PROPOSAL ON OBJECTIVE SPEECH QUALITY ASSESSMENT FOR WIDEBAND IP TELEPHONY

Chiharu Morioka, Atsuko Kurashima and Akira Takahashi

NTT Service Integration Laboratories
E-mail: morioka.chiharu@lab.ntt.co.jp

ABSTRACT

We investigated the performance of the objective speech quality assessment method called "wideband PESQ", which has been proposed to ITU-T for quality assessment of wideband speech, for quality ranging from narrowband to wideband speech with packet-loss degradation. The experimental results show that although the relationship between wideband PESQ and subjective quality depends on the codec for wideband speech, wideband PESQ can be used quite well as an objective quality measurement for speech with packet-loss degradation as long as the same codec is used. This codec dependence, which is especially observed when evaluating wideband codecs, makes it difficult to consistently estimate the quality of speech coded by various kinds of codecs. To avoid this difficulty, we propose correcting objective scores based on a function prepared for each codec. This correcting function is predetermined based on the results of a subjective quality experiment. We show the validity of the proposed method in evaluating the speech quality of various codecs with packet-loss degradation.

1. INTRODUCTION

IP telephony is one of the most promising applications over IP networks. 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 IP telephony services constantly and take necessary action to maintain the level of service.

The quality of service should be discussed in terms of subjective quality, which corresponds to users' perceptions of transmitted speech. However, since subjective quality assessment is time-consuming and expensive, a method that estimates subjective quality by measuring physical characteristics of the terminals and networks, called "objective quality assessment" [1] has been investigated.

The expansion of IP-telephony services will accelerate the market need for wideband speech communications in place of the conventional telephone-band speech communications. Not only wideband speech coding methods for a single bandwidth but also bandwidth-scalable coding methods, which change the speech bandwidth according to network conditions, are being developed [2][3]. To evaluate the quality of services provided using these coding methods, we need an objective speech quality assessment method for consistently estimating the quality of speech with various bandwidths. In addition, we also need a method of evaluating the trade-off between the quality enhancement achieved by widening the speech bandwidth and quality degradation caused by packet loss, for example.

For telephone-band speech services, Recommendation P.862 "PESQ" is used widely as an objective speech quality assessment method for estimating the listening quality. Its extension to the evaluation of wideband (7 kHz) speech, which is called "wideband PESQ" hereafter, was proposed [4]. Although the performance of wideband PESQ for wideband speech quality assessment was investigated [4], that for evaluating narrowband and wideband speech with packet-loss degradation on the same quality scale is not necessarily sufficient.

In this paper, we first describe the subjective quality experiment used for evaluating the performance of wideband PESQ in our investigation. Next, we discuss the performance of wideband PESQ from the viewpoint of estimating the quality of narrowband and wideband speech consistently. Finally, we propose a method of correcting the codec dependence observed in the evaluation of wideband codecs.

2. EXPERIMENTS

We conducted subjective and objective quality experiments, including narrowband and wideband speech conditions, to investigate the performance of wideband PESQ for quality ranging from narrowband to wideband speech.

0-7803-8874-7/05/$20.00 ©2005 IEEE

2.1. Source speech signals

We used Japanese speech in the database provided by [5]. We concatenated three different short sentences with some pause periods for each of four talkers (two males, two females).

The resultant speech samples (referred to as “source speech” hereafter) were 16 seconds long, with a speech activity rate of approximately 40%. This speech activity rate is almost equal to that in a channel of an actual telephone call. Accordingly, the probability that packet loss occurred during active speech parts in each speech sample was nearly equal to the probability in an actual telephone call.

2.2. Coding conditions

Table 2.1 shows the speech processing conditions in the subjective experiment. The codecs we used were i) Recommendations G.711 with PLC (the packet-loss concealment algorithm conforming to its Appendix I) and G.729 for narrowband speech and ii) G.722 (64 kbit/s mode), G.722.1 (32 kbit/s mode) and G.711 with PLC[*1] for wideband speech. Figures 2.1 and 2.2 show the speech processing for coding and packet-loss conditions.

