OPINION MODEL FOR ESTIMATING CONVERSATIONAL QUALITY OF VOIP

Akira Takahashi

NTT Service Integration Laboratories
E-mail: takahashi.akira@lab.ntt.co.jp

ABSTRACT

VoIP (Voice over IP) is one of the key technologies for recent telecommunication services. Since IP networks do not generally guarantee the transmission quality, it is extremely important to design and manage the QoS (Quality of Service) properly. To do this, it is indispensable to develop a systematic method which estimates subjective quality based on physical characteristics of the VoIP system. This paper investigates the performance of the existing opinion model called E-model, and proposes a method to improve it. The experimental results showed that the proposed model surely improves the estimation accuracy in evaluating practical VoIP systems.

1. INTRODUCTION

VoIP (Voice over IP), which integrates conventional telephone services with a growing number of IP-based applications, is one of the most important technologies for recent telecommunication services. In addition to the cost-reduction effect achieved by sharing telecommunication devices, VoIP is expected to accelerate the development of rich multimedia services.

Since the quality of IP networks is not generally guaranteed, it is important to properly design networks and/or terminals before providing services, as well as to monitor the quality of VoIP services constantly and take necessary action to maintain the level of service.

The quality of service (QoS) of VoIP should be discussed in terms of subjective quality, which corresponds to users' perceptions of transmitted speech. Since subjective quality assessment is time-consuming and expensive, however, a method that estimates subjective quality by measuring physical characteristics of the terminals and networks is needed. This method is called "objective quality assessment."

In particular, a computational model that estimates the overall speech communication quality taking into account not only listening quality but also conversational factors such as delay and echo is called the "opinion model." ITU-T (International Telecommunication Union - Telecommunication Sector) standardized an opinion model called "E-model" as Recommendation G.107 in 1998. Although E-model is the most widely used opinion model for network planning of VoIP services, its performance has not necessarily been completely investigated [1].

This paper first investigates the estimation accuracy of E-model for the major quality factors in VoIP which are speech distortion, delay, talker echo, and loudness. Then, it proposes a new model based on E-model for improving performance, and demonstrates the validity of the new model in estimating the conversational quality of practical VoIP systems.

2. PERFORMANCE OF E-MODEL

2.1. Subjective quality experiments

We conducted subjective quality experiments to investigate the performance of E-model for individual quality factors such as loudness, delay, and talker echo, as well as the interaction between speech distortion and delay. Table 2.1 shows the experimental parameters in each experiment. Other experimental conditions were determined so that they complied with the values which E-model assumed as its default settings.

Table 2.1 Experimental settings for Exp. #1 and #2.

Exp. #1	
Subset A	
OLR [dB]	3, 5, 10, 13, 15, 25, 30
Subset B	
Ta [msec]	10, 30, 50, 70, 100, 300, 500
Subset C	
MNRU [dB]	12, 24, 30, 36
Ta [msec]	10, 30, 50, 70, 100, 300, 500
Subset Ref.	
MNRU-Q[dB]	0, 6, 12, 18, 24, 30, 36, 40
Exp. #2	
Subset A	
Ta [msec]	10, 30, 50, 70, 100, 300, 500
TELR [dB]	15 - 65[*1]
Subset Ref.	
MNRU-Q[dB]	0, 6, 12, 18, 24, 30, 36, 40

[*1] Five "TELR" conditions for each "Ta" condition

0-7803-8484-9/04/$20.00 ©2004 IEEE

Table 2.2 Subjective experimental conditions.

No. of subjects	40
Methodology	Rec. P.800 "ACR test"
Duration of conversation	1 min./condition
Ambient noise	Hoth noise@35 dB(A)
Conversational task	Free conversation[*1]

[*1] Subjects were allowed to perform a task in which they determined the shape of a figure by receiving oral information.

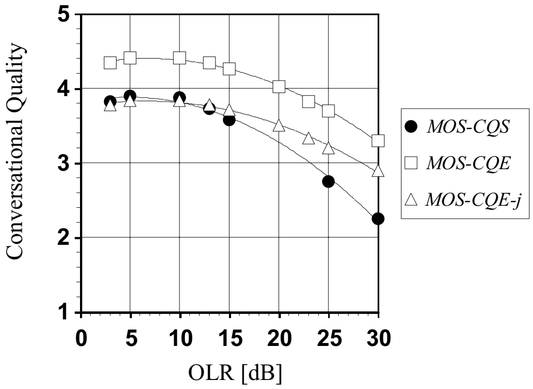


Figure 2.1 Accuracy of E-model (Loudness)

In the first experiment (Exp. #1), we evaluated the subjective quality for loudness and delay individually. "*OLR*" and "*Ta*" denote the overall loudness rating and one-way absolute delay, respectively. In addition, we investigated the additive property of the impairment caused by delay and speech distortion. The amount of speech distortion was controlled by the Q-value of the reference MNRU (Modulated Noise Reference Unit) defined by ITU-T Recommendation P.810.

