IEICE TRANS. COMMUN., VOL.E90–B, NO.12 DECEMBER 2007

PAPER

QoE Estimation Method for Interconnected VoIP Networks Employing Different Codecs

Akira TAKAHASHI[†a)], **Noritsugu EGI**[†], ***and*** **Atsuko KURASHIMA**[†], ***Members***

SUMMARY VoIP is one of the key technologies for recent telecommunication services. In addition to the migration from the conventional PSTN to IP networks, mobile networks will follow the PSTN in moving to an IP-based infrastructure. Due to limited radio resources, the speech bitrate in mobile networks must be more strongly compressed than is true in PSTN. This will lead to a heterogeneous network environment, in which different speech codecs are employed in fixed and mobile networks. Therefore, from the viewpoint of designing and managing the QoE (Quality of Experience) of end-to-end telephony services, establishing a method to evaluate the quality of VoIP in such a heterogeneous network environment is very important. The quality of speech communication services should be discussed in subjective terms. Subjective quality assessment is time-consuming and expensive, however, so objective quality assessment which estimates subjective quality without carrying out subjective quality experiments is desirable. To establish an objective method to evaluate the end-to-end quality of speech in a heterogeneous network environment, this paper proposes a method for estimating the end-to-end listening quality based on the quality in each individual segment. This method is very important because conventional technologies such as the E-model, which was standardized as ITU-T Recommendation G.107, cannot accurately estimate overall quality based on segmental qualities. The experimentals show that the proposed method offers better performance in terms of quality estimation than the conventional method.

key words: *VoIP, objective quality assessment, speech quality, subjective quality, QoE, E-model*

1. Introduction

IP-telephony has become a major telecommunication service. In addition to ordinary wireline services, many services such as voice over wireless LAN and IP-based mobile phone have been investigated. Therefore, seamlessly interconnecting these services is very important to provide users with such a multi-carrier service as a transparent telephony service.

The quality of IP-telephony services should be evaluated in terms of subjective quality, which reflects users' perceptions of speech quality. The fundamental method to measure subjective quality is the subjective quality assessment method, in which users are asked to evaluate the quality of speech communication. Although subjective quality assessment is the most reliable way to obtain subjective quality, it is time-consuming and expensive. In addition, it requires special assessment facilities such as an acoustically shielded chamber. Therefore, an objective means for estimating subjective quality without carrying out a subjective quality assessment is desired. This is called "objective quality assessment [1]."

Objective quality assessment methods can be categorized into five types (see Table 1). The "media-layer model" estimates subjective quality by using speech signals. The most well-known media-layer model is ITU-T Recommendation P.862 "PESQ (Perceptual Evaluation of Speech Quality [2]." The merit of such a model is that subjective quality can be estimated without any *a priori* knowledge about the configuration of terminals/networks, e.g., type of codec and packet-loss rate. This is why media-layer models are called "black-box approaches [3]." The second type is the "packet-layer model," which estimates subjective quality from packet-header information [4], [5]. ITU-T standardized Recommendation P.564 [6] that determines the performance criteria for packet-layer models. The third type is the "parametric model," which takes quality design and management parameters as its inputs. The most widely used parametric model is the E-model that was standardized as ITU-T Recommendation G.107 [7]. The fourth type is the "bitstream layer model," which utilizes payload information (but not decoded media information). This is expected to improve the performance of a packet-layer model by reflecting the content-dependence of quality degradation [8]. The fifth type is the "hybrid model," which takes a combination of the above model inputs to utilize as much information as possible. ITU-T launched a new Question under Study Group 12 to standardize a hybrid model, which is provisionally called Recommendation P.CQO [9].

From the viewpoint of designing and managing multi-carrier services, it is highly important to divide the entire network into individual segments that can be operated by a single carrier, and accumulate the quality degradation within each segment to know the end-to-end QoE (Quality of Experience). Although the delay degradation can be accumulated by a simple sum of the delay in each segment, it is not straightforward to accumulate the effects of coding distortion and packet loss, which are primary quality factors in IP telephony. In particular, that becomes difficult when different segments use different speech coding schemes because coding distortion has a nonlinear nature and packet-loss degradation with different codecs cannot be estimated by simply accumulating the packet-loss rate in each segment.

