Objective Quality Assessment of Wideband Speech Coding using W-PESQ Measure and Artificial Voice

Nobuhiko Kitawaki, Kou Nagai, and Takeshi Yamada

University of Tsukuba
1-1-1, Tennoudai, Tsukuba-shi, 305-8573 Japan.
Tel. & Fax. +81 29 853 5526, E-mail:kitawaki@is.tsukuba.ac.jp

Abstract

An objective quality measurement methodology for wideband-speech coding has been studied, its essential components being an objective quality measure and an input test signal. Wideband-PESQ conforming to draft Recommendation P.862 has been studied as the objective quality measure. The Wideband-PESQ has been verified from the viewpoint of the consistency between subjectively evaluated MOS and objectively estimated MOS. Experimental results show that the correlation between them is strong, and the RMSE is relatively small. It is concluded that Wideband-PESQ is a promising measure for the objective quality assessment of wideband-speech coding. However, the mapping function from Wideband-PESQ values to subjective MOS requires further study.

This paper also describes the verification of artificial voice conforming to Recommendation P.50, as the input test signal for such measurements, by evaluating the consistency between the objectively estimated MOS using a real voice and that obtained using an artificial voice. Experimental results show that the correlation between them is very strong, and the RMSE is very small. It is concluded that an artificial voice can be used, in place of a real voice, for the objective quality assessment of wideband-speech coding.

Key Words: objective quality assessment, wideband speech, W-PESQ, artificial voice

1. Introduction

IP telephony was first introduced to provide free or inexpensive telephony services for Internet users. The quality was primarily based on the best effort type. With the increased number of Internet users and increased penetration of broadband access environments, the QoS of IP telephony has become a significant issue. In designing, providing and supporting networks and terminals for such services, it is important that this issue is addressed. We proposed an objective QoS measurement scheme involving an objective quality measure and a test speech signal for voiceband codecs in 1982, as shown in Fig.1 [1].

The coded speech quality at a low-bit-rate depends on the talker, because it contains less information than is possible with higher-bit-rate coding. Thus, for test purposes, real speech signals should be selected taking account of this dependency on the talker. In practice, an insufficient number of speech signals were used for the objective quality measurement. Therefore, we proposed another approach, which uses an artificial voice reflecting the average characteristics of the human voice as the test signal instead of a real voice [2]. The artificial voice was standardized as ITU-T Recommendation P.50 in 1988, as shown in Fig.2. This paper describes an artificial voice conforming to Recommendation P.50 for use in objective quality measurements of wideband speech coding, in terms of the consistency between the objectively estimated MOS using a real voice and that obtained using an artificial voice.

Recently, wideband speech communication has become increasingly necessary for use in advanced IP telephony using PCs, since for this hands-free communication using separate microphones and loudspeakers is indispensable, and in this situation wideband speech is particularly helpful in enhancing the naturalness of communication. This paper also evaluates Wideband-PESQ conforming to draft Recommendation P.862 as a method of providing an objective quality assessment of 7-kHz wideband-speech coding, as shown in Fig.3 [3].

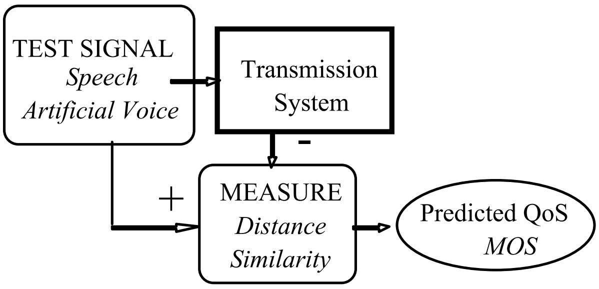


Fig.1 Objective QoS measuring scheme using an artificial voice and a measure as a criterion.