2.3. Packet loss conditions

Table 2.1 also shows the packet length and packet-loss rate conditions for each codec. The packet-loss rate is defined as the proportion of lost packets out of all transmitted packets.

The packet-loss pattern was controlled by using the Discrete Gilbert Elliot Channel Model employed in Recommendation G.191 “Software tools for speech and audio coding standardization.” The burst ratio “b” in the model was 0.2 for the random packet-loss conditions and 0.8 for the burst ones. Four random packet-loss patterns and four burst packet-loss patterns were used for each codec and packet length pair in Table 2.1. A random packet-loss pattern and a burst packet-loss pattern were assigned to a group of eight or sixteen subjects.

2.4. Subjective experimental conditions

The subjective experiment including narrowband and wideband speech was conducted using the conventional listening ACR (Absolute Category Rating) method defined in ITU-T Recommendation P.800. In this experiment, the subjects were instructed to evaluate both narrowband and wideband speech samples as “telephone communication quality”. This implies that the subjects took into account the speech bandwidth, as well as coding and packet-loss distortion. Table 2.2 shows the conditions of the subjective experiment.

Table 2.1 Testing conditions.

Band	Codec	Packet length [ms]	Packet loss rate [%]
Narrowband	G.711PLC	10	1, 3, 5, 10
		20	1, 3, 5, 10
		40	1, 3, 5, 10
		100	1, 3, 5, 10
	G.729	10	1, 3, 5, 10
		20	1, 3, 5, 10
		50	1, 3, 5, 10
	G.711PLC	N/A	N/A
	G.729	N/A	N/A
	Narrowband MNRU (Q=5, 15, 25, 35, 40)	N/A	N/A
Wideband	Wideband G.711PLC[*1]	10	1, 3, 5, 10
		20	1, 3, 5, 10
		40	1, 3, 5, 10
	G.722	10	1, 3, 5, 10
		20	1, 3, 5, 10
		40	1, 3, 5, 10
	G.722.1	20	1, 3, 5, 10
		40	1, 3, 5, 10
		80	1, 3, 5, 10
	Wideband G.711PLC[*1]	N/A	N/A
	G.722	N/A	N/A
	G.722.1	N/A	N/A
	Wideband MNRU (Q=15, 25, 30, 35, 40)	N/A	N/A

Table 2.2 Subjective experimental conditions.

Testing methodology	Listening ACR test complying with Rec. P.800
No. of subjects	48 (24 male and 24 female)
Listening level	-15 dBPa
Receiving characteristics (narrowband)	Rec. P.830 modified IRS
Receiving characteristics (wideband)	ITU-R Rec. BS.1116
Ambient noise in listening room	none

[*1] We simply applied 16-kHz-sampled speech as input to Recommendation G.711. The frame size for speech analysis and the frequency range for searching the pitch in a frame were changed accordingly.

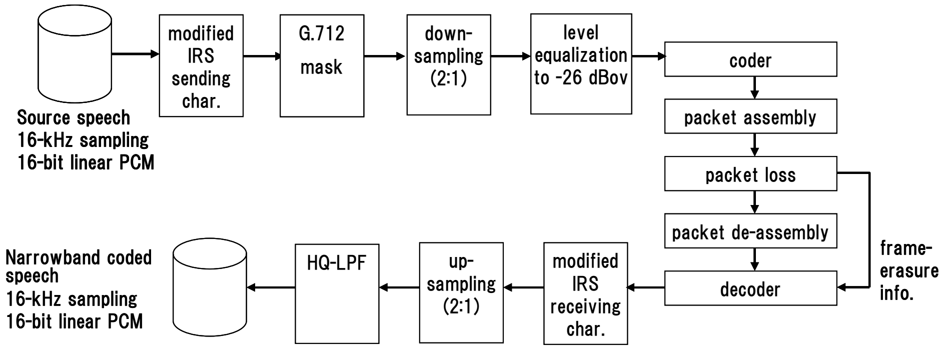

Figure 2.1 Signal processing for narrowband coding conditions.

