Conversational Quality Estimation Model for Wideband IP-Telephony Services

Hitoshi Aoki, Atsuko Kurashima, Akira Takahashi

NTT Service Integration Laboratories
NTT, Tokyo, Japan
{aoki.hitoshi}{kurashima.atsuko}{takahashi.akira}@lab.ntt.co.jp

Abstract

As broadband and high-speed IP networks spread, IP-telephony services have become a popular speech communication application over IP networks. Recently, the speech quality of IP-telephony services has become close to that of conventional PSTN services. To provide better speech quality to users, speech communication with wider bandwidth (e.g., 7 kHz) is one of the most promising applications. To ensure desirable quality, we should design the quality before services start and manage it while they are being provided. To do this, an effective means for estimating users' perceptions of speech quality is indispensable. This paper describes a model for estimating the conversational quality of wideband IP-telephony services from physical characteristics of terminals and networks. The proposed model takes into account the quality enhancement effect achieved by widening speech bandwidth and has the advantage that it can evaluate the quality of both wideband and telephone-band speech on the same scale. Based on subjective conversational quality evaluation experiments, we show that the proposed model can accurately estimate the subjective quality for wideband speech as well as for telephone-band speech.

Index Terms: wideband speech, quality estimation, quality evaluation, conversational quality, IP-telephony

1. Introduction

Speech communication services over Internet Protocol (IP) networks have been growing through the spread of broadband Internet in recent years. In particular, the evolution of Voice over IP (VoIP) technology has attracted users' interests in high-quality speech communication services. That is, since IP-telephony services are achieving speech quality close to that of conventional PSTN services, users expect the quality to be enhanced so that they can get the benefits of broadband IP network services. Widening the speech bandwidth (to 7 kHz) is one of the most promising applications because wideband speech can enhance the speech intelligibility and naturalness in comparison with conventional telephone-band (300 - 3400 Hz) speech.

In order to provide customers with appropriate quality of service, effective means for design and management of the quality of speech communication services is needed. From this viewpoint, establishing a quality evaluation method is important.

The fundamental criterion for the quality of speech communication services is subjective quality, i.e., the user's perception of service quality, regardless of the speech bandwidth. Subjective quality is evaluated in psychological experiments, in which subjects evaluate the perceived quality of transmitted speech. The opinion test is widely used for subjective quality evaluation, and results derived from opinion tests are called the mean opinion score (MOS). In subjective quality evaluation, it is necessary to prepare exclusive test facilities such as acoustically shielded chambers so that the quality evaluation test is conducted stably and reproducibly. Moreover, many subjects are needed to appropriately evaluate various networks and terminal conditions. These make subjective quality evaluation time-consuming and expensive, although it is the most reliable method.

Therefore, a method that estimates subjective quality by measuring physical characteristics of terminals and networks is desirable. Such a method is called "objective quality assessment[1]." International Telecommunication Union (ITU) has already standardized objective quality assessment methods for telephone-band speech communications services such as ITU-T Recommendations P.862 "PESQ" and G.107 "E-model," which estimate subjective listening quality and subjective conversational quality, respectively. For wideband speech, although a wideband extension of PESQ has been standardized in ITU-T as Recommendation P.862.2, a method that estimates conversational quality has not been established yet.

In this paper, we first introduce our investigation of extending a telephone-band subjective conversational quality estimation model[2] to the evaluation of wideband speech communication services. Then, we describe subjective conversational quality assessment experiments used to validate the proposed model. Finally, we demonstrate the performance of the proposed model based on these experimental results, which are unknown for the model.

2. Proposed conversational quality estimation model

We previously developed a conversational quality estimation model specifically for telephone-band speech communication services[2]. This model is an opinion model[1] that integrates various quality factors on a common scale[3, 4]. In the development of this model, we investigated the effects of delay, talker echo, and speech distortion due to speech coding and packet loss, which are the primary quality factors in IP-telephony services. Based on the subjective quality evaluation characteristics of these factors, we took the following features into consideration in the model:

- proposing new equations representing the effects of delay and echo and
- quantifying the effect of interaction between delay and speech distortion.

Since the model has been developed based on the characteristics obtained from extensive subjective experimental data and achieved sufficiently high accuracy in terms of the consistency between subjective MOS and its estimates, we have come to the conclusion that this method is applicable not only to quality planning but also to quality benchmarking and management.

In extending its scope to the evaluation of wideband speech, there are two more issues that we must consider:

1. quantifying the quality enhancement effect achieved by widening the speech bandwidth and
2. determining the quality evaluation characteristics of various wideband speech codecs.