The E-model uses a so-called "impairment factor framework [10]" to accumulate nonlinear distortion such

Manuscript received April 16, 2007.
Manuscript revised July 3, 2007.

[†]The authors are with NTT Service Integration Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
a) E-mail: takahashi.akira@lab.ntt.co.jp

DOI: 10.1093/ietcom/e90–b.12.3572

Table 1 Categories of objective models.

category	media-layer model	packet-layer model	parametric model	bitstream-layer model	hybrid model
input information	speech signal	RTP header and RTCP	quality design and management parameters	RTP payload information (without decoding)	combination of other model inputs
application	benchmarking	monitoring/management	planning and management	monitoring/management	monitoring/management
ITU-T Recommendation	Recs. P.862, P.563	Rec. P.564	Rec. G.107	–	Rec. P.CQO

as coding and packet-loss distortion. In this approach, the model first calculates the quality degradations in individual segments on a psychological scale, which is the equipment impairment factor called Ie,eff, and takes their sum over all the segments. The premise of this approach is that any degradation is additive on a psychological scale. This concept fits the planning purpose nicely because simple addition makes the network planning fairly easy. However, this additive property that the E-model assumes is sometimes questionable [7].

The E-model is a quality planning tool, so quality planning parameters are required as inputs. These parameters include the type of codec and packet-loss rate. However, it is sometimes difficult to directly obtain these values when one does not have access to IP-layer information. In such a case, the speech-layer objective measure, such as ITU-T Recommendation P.862 [2] "PESQ," is very useful because it only requires the speech signal. In particular, if we could estimate the end-to-end listening quality based on individual PESQ measurement results for multiple service segments, that would be very valuable in a multi-carrier environment.

The goal of our study is to establish an objective quality assessment method that can be applied to the evaluation of speech quality in a heterogeneous network environment, in which different networks use different speech-coding schemes.

In this paper, we first investigate the subjective quality evaluation characteristics for speech degraded by packet loss in multiple segments that employ different speech codecs. Then, we propose a method to accumulate individual quality degradations on a psychological scale in a nonlinear manner. The validity of the proposed model is evaluated in terms of the consistency between subjective quality and its objective estimation. Finally, we apply the proposed method to integrate multiple PESQ measurement results obtained for different segments.

2. Subjective Evaluation Characteristic and Its Modelling

This section first describes a subjective quality experiment in which various speech-coding schemes under various packet-loss conditions are examined. Next, the experimental results are investigated from the viewpoints of the effect of coding order and degradation accumulation on a psychological scale. Then, we propose a model for estimating the overall quality from segmental qualities.

2.1 Experimental Condition

In our investigation, we focus on the evaluation of coding distortion and packet-loss degradation, which are primary quality factors in IP-telephony services. In this experiment, we used G.711 with packet-loss concealment (PLC) [11], [12], G.729 [13], G.723.1 [14] at a bitrate of 6.3 kb/s, Adaptive Multi-rate (AMR) Codec at a bitrate of 12.2 kb/s [15]. We controlled the packet-loss rate between 0 and 5%.

Table 2 Coding and packet-loss conditions.

Cond. #	Codec #1	PLR #1	Codec #2	PLR #2
1-9	G.711PLC	1, 4, 8	G.729	1, 4, 8
10-18	G.729	0, 3, 5	G.723.1	0, 3, 5
19-27	G.729	0, 3, 5	AMR	0, 3, 5
28-36	G.729	1, 4, 8	G.711PLC	1, 4, 8
37-45	G.723.1	0, 3, 5	G.729	0, 3, 5
46-54	AMR	0, 3, 5	G.729	0, 3, 5
55-57	G.711PLC	1, 4, 8	-	-
58-60	G.729	0, 3, 5	-	-
61-63	AMR	0, 3, 5	-	-
64-66	G.723.1	0, 3, 5	-	-
67	G.711@64	0	-	-
68	GSM-EFR	0	-	-
69	G.726	0	-	-
70	G.728@16	0	-	-
71	G.729	0	-	-
72	G.726*2	0	-	-
73	G.728*2	0	-	-
74	GSM-FR	0	-	-
75	G.729*2	0	-	-
76	PDC-VSELP	0	-	-
77	G.726@24	0	-	-
78	G.729*3	0	-	-
79	GSM-FR*2	0	-	-
80	G.726@16	0	-	-
81-89	MNRU (Q=0 - 40 dB; 5 dB step)	-	-	-
90	MNRU (Q=99 dB)	-	-	-