In the second experiment (Exp. #2), we evaluated the effect of talker echo for various *TELR* (Talker Echo Loudness Rating) and delay combinations.

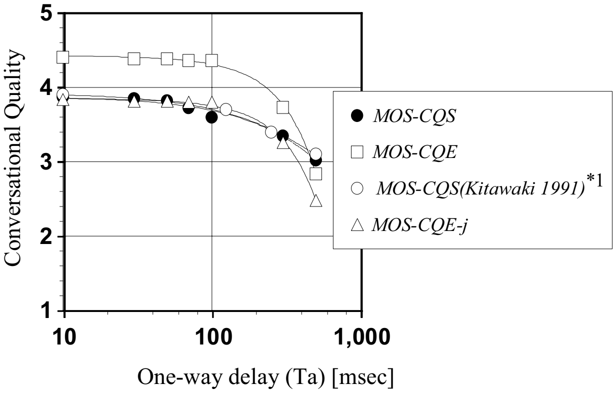
The subjective experiments were conducted using the conversational ACR (Absolute Category Rating) method defined in ITU-T Recommendation P.800. The associated testing conditions are summarized in Table 2.2.

2.2. Quality estimation by E-model

E-model has 20 parameters that represent the network, terminal, and environmental quality factors. The experimental system parameters we used in subjective assessment basically comply with the default settings defined in ITU-T Recommendation G.107, unless otherwise specified. We determined the "mean one-way delay of the echo path (*T*)" to be equal to the "one-way absolute delay (*Ta*)" and the "round trip delay in a four-wire Loop (*Tr*)" to be 2* *Ta*.

Although ITU-T Recommendation G.107 Annex B provided the relationship between the R-value, which is the output of E-model, and the estimated MOS (MOS-CQE as defined in ITU-T Recommendation P.800.1), it was not directly applied to the Japanese experimental results. This is because a systematic difference exists between the Western MOS and the Japanese MOS. Therefore, we estimated the Japanese MOS-CQE, which is denoted by MOS-CQE-j hereafter, from the R-value by using the following transformation. We transformed the R-value to MOS-CQE by applying the formula defined in ITU-T Recommendation G.107 Annex B:

$$\text{"MOS - CQE - j"} = 0.8681 \cdot \text{"MOS - CQE"} + 0.0271 \qquad (1)$$



**1 In "MOS-CQS(Kitawaki 1991)", the plot at Ta = 10 indicates Ta = 0.*

Figure 2.2 Accuracy of E-model (Delay)

2.3. Analysis

2.3.1 Loudness

Figure 2.1 demonstrates the relationship between the subjective MOS (MOS-CQS) and its E-model estimates (MOS-CQE and MOS-CQE-j) for different OLR conditions (Subset A in Exp. #1). The MOS-CQE-j can be readily seen to better fit the subjective MOS-CQS than the raw estimate, which is MOS-CQE.

In comparing MOS-CQS and MOS-CQE-j, we can say that the E-model estimates the subjective quality fairly well around the nominal *OLR* value, which is 10 dB.

2.3.2 Delay

Figure 2.2 demonstrates the relationship between the subjective MOS and its E-model estimates for different delay conditions (Subset B in Exp. #1).

Although the E-model assumes that no degradation due to delay occurs up to 100 msec, the real subjective quality shows different characteristics. In addition, the E-model tends to underestimate the quality for delays beyond 300 msec. In Figure 2.2, the experimental results by Kitawaki and Itoh in 1991 [2] are also plotted as a reference. This data is very close to that obtained in our experiment. It thus supports our claim.