We propose the computational algorithm for estimating quality of wideband speech illustrated in Figure 1. The model takes four parameters, i.e., echo loudness, one-way delay, codec type, and packet-loss rate as inputs. Then it derives quality factors, such as echo impairment, delay impairment, and speech distortion, using input parameters. In the proposed model, the amount of quality impairment of each quality factor is defined on the psychological scale. Delay impairment and speech distortion are combined so that the model takes into account the interaction between the effects of delay and distortion. Next the proposed model integrates echo impairment and combined delay and distortion impairment and outputs an intermediate quality index. The final output of the proposed model is the estimated subjective conversational MOS mapped from the intermediate quality index.

In the following subsections, we discuss our solutions for the above-mentioned issues.

2.1. Quality enhancement effect by wideband speech

The proposed model uses a psychological scale R_n as an intermediate quality index, which was described in Ref [2]. For telephone-band speech, the maximum value of R_n is about 93. In this investigation, we kept the same R_n value for the conventional telephone-band speech. In other words, we added the difference between clean wideband speech quality and clean telephone-band speech quality as the quality enhancement effect of wideband speech to the maximum value (93) of the quality index for the telephone-band speech. In this way, we can ensure the compatibility between the evaluation results obtained by the originally proposed method and those obtained by its extension.

We conducted a subjective speech quality evaluation experiment, in which both wideband and telephone-band speech were evaluated in the same context[5]. This experiment gave MOS = 4.2 and MOS = 3.4 to clean wideband speech and clean telephone-band speech, respectively. By transforming these MOSs to the psychological scale R_n, we obtained $R_n = 86$ and $R_n = 67$, respectively. From this finding, we concluded that the quality enhancement effect achieved by widening speech bandwidth is 19 on the R_n scale. Therefore, in evaluating the wideband speech, we should subtract 19 from the speech distortion index to reflect the quality enhancement in the proposed model.

2.2. Quality evaluation characteristics of wideband speech codecs

The proposed method quantifies speech distortion caused by speech coding and packet loss as *Ie,eff* in the same manner as the E-model[6]. *Ie,eff* is defined by Equation (1) by using *Ie* representing the coding distortion, *Bpl* representing the robustness against packet loss, and *Ppl* representing the packet-loss rate.

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

Table 1: Codec and packet-loss rate conditions.

Bandwidth (Hz)	Codec	Bitrate (kb/s)	Packet-loss rate (%)
100 – 7000	G.722	64	0, 1, 3, 5, 10
	G.722.1	32	
		24	
	G.722.2	6.6	
		8.85	
		12.65	
		14.25	
		15.85	
		18.25	
		19.85	
		23.05	
		28.85	
	MNRU	(Q = 15, 20, 25, 30, 35, 45, 99 dB)	
300 – 3400	telephone-band reference conditions defined in Recommendation P.833		
	MNRU	(Q = 15, 20, 25, 30, 35, 45, 99 dB)	

Although the *Ie* and *Bpl* values for telephone-band speech codecs are provided in Recommendation G.113 Appendix I[7], there are only a few studies[8, 9] with respect to the *Ie* values for wideband codecs and none for the *Bpl* values. Therefore, we derived the *Ie* and *Bpl* values for typical wideband speech codecs standardized in ITU-T based on the subjective listening quality assessment.

Table 1 shows the subjective testing conditions. Since we basically followed the methodology defined in Recommendation P.833 in determining *Ie* and *Bpl*, we included the reference testing conditions consisting of telephone-band speech codecs such as G.711, G.726, and G.729. MNRU is a reference system defined in Recommendation P.810, and its quality is controlled by the Q-value, which represents the SNR of MNRU.

The procedures for determining *Ie* and *Bpl* based on the results of this subjective experiment are described below.

2.2.1. Experimental Ie,eff

First, we transformed the raw MOS values obtained in the experiment onto the *Rn* scale. To do this, we applied the transformation function provided in Annex B to Recommendation G.107. Second, these *Rn* values were scaled by a factor of $\alpha = 112/Rn\_direct$, where *Rn_direct* denotes the *Rn* value for wideband speech without any degradation. That is, this scaling forces the maximum possible *Rn* value to be 112, as we determined in the previous subsection. The "experimental *Ie,eff*" value is defined as the difference between the scaled *Rn* value for each condition and the scaled *Rn_direct*.

2.2.2. Removal of experimental bias

To remove the bias between the subjective quality obtained in this experiment and that in the reference experiment on which P.833 is based, we derived the "correction function" defined in P.833. This was done by deriving a linear regression function between the reference *Ie* value in P.833 and the experimental *Ie,eff* obtained in this experiment for the same telephone-band coding conditions. Here, we subtracted 19 from the experimental *Ie,eff* because we do not need to reflect the advantage of wideband speech.