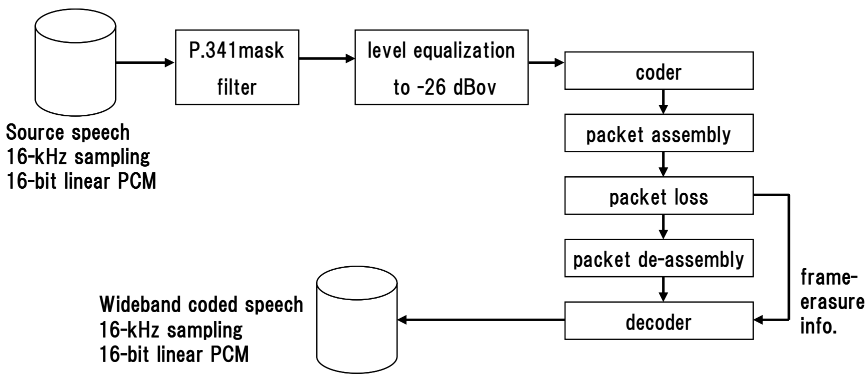

Figure 2.2 Signal processing for wideband coding conditions.

2.5. Objective experimental conditions

We calculated the objective quality value for each codec and packet-loss condition shown in Section 2 by using wideband PESQ. The resulting measured value is referred to as "WPESQ" hereafter. WPESQ was defined as the mean value for multiple speech samples used for the same codec and packet-loss condition shown in Table 2.1.

Since we investigated the performance of wideband PESQ for consistently estimating the quality of narrowband and wideband speech, a wideband speech sample was always used as the original input signal for wideband PESQ algorithm in this investigation.

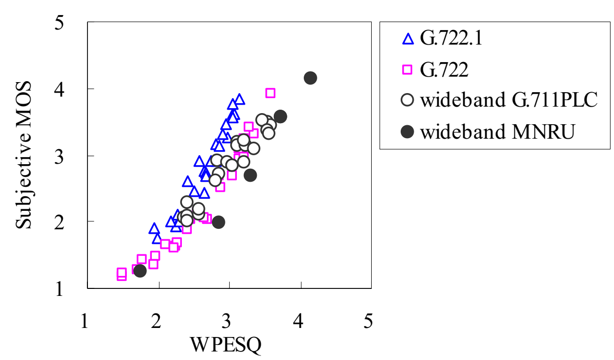

Figure 3.1 Relationship between subjective MOS and WPESQ for different wideband codec conditions

3. DISCUSSION

3.1. Quality estimation of wideband speech

Figure 3.1 demonstrates the relationship between subjective MOS and WPESQ for different wideband codec conditions. It shows that the relationship between subjective MOS and WPESQ depends on the wideband codec conditions. On the other hand, the consistency between subjective MOS and WPESQ is quite good for the same wideband codec conditions. Therefore, although the codec dependence makes it difficult to estimate the quality of wideband speech for any kind of codec, wideband PESQ can be used as an objective quality index for the wideband speech as long as the same codec is used.