A number of combinations of different speech coding schemes were tested, as shown in Table 2. In Table 2, we used the G.729 codec, which has been widely used in enterprise VoIP systems, as a "pivot." G.723.1 and AMP were used in the tests because they are often used in enterprise VoIP systems and 3G mobile systems, respectively. In Conditions 1 through 27, two different codecs are connected in tandem to simulate a heterogeneous network environment such as the interconnection of fixed and mobile networks. Conditions 28 through 54 are connections in tandem with sequences opposite from those in Conditions 1 through 27. Conditions 55 through 66 use a single codec. Conditions 67 through 80 are reference conditions defined in ITU-T Recommendation P.833 [16] for deriving the equipment impairment factor for unknown conditions. Conditions 81 through 90 are additional reference conditions generated by ITU-T Recommendation P.810 "MNRU (Modulated Noise Refer-

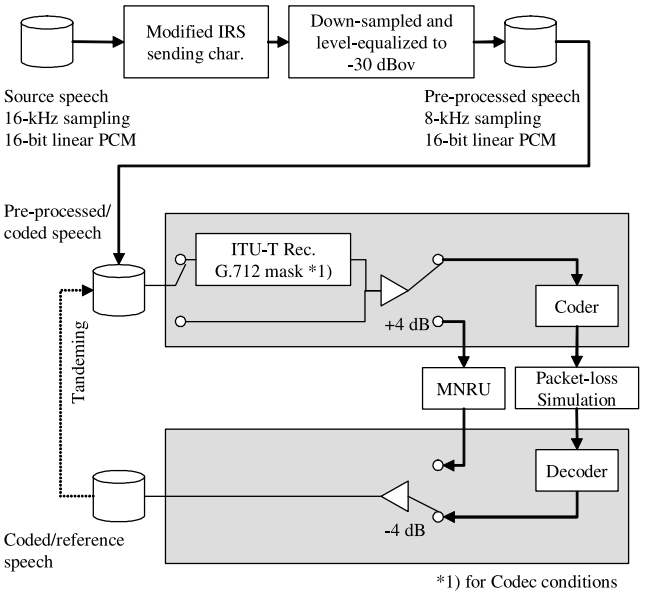


Fig. 1 Signal processing procedure.

Table 3 Experimental conditions.

Subject	20 male and 20 female
Source speech	Concatenated sentences (8 s)
Talker	2 male and 2 female
handset characteristics	modified IRS receiving [19]
ambient noise at receiving end	Hoth noise at 35 dB(A)
Listening level	−15dBPa

ence Unit) [17].” PLR stands for packet-loss rate†.

The signal-processing procedure is illustrated in Fig. 1. Source speech files are preprocessed by applying the modified IRS sending characteristics and the level is equalized to −30 dBov (dB relative to the digital overload level). Then, the preprocessed speech is passed through a prefilter defined in ITU-T Recommendation G.712 [18] and fed into a coder. Packet loss is randomly inserted in the coded bit-stream. We assume that the decoder always detects the packet loss and its packet-loss concealment algorithm works. For transcoding conditions, the coded speech with packet loss is again fed into a codec, as shown by the dotted line in Fig. 1.

The experimental conditions are summarized in Table 3. We used the 5-point ACR listening test method defined in ITU-T Recommendation P.800 [20]. The source speech was a Japanese sentence-pair with a duration of 8 s. For each coding and packet-loss condition shown in Table 2, we used four sentence-pairs spoken by different talkers. That is, we obtained 160 opinion votes (4 sentence-pairs * 40 subjects) for each condition.