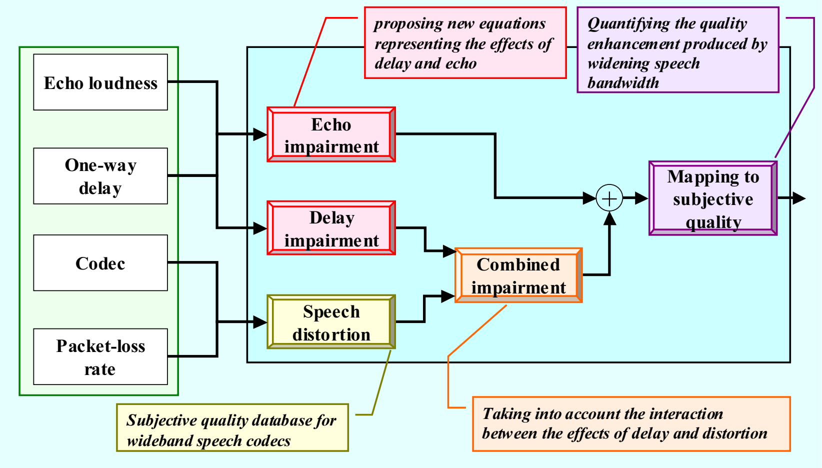
proposing new equations representing the effects of delay and echo
Quantifying the quality enhancement produced by widening speech bandwidth
Echo loudness
One-way delay
Codec
Packet-loss rate
Echo impairment
Delay impairment
Speech distortion
Combined impairment
+
Mapping to subjective quality
Subjective quality database for wideband speech codecs
Taking into account the interaction between the effects of delay and distortion

Figure 1: *Structure of the proposed model.*

Table 2: *Ie* and *Bpl* values for wideband codecs.

Codec	Bitrate (kb/s)	Wideband *Ie*	Wideband *Ie* (corrected)	*Bpl* (random)	*Bpl* (bursty)
G.722	64	9.50	9.50	5.19	5.76
G.722.1	32	15.04	15.04	14.77	12.70
	24	18.38	18.38	14.69	13.90
G.722.2	6.6	37.68	34.70	19.83	14.82
	8.85	27.09	27.58	25.31	18.96
	12.65	15.04	18.91	21.10	22.73
	14.25	14.30	16.02	20.95	15.27
	15.85	12.33	13.44	18.11	15.34
	18.25	9.34	10.02	15.05	13.32
	19.85	11.24	7.98	14.96	11.95
	23.05	6.01	4.35	12.77	10.80
	28.85	12.48	—	16.28	14.35

2.2.3. *Derivation of Ie and Bpl*

The experimental *Ie,eff* values for the packet-loss free conditions directly represent the *Ie* values for the associated coding conditions (see Eq. (1)). By applying *Ie* and *Ppl* values and solving Eq. (1) with respect to *Bpl*, we can derive the *Bpl* value for each packet-loss condition. Ideally, these *Bpl* values should be the same for the same codec. In practice, however, they differ due to experimental errors. In our investigation, we defined a *Bpl* value for each codec by taking the average of multiple *Bpl* values for multiple packet-loss conditions.

Table 2 summarizes the resultant *Ie* and *Bpl* values. We applied an exponential regression function to smooth the bitrate vs. *Ie* relationship for the G.722.2 codec. In this analysis, we excluded the *Ie* for a bitrate of 23.85 kbit/s because the *Ie* value for this condition seems to be an outlier. The smoothed *Ie* values are denoted by “Wideband *Ie* (corrected)” in the table. The *Bpl* values are derived separately for random and bursty packet-loss conditions.

3. Validation experiments

Since various degradations caused by terminals and networks often occur simultaneously in the real IP-telephony services environments, we need to validate the proposed model for such combined effects of various degradations. So we conducted extensive conversational subjective experiments. First we describe the subjective experiments carried out for the purposes of validation. We investigated the performance of the proposed model from the viewpoint of estimation accuracy, which is represented by the correlation between the subjective quality estimated by the proposed model and the actual subjective quality based on these conversational experiments.

3.1. Experimental conditions

Table 3 shows the subjective experimental conditions. The numbers of testing conditions in Experiments 1 and 2 were 41 and 43, respectively. In Experiment 1, we evaluated the combined effects of speech distortion, packet-loss, and delay. In Experiment 2, we evaluated the combined effects of speech distortion, packet-loss, delay, and talker echo. We used different subjects for Experiments 1 and 2. It might be necessary to extend the duration of conversation to reflect the characteristics of actual conversations over telephone services. We think, however, a duration of 60 seconds is sufficiently long for subjects to evaluate the effects of the degradations tested in our experiments.