2.3.3 Talker Echo

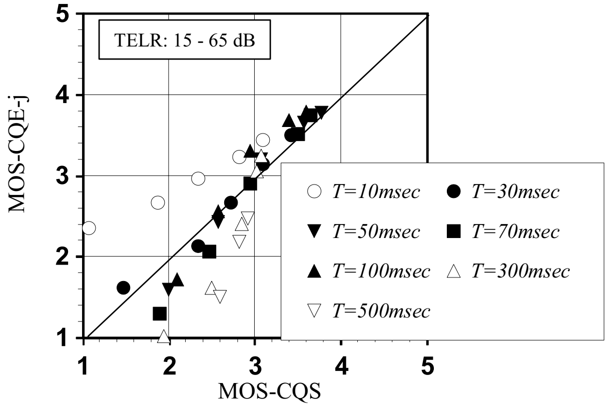


Figure 2.3 Accuracy of E-model (Talker echo)

Figure 2.3 demonstrates the relationship between the subjective MOS and its E-model estimates for different echo loudness and delay conditions (Subset A in Exp. #2). In Exp. #2, the subjects experience the echo of their own voice delayed by 2**T* msec, as well as the delay that is experienced as a delay in far-end speech.

The estimate tends to fit relatively well with the subjective quality for "good" (MOS-CQS > 3.0) talker-echo conditions. For conditions in which the talker echo is objectionable, however, the estimate diverges from the actual subjective quality.

2.3.4 Interaction between delay and speech distortion

E-model assumes that individual quality factors such as loudness, delay, talker echo, and speech distortion are mutually independent on the psychological scale. We investigate the interaction between the delay and speech distortion using the data from Subset C in Exp. #1.

The *Ie,eff* (Effective Equipment Impairment Factor), which represents the speech distortion by low bitrate coding and packet loss, is determined by the kind of CODEC and the packet-loss rate, as given in ITU-T Recommendation G.113 Appendix I, in which the *Ie,eff* for the MNRU conditions is not provided. In our investigation, we defined *Ie,eff* under the MNRU conditions in the following way:

1) Transform the MOS-CQS for the MNRU conditions with *Ta* = 10 msec (Subset Ref. in Exp. #1) into the R-value. This is done by first transforming the MOS-CQS to MOS-CQE based on Equation (1), and then converting it to the R-value based on the function provided by ITU-T Recommendation G.107 Appendix I.
2) Calculate the R-value for *Ta* = 10 msec based on the E-model, setting all the other parameters to the default value in ITU-T Recommendation G.107. Let this value be R0.
3) The difference between R0 and the value derived for the MNRU condition is defined as *Ie,eff* for the associated MNRU condition.

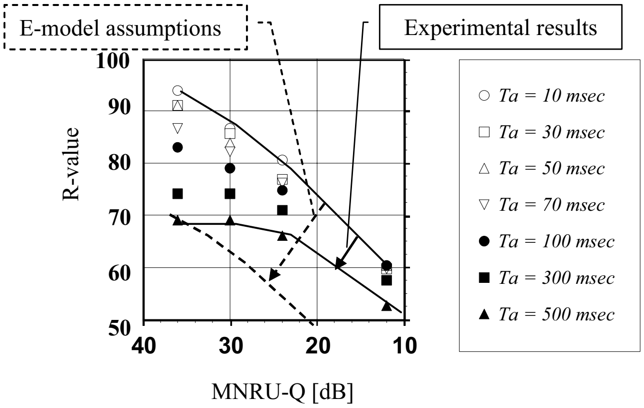


Figure 2.4 Interaction between pure delay and speech distortion.

Figure 2.4 shows the impact of speech distortion on the R-value for various delay conditions. This figure implies that the effect of speech distortion is dependent on a delay condition in terms of the R-value. That is, the additive property that the E-model assumes does not necessarily hold, especially for a low-quality region.

3. PROPOSED OPINION MODEL

Taking into account the problems pointed out in the previous section, we modified the E-model so that it better estimates the subjective quality. We call it the modified E-model in this paper. We did not modify the E-model equations for loudness, because the E-model estimation for loudness seems to be reasonable for the regular *OLR* range.

3.1. Delay and talker echo

Since the subjective quality in Figure 2.2 seems to be modeled by a simple 2nd order polynomial, we applied the following equation for the delay impairment factor (*Idd*):

$$Idd = a_1 \cdot Ta^2 + a_2 \cdot Ta \qquad (2)$$

in which a_1 and a_2 are constants.

After investigating the effect of talker echo as a function of echo path delay (*T*) and the echo loudness (*TELR*), we modeled the talker echo impairment factor (*Idte*) by a combination of logarithmic and exponential functions:

$$Idte = \max(b_1 \cdot \mathrm{e}^{b_2 \cdot TELR} + b_3 \cdot \ln(b_4 \cdot T), 0) \qquad (3)$$

in which b_1, b, b_3, and b_4 are constants.

We assumed regular terminal and environmental conditions as given by the default parameters of E-model. Modeling *Idte* for other conditions than those assumed in our experiments is for further study. This assumption is, however, practical in most cases of evaluating handset communications.