2.2 Results

2.2.1 Order Effect

First, we investigate the effect of transcoding order. Reference [10] points out that the order of transcoding affects the resultant speech quality. That is, the speech quality for a condition where codec A is connected before codec B is sometimes different from that for a condition where codec A is connected after codec B. If this effect is significant, we need to take that into account in modelling the speech quality degraded by transcoding.

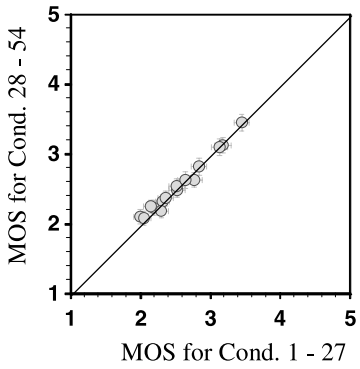


Fig. 2 Effect of coding order.

To investigate this issue, we compared the MOS for Conditions 1 through 27 with those for Conditions 28 through 54 in Table 2. The results are plotted in Fig. 2. The error bars indicate the 95% confidence intervals of subjective MOS. Although there are a few cases where statistically significant differences are observed, we concluded that the order effect is not so significant. Therefore, we decided not to model the order effect in our proposed method described in Sect. 2.3.

2.2.2 Additive Property

The quantities used in this section are illustrated in Fig. 3. In the following analysis, we transformed the Mean Opinion Score (MOS) obtained in the subjective experiment into Ie,eff, which represents the degree of degradation relative to the reference G.711 coding condition on a psychological scale, by applying the method provided in ITU-T Recommendation P.833 [16]. We call this “Ie,eff_{sub}” hereafter. In addition, we objectively calculated Ie,eff for each segment based on Eq. (1) defined in Recommendation G.107 [7]:

$$Ie,eff = Ie + (95 - Ie)\frac{Ppl}{\frac{Ppl}{BurstR} + Bpl}, \quad (1)$$

where Ie, Bpl, Ppl, and $BurstR$ represent the basic equipment impairment factor and packet-loss robustness factor for a given codec, which are provided in ITU-T Recommendation G.113 Appendix I [21]††; the packet-loss rate [%]; the burstiness parameter, which is fixed to “1” in this paper, respectively.

†The packet-loss rates for Codecs #1 and #2 are the same for Conditions 37 through 54.

††The Ie and Bpl values for AMR codec is not provided in Recommendation G.113. Therefore, we conducted another experiment to determine these values based on Recommendation P.833, and obtained Ie = 2.2, Bpl = 8.2 for the 12.2 kb/s mode of the AMR codec.

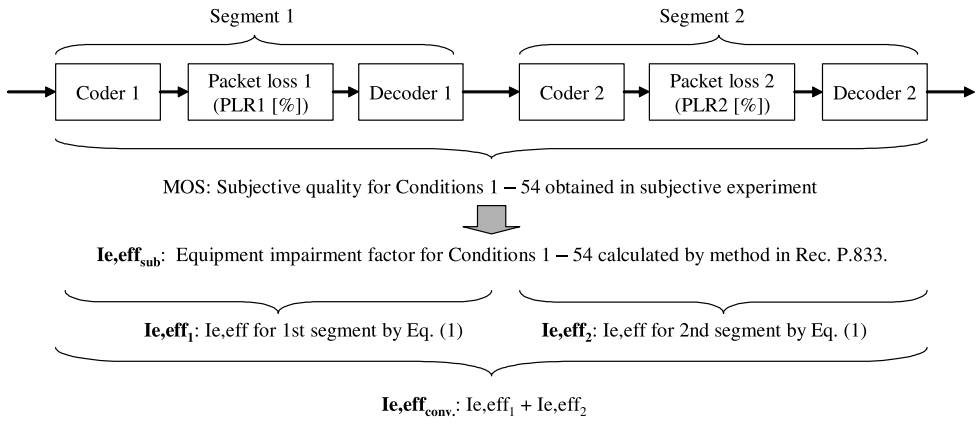


Fig. 3 Quantities used in analysis.