3.2. Results

Figure 2 demonstrates the relationship between MOS estimated by the proposed model and that obtained in the actual subjective experiments. Although we optimized the mapping function from R_n to MOS by using these databases, they were not used in optimizing the derivation of R_n in the proposed model. In that sense, these databases are unknown data sets for the proposed method.

The figure indicates that the estimated MOS correlates well with the actual subjective MOS. The cross-correlation coefficient and root mean square error (RMSE) between estimated MOSs and subjective MOSs were 0.84 and 0.27, respectively. Taking into consideration that the mean of the 95% confidence interval of actual subjective quality evaluation experiments was 0.28, we con-

Table 3: Subjective testing conditions for experiments 1 and 2.

Common conditions	
Subjects	20 females and 20 males
Duration of conversation	60 s
Conversational task	Free conversation or guess the shape of a figure
Terminal	ITU-R Recommendation P.311 sensitivity/frequency characteristics handset
MNRU	Q = 12, 18, 24, 30, 36, 42, 48 dB
Experiment 1	
Codec	G.722 (64 kb/s), G.722.2 (23.05 kb/s), G.722.2 (8.85 kb/s)
Packet-loss rate	0, 1, 3, 5 %
Packet-loss patterns	random, bursty
Delay	100, 160, 200, 300, 400 ms
Echo (TELR)	65 dB
Experiment 2	
Codec	G.722 (64 kb/s), G.722.2 (23.05 kb/s)
Packet-loss rate	0, 3 %
Packet-loss pattern	random
Delay	100, 200, 400 ms
Echo (TELR)	35, 40, 45, 55, 60 dB

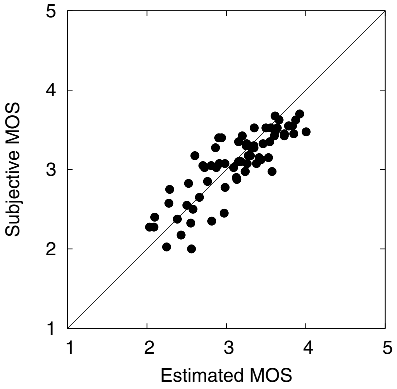

Figure 2: *Relationship between estimated and subjective MOS.*

clude that the proposed estimation model has sufficient estimation accuracy for practical use.

4. Conclusion

In this paper, we proposed a quality estimation model for wideband speech communication services. The proposed model is an extension of the quality estimation model for telephone-band speech communication services. The model can be applied not only to wideband speech but also telephone-band speech, enabling a direct comparison between these different media. A performance evaluation based on the extensive conversational subjective experiments demonstrated that the proposed model could predict the subjective quality for combined effects of various degradations even for unknown data sets. By applying the proposed model in designing the quality of networks and terminals, we can directly evaluate the advantage of a wideband speech communication service under consideration and how it can be improved. In addition, by deriving *Ie,eff* by using an objective quality measure such as Recommendation P.862.2 “Wideband PESQ,” it is possible to estimate the end-to-end conversational quality taking into account the performance of the actual codec implementation. We will further investigate the performance of the proposed model from such a viewpoint.

5. References

[1] A. Takahashi, H. Yoshino, and N. Kitawaki, “Perceptual QoS assessment technology for VoIP,”IEEE Commun. Mag., pp. 28-34, July 2004.

[2] A. Takahashi, “Opinion Model for estimating conversational quality of VoIP,” Proc. IEEE ICASSP 2004, vol. III, pp. 1072–1075, May 2004.

[3] N. Osaka, K. Kakehi, S. Iai, and N. Kitawaki, “A model for evaluating talker echo and sidetone in a telephone transmission network,” IEEE Trans. Commun., vol. 40, no. 11, pp. 1684-1692, 1992.

[4] N. O. Johannesson, “The ETSI computation model: A tool for transmission planning of telephone networks,” IEEE Commun. Mag., pp. 70-79, Jan. 1997.

[5] A. Takahashi, A. Kurashima, and H. Yoshino, “Subjective quality index for compatibly evaluating narrowband and wideband speech,” MESAQIN2005, June 2005.

[6] ITU-T Recommendation G.107, “The E-model, a computational model for use in transmission planning,” March 2005.

[7] ITU-T Recommendation G.113 Appendix I, “Provisional planning values for the equipment impairment factor Ie and packet-loss robustness factor Bpl,” May 2002.

[8] S. Möller and A. Raake, “Preliminary equipment impairment factor for wideband speech codecs,” ITU-T COM 12-D29-E, January, 2005.

[9] N. Kitawaki, “Ie and R values of wideband speech coding,” ITU-T COM 12-D77-E, October, 2005.

[10] ITU-T Recommendation P.833, “Methodology for derivation of equipment impairment factors from subjective listening only test,” February 2001.