Table 4 Settings for Exp. #3.

Exp. #3	
Subset A	
CODEC	G.711 with PLC
Ppl [%]	0, 1, 3, 5
Ta [msec]	100, 150, 300
Subset B	
CODEC	G.729
Ppl [%]	0, 1, 3, 5
Ta [msec]	100, 400
Subset C	
CODEC	G.711
Ppl [%]	0, 1, 3, 5
Ta [msec]	100, 200
Subset D	
CODEC	G.729
Ppl [%]	0, 3
Ta [msec]	100, 150, 200, 300, 400
TELR [dB]	35 - 55（select 2 values for each Ta condition）
Subset Ref.	
MNRU-Q [dB]	0, 6, 12, 18, 24, 40

3.2. Interaction between delay and speech distortion

E-model assumes that the effects of delay and speech distortion are independent. Therefore, the combined impairment (we call this *LQd* hereafter) is expressed as follows:

$$LQd = Idd + Ie,eff \quad (4)$$

However, we modeled *LQd* with a 2nd-order polynomial of *Idd* and *Ie,eff*, because we observed some dependence between them. Here, we apply the new *Idd* defined by Equation (2):

$$LQd' = LQd + Idd \cdot Ie,eff \cdot (c_1 + c_2 \cdot Ie,eff + c_3 \cdot Idd + c_4 \cdot Idd \cdot Ie,eff) \quad (5)$$

in which c_1, c_2, c_3, and c_4 are constants.

4. VALIDITY OF THE PROPOSED MODEL

We compared the performances of the original and the modified E-models by applying them to an evaluation of conversational quality in practical VoIP systems for various delay, echo, CODEC, and packet-loss conditions (Exp. #3). This is unknown data to both the original E-model and the modified E-model.

We used commercial VoIP gateway products with an analog two-wire interface. The network emulator inserted between the gateways controlled the packet delay and packet loss. The experimental conditions are listed in Table 4. In Subsets A, B, and C, the combinations of packet-loss rate (*Ppl* [%]) and delay (*Ta* [msec]) were tested for G.711 PCM CODEC with a packet loss concealment (PLC) algorithm, G.729 CODEC, and G.711 CODEC without PLC, respectively. The *TELR*, as well as the delay and packet-loss rate, were controlled for G.729 CODEC in Subset D. The packet length was 20 msec for each CODEC. All the other conditions were the same as the default values of ITU-T Recommendation G.107.

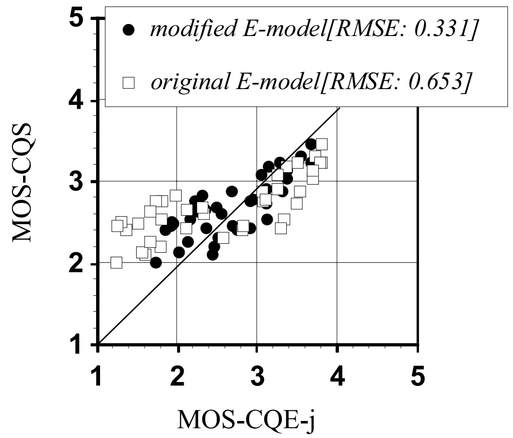


Figure 4 Relationship between subjective and estimated quality.

The estimation accuracy of the original and modified E-models is demonstrated in Figure 4. The figure shows that even if E-model estimates the subjective quality to be 2.5, for instance, the actual subjective quality may vary from 1.5 to 3.5. In terms of Root Mean Square Error (RMSE), the modified E-model reduces the estimation error by about 1/2. The cross-correlation coefficients were 0.763 and 0.793 for the original and modified E-models, respectively.

5. CONCLUSION

We investigated the performance of E-model standardized as ITU-T Recommendation G.107. The experimental results showed that the E-model prediction sometimes diverges from the actual subjective quality in evaluating delay, talker echo and the interaction between delay and speech distortion although it accurately predicts subjective quality in evaluating loudness.

Then, we proposed a new opinion model based on E-model and demonstrated the improvement achieved by the proposal in the evaluation of practical VoIP systems.

6. REFERENCES

[1] S. Möller, “Assessment and Prediction of Speech Quality in Telecommunications,” Kluwer Academic Publishers, Boston, 2000.

[2] N. Kitawaki, and K. Itoh, “Pure delay effects on speech quality in telecommunications,” IEEE Journal on Selected Areas in Communications, Vol. 9 Issue 4, pp. 586-593, May 1991.