = 0.386617 \times MOS_A + 0.335879 \times MOS_V + 0.0910265 \times MOS_A \cdot MOS_V + 0.078473 ,$$

If  $MOS_A \geq 2, MOS_V < 2$ ,

$$MOS_{AV} = 0.403925 \times MOS_A + 1.32839 \times MOS_V - 0.157422 \times MOS_A \cdot MOS_V - 0.745953 ,$$

If  $MOS_A \geq 2, MOS_V \geq 2$ ,

$$MOS_{AV} = -0.315292 \times MOS_A + 0.0188303 \times MOS_V + 0.19284 \times MOS_A \cdot MOS_V + 1.99556 , \quad (6)$$

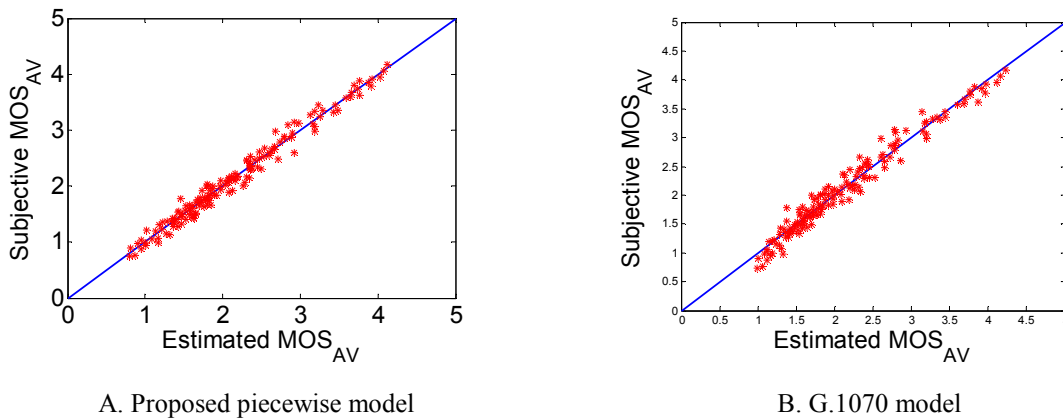


Fig. 7. Relationship between estimated and subjective  $MOS_{AV}$

To evaluate the performance of the proposed model, the  $MOS_{AV}$  is estimated using the proposed method and the G.1070 model, respectively. The relationship between estimated multimedia qualities and the actual subjectively measured  $MOS_{AV}$  is shown in Fig. 7. Compared with the multimedia model reported in G.1070, the proposed model provides a better fit with the actual data. Specifically, the Pearson correlation coefficient (PCC) and the root mean square error (RMSE) between estimated multimedia qualities and the actual data are 0.99 and 0.12. Comparatively, the PCC and RMSE between the model recommended by G.1070 and the actual data are 0.98 and 0.15. Clearly, the proposed model achieves better performance in accuracy.

## 5. Conclusions

This paper addresses how to accurately assess the perceived multimedia quality given the individual audio and visual quality. Extensive subjective tests have been performed to evaluate the model recommended in G.1070, where analyses reveal that the corresponding constant coefficients should actually be adapted to the different levels of the unimodal quality. Accordingly, we propose to formulate the multimodal quality integration using a piecewise function. The proposed model has been shown to be more accurate compared to the G.1070 model.

## 6. Acknowledgments

This work was supported by the National Science Foundation of China (60902052, 60902081), the Doctoral Fund of Ministry of Education of China (No. 20096102120032) and the NPU Foundation for Fundamental Research (JC201038).

## Reference

- [1] ITU-T Rec. G.107, "The E-model, a computational model for use in transmission planning," March 2005.
- [2] A. Takahashi, A. Kurashima, and H. Yoshino, "Objective Assessment Methodology for Estimating Conversational Quality in VoIP," IEEE Trans. on Speech Audio and Language Processing, vol. 14, no. 6, pp. 1983-1993, Nov. 2006.
- [3] K. Yamagishi and T. Hayashi, "Opinion Model for Estimating Video Quality of Videophone Services," IEEE GLOBECOM 2006, Nov. 2006.
- [4] N. H. Anderson, Foundations of Information Integration Theory. New York: Academic, 1981.
- [5] D. S. Hands, "A Basic Multimedia Quality Model," IEEE Transactions on Multimedia, vol. 6, no. 6, pp. 806-816, Dec. 2004.
- [6] C. Jones and D. J. Atkinson, "Development of opinion-based audiovisual quality models for desktop video-teleconferencing," IEEE Sixth Int. Workshop on Quality of Service (IWQoS'98), pp.196 - 203, May 1998.
- [7] S. Winkler, C. Faller, "Audiovisual Quality Evaluation of Low-Bitrate Video," Proc. SPIE/IS&T, Human Vision and Electronic Imaging, pp. 139-148, San Jose, CA, Jan. , 2005.
- [8] X. Tan, J. Gustafsson and G. Heikkilä, "Perceived Video Streaming Quality under Initial Buffering and Rebuffering Degradations," MESAQIN 2006 conference.
- [9] ITU-T Rec. G.1070, "Opinion model for video-telephony applications," April 2007.
- [10] Takanori Hayashi, Kazuhisa Yamagishi, Toshiko Tominaga, and Akira Takahashi, "Multimedia Quality Integration Function for Videophone Services," IEEE Global Telecommunications Conference, 26-30 Nov. 2007.
- [11] <http://www.cdvl.org/index.php>
- [12] [http://www.3gpp2.org/Public\\_html/specs/alltsgscfm.cfm](http://www.3gpp2.org/Public_html/specs/alltsgscfm.cfm)
- [13] <http://ffmpeg.org/>
- [14] ITU-T Rec.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," April 2008