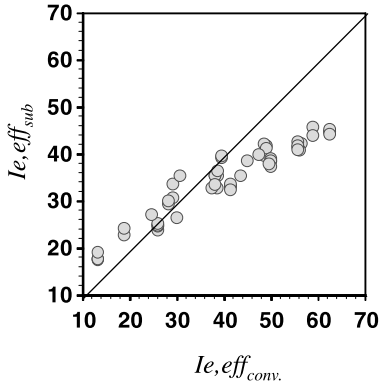


Fig. 4 Relationship between objectively estimated score by conventional method and subjectively evaluated score.

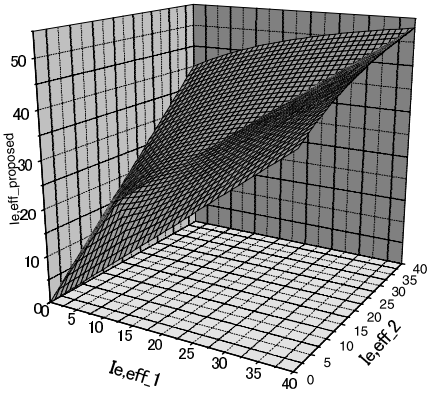


Fig. 5 Characteristics of proposed model.

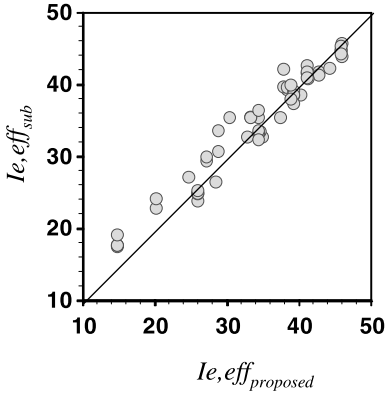


Fig. 6 Relationship between subjectively evaluated score and objectively estimated score by proposed method.

Conventionally, the end-to-end equipment impairment factor "$Ie, eff_{conv.}$" is defined as the sum of Ie, eff in each segment:

$$Ie, eff_{conv.} = Ie, eff_1 + Ie, eff_2. \quad (2)$$

Figure 4 demonstrates the relationship between Ie, eff_{sub} obtained in the subjective experiment and $Ie, eff_{conv.}$ estimated by the conventional method. Because the "Ie,eff" represents the degree of degradation, the more the value is, the lower the quality is. Therefore, this figure indicates that the end-to-end degradation is less than the simple sum of degradation in each segment. This is true especially for the low-quality region (i.e., Ie, eff_{sub} is more than 40). The Root Mean Square Error (RMSE) is 9.19.

2.3 Proposed Model

Taking into account the findings in the previous section, we propose the following model for estimating the end-to-end quality from the quality of individual segments. We found that there is no "order effect" in the subjective quality of transcoded speech. To the extent that the objective estimation for individual segments is consistent with the subjective quality, there should not be an order effect in the objective model either. This assumption is reasonable because the validity of "Ie,eff" has been well verified in standardizing ITU-T Recommendations. Therefore, the proposed formula is symmetric with respect to "Ie, eff_1" and "Ie, eff_2." The characteristics of the proposed model as a function of Ie, eff_1 and Ie, eff_2 are shown in Fig. 5. Eq. (3) was derived so that the combined Ie,eff would better fit the subjective quality evaluation characteristics. We observed that the combined effects of two codecs in tandem were weaker than the simple sum

of individual "Ie,eff" in two segments, so we tried to model this by this equation.

$$Ie,eff_{proposed} = max(Ie,eff_1, Ie,eff_2) + 26\frac{min(Ie,eff_1, Ie,eff_2)}{max(15, Ie,eff_1 + Ie,eff_2)} \quad (3)$$

Figure 6 demonstrates the performance of the proposed model. The quality estimated by the proposed model is shown to be fairly consistent with the actual subjective quality in comparison with Fig. 4. The RMSE is 2.06.

3. Validation of Proposed Model

In the previous section, we investigated the performance of the proposed model using the training data set. This section further validates the model in an application to an unknown data set.