3.2. Quality taking into account speech bandwidth

Figure 3.2 demonstrates the relationship between

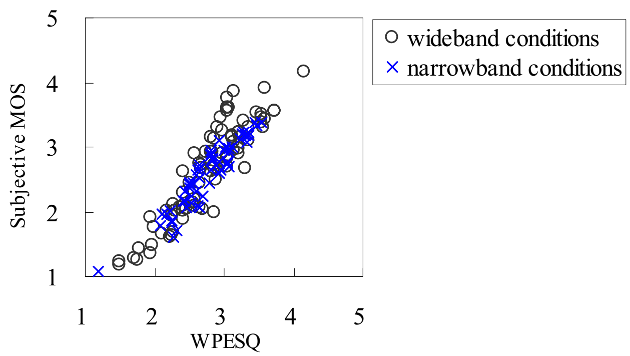

Figure 3.2 Relationship between subjective MOS and WPESQ

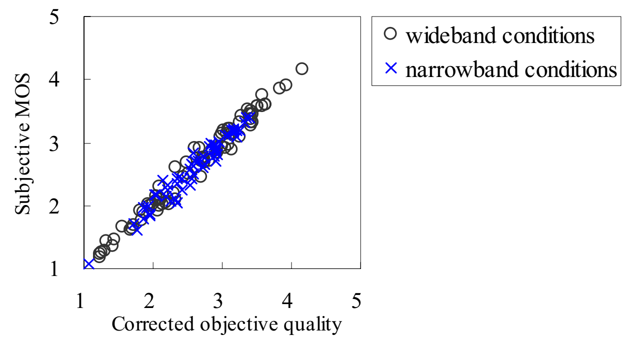

Figure 3.3 Relationship between subjective MOS and corrected objective quality

subjective MOS and its WPESQ for wideband and narrowband conditions. The consistency between subjective MOS and WPESQ is poor due to the codec dependence for the wideband conditions (c.f. section 3.1), as mentioned above. However, the correspondence between subjective MOS and WPESQ for narrowband conditions is good and quite similar to that for some wideband conditions. Therefore, if the variation due to the kind of wideband codec becomes smaller, it may be possible to estimate the quality of narrowband and wideband speech on the same scale using wideband PESQ.

4. PROPOSAL

We propose applying a correcting function optimized for each wideband codec. This makes it possible to compare the wideband speech quality coded by different codecs with packet-loss degradation using wideband PESQ. In our investigation, the correcting function was derived taking into account the relationship between subjective MOS and WPESQ for each wideband codec in the subjective experiment (c.f. section 3.1). WPESQ was transformed into the corrected objective quality value by using the correcting function (third-order polynomial function).

For narrowband speech, since the correspondence between subjective MOS and WPESQ is good (c.f. section 3.2), one correcting function common for all kinds of codecs was used.

Figure 3.3 demonstrates the relationship between subjective MOS and corrected objective quality value for narrowband and wideband speech conditions. Compared with Fig. 3.2, the variation for wideband conditions is much smaller. Therefore, if the kind of the wideband speech codec used in a system under evaluation is known and the speech quality of the codec is evaluated in a subjective quality experiment in advance, the proposed method works fairly well to consistently compare the speech quality with packet-loss degradation between different wideband codecs and between narrowband and wideband using wideband PESQ.

5. CONCLUSIONS

We investigated the performance of wideband PESQ for estimating the quality of narrowband and wideband speech with coding and packet-loss degradation. The experimental results show that the relationship between subjective MOS and WPESQ sometimes depends on the kind of codec for wideband speech, although wideband PESQ estimates the quality of narrowband speech for any codec.

Taking into account these findings, we proposed applying a correcting function for each wideband codec which is predetermined from a subjective experiment. We showed the validity of the proposed method by applying it to the evaluation of the wideband and narrowband speech quality with packet-loss degradation. This method lets service designers evaluate the speech quality enhancement achieved by widening the speech bandwidth under packet-loss conditions.

REFERENCES

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

[2] ISO/IEC 14496-3, "Coding of Audio-Visual Objects -Part 3: Audio," Dec. 1999.

[3] Y. Hiwasaki, T. Mori, H. Omuro, J. Ikedo, D. Tokumoto and A. Kataoka, "Scalable Speech Coding Technology for High-Quality Ubiquitous Communications," NTT Technical Review, vol. 2. no. 3, pp. 53-58, Dec. 2003.

[4] "Proposed modification to draft P.862 to allow PESQ to be used for quality assessment of wideband speech (BT, United Kingdom, and KPN, The Netherlands)," ITU-T SG12 Delayed Contribution COM-D007-E, Feb. 2001.

[5] "Multi-Lingual Speech Database for Telephonometry 1994," NTT Advanced Technology Corp., June 1994.