3.1 Experimental Condition

The experimental conditions are listed in Table 4. We used G.711, G.729, AMR at a bitrate of 12.2 kb/s, and G.723.1 at a bitrate of 6.3 kb/s. In Table 4, G.711PLC, which has been employed in consumer IP-telephony services in Japan, was used as the primary codec. The packet-loss rate varies between 0 and 8%. Other testing conditions are the same as those in Sect. 2.1.

Table 4 Coding and packet-loss conditions (validation).

Cond. #	Codec #1	PLR #1	Codec #2	PLR #2
1-9	G.711PLC	0, 3, 5	G.729	0, 3, 5
10-18	G.711PLC	0, 3, 5	AMR	0, 3, 5
19-27	G.711PLC	0, 3, 5	G.723.1	0, 3, 5
28-36	G.729	0, 3, 5	G.711PLC	0, 3, 5
37-45	AMR	0, 3, 5	G.711PLC	0, 3, 5
46-54	G.723.1	0, 3, 5	G.711PLC	0, 3, 5
55-57	G.711PLC	0, 3, 5	-	-
58-60	G.729	0, 3, 5	-	-
61-63	AMR	0, 3, 5	-	-
64-66	G.723.1	0, 3, 5	-	-
67-80	Same as in Table 2			
81-89	MNRU (Q = 0 - 40 dB; 5 dB step)	-	-	-
90	MNRU (Q = 99 dB)	-	-	-

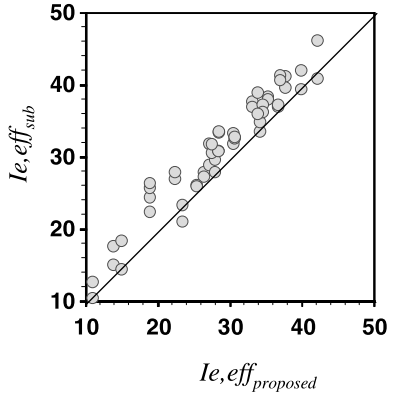


Fig. 7 Relationship between objectively estimated score by proposed method and subjectively evaluated score (Unknown data set).

3.2 Performance Analysis

Figure 7 demonstrates the quality estimation performance of the proposed model obtained using the unknown data in the subjective quality assessment described in the previous section. The RMSE became 3.07, indicating that the proposed method also works fairly well for an unknown data set. In addition, the proposed method can deal with other combinations of speech codecs in tandem than those tested in the training data set. The validation data is still limited in terms of codec variation and packet-loss conditions. For example, we need to further validate the model by using bursty packet-loss conditions because the packet loss was always inserted randomly in this investigation. Such issues are for further study.

4. Applicability to PESQ Evaluation

In the previous sections, we determined the segmental Ie,eff by using Eq. (1), assuming that Ie and Bpl values are known in advance. However, in some new applications, an unknown codec is used and it is difficult to prepare Ie and Bpl values for such a codec. In addition, it is often difficult to measure the packet-loss rate in a delay jitter buffer, whose impact on the end-to-end quality is not negligible. In these cases, exploiting speech-layer measurement, which requires neither *a priori* knowledge about the kind of codec nor the end-to-end packet-loss rate, is very effective.

In this section, we evaluate the validity of the proposed method in conjunction with the speech-layer objective quality measure recommended in ITU-T Recommendation P.862 [2], which is called PESQ (Perceptual Evaluation of Speech Quality). PESQ is the most widely used speech-layer objective-quality measure.

Figure 8 illustrates the flow chart to derive MOS_LQO based on segmental PESQ measurement. First, the PESQ value is calculated for each segment by using testing conditions 55 through 66 in Tables 2 and 4. Then, the average PESQ value taken over four talkers is converted to the Ie,eff value by applying ITU-T Recommendation P.834 [22]. Here, the Ie,eff value sometimes became negative, and we forced those values to zero. The resultant segmental Ie,eff values are integrated by Eq. (3) to generate the end-to-end Ie,eff. Because this represents the total degradation caused in two different segments, we subtract that from the reference R value indicating the maximum quality. Finally, the R value is mapped to the MOS_LQO scale based on ITU-T Recommendation G.107 Annex B.

The results are compared with the subjective MOS (MOS_LQS) in Figs. 9 and 10. As reference, the MOS_LQO estimated by applying Eq. (2) instead of Eq. (3) is also plotted in the figures (referred to as "conventional" in the fig-

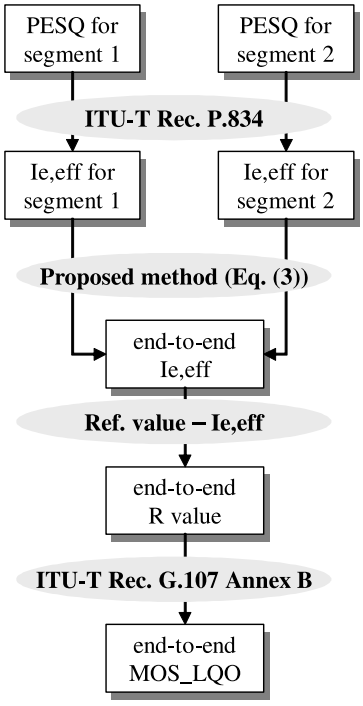


Fig. 8 Derivation of MOS_LQO based on segmental PESQ measurement.

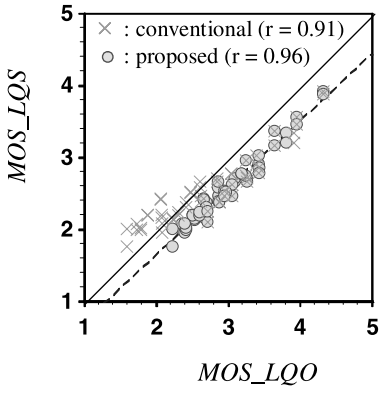


Fig. 9 Subjective MOS and its objective estimates by PESQ (Exp. 1).

ures).

From these figures, we claim that the proposed method also works well with PESQ, enabling the end-to-end quality estimation based on the individual quality measurement for each segment. The cross-correlation coefficients between subjective and objective MOS were 0.96 and 0.98, as shown in Figs. 9 and 10, respectively. In addition, the proposed method improves the accuracy of quality estimation in comparison with the conventional method, which is a simple sum of quality degradation in individual segments. The cross-correlation coefficients between subjective and objective MOS were 0.91 and 0.93, as obtained by the conventional method.

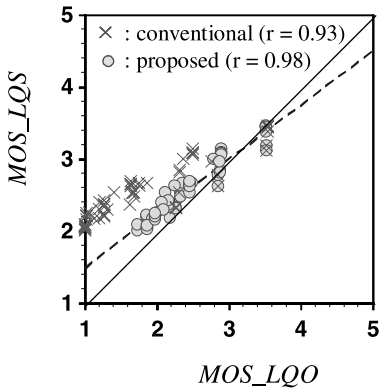


Fig. 10 Subjective MOS and its objective estimates by PESQ (Exp. 2).

5. Conclusion

Toward establishing quality-evaluation methodologies for speech communication over heterogeneous networks, this paper proposed a model for estimating end-to-end quality based on qualities in individual network segments.

First, we investigated the subjective quality assessment characteristics of multiple codecs connected in tandem and found that the order of codecs does not affect the subjective quality. In addition, we showed that the conventional method introduced in ITU-T Recommendation G.107 (E-model) has difficulty in evaluating the speech quality in a heterogeneous network environment.

Then, we proposed a method to accumulate quality degradation in each network segment and showed that the proposed method outperforms the conventional E-model approach. Moreover, we applied the proposed method to speech-layer objective quality assessment based on ITU-T Recommendation P.862 and verified its validity.

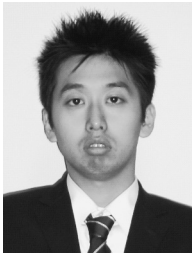
The method proposed in this paper is very useful in evaluating end-to-end VoIP quality even if networks using different speech-coding technologies are interconnected with each other. The method and findings described in this paper are a good starting point for improving the current E-model from the viewpoint of evaluating a multi-carrier environment.

References

[1] A. Takahashi, H. Yoshino, and N. Kitawaki, "Perceptual QoS assessment technologies for VoIP," IEEE Commun. Mag., vol.42, no.7, 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 narrow-band telephone networks and speech codecs," Feb. 2001.
[3] A. Rix, J.G. Beerends, D. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality — Technology and applications," IEEE Trans. Audio Speech and Language Processing, vol.14, no.6, pp.1890–1911, Nov. 2006.
[4] A. Clark, "Modeling the effects of burst packet loss and recency on subjective voice quality," IP Telephony Workshop, April 2001.
[5] S.R. Broom, "VoIP quality assessment: Taking account of the

edge-device," IEEE Trans. Audio Speech and Language Processing, vol.14, no.6, pp.1977–1983, Nov. 2006.
[6] ITU-T Recommendation P.564, "Conformance testing for narrow-band voice over IP transmission quality assessment models," July 2006.
[7] ITU-T Recommendation G.107, "The E-model, a computational model for use in transmission planning," March 2005.
[8] S. Uemura, N. Fukumoto, H. Yamada, and H. Horiuchi, "A note on a speech quality assessment based on lost packet discrimination," Proc. IEICE Gen. Conf. '07, B-11-2, March 2007.
[9] A. Takahashi, A. Kurashima, and H. Yoshino, "Objective assessment methodology for estimating conversational quality in VoIP," IEEE Trans. Audio Speech and Language Processing, vol.14, no.6, pp.1984–1993, Nov. 2006.
[10] S. Möller, Assessment and prediction of speech quality in telecommunications, Kluwer Academic Publishers, 2000.
[11] ITU-T Recommendation G.711, "Pulse code modulation (PCM) of voice frequencies," Nov. 1988.
[12] Appendix I to ITU-T Recommendation G.711, "A high quality low-complexity algorithm for packet loss concealment with G.711," Sept. 1999.
[13] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)," March 1996.
[14] ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," March 1996.
[15] ETSI TS 126 071 v6.0.0, "AMR speech codec; General description," Dec. 2004.
[16] ITU-T Recommendation P.833, "Methodology for derivation of equipment impairment factors from subjective listening only tests," Feb. 2001.
[17] ITU-T Recommendation P.810, "Modulated noise reference unit (MNRU)," Feb. 1996.
[18] ITU-T Recommendation G.712, "Transmission performance characteristics of pulse code modulation channels," Nov. 2001.
[19] ITU-T Recommendation P.830, "Subjective performance assessment of telephone-band and wideband digital codecs," Feb. 1996.
[20] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality," Aug. 1996.
[21] Appendix I to ITU-T Recommendation G.113, "Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl," May 2002.
[22] ITU-T Recommendation P.834, "Methodology for the derivation of equipment impairment factors from instrumental models," July 2002.

Akira Takahashi received a B.S. degree in mathematics from Hokkaido University in 1988, an M.S. degree in Electrical Engineering from the California Institute of Technology in 1993, and a Ph.D. in Engineering from the University of Tsukuba in 2007. He joined NTT Laboratories in 1988 and has been engaged in research into the quality assessment of speech and audio telecommunications. Currently, he is primarily working on the quality assessment of speech, audio, and video over IP networks. He has been contributing to ITU-T SG12 since 1994. He has been a co-Rapporteur of Question 13 of SG12, which studies QoE requirements and assessment methodologies, since 2005. Mr. Takahashi received the Telecommunication Technology Committee Award in Japan in 2004 and the ITU-AJ Award in Japan in 2005. He also received the Best Tutorial Paper Award from the IEICE Communication Society in Japan in 2006.



Noritsugu Egi received his B.E. and M.E. degrees in Electrical Communication Engineering from Tohoku University in 2003 and 2005. He joined NTT Service Integration Laboratories in 2005. Currently, he is researching subjective and objective quality assessment of telephony speech.

Atsuko Kurashima graduated from Tsukui High School and joined NTT Laboratories in 1980. She is currently researching objective quality assessment of telephony speech. Ms. Kurashima received the Electrical Science and Engineering Award in Japan in 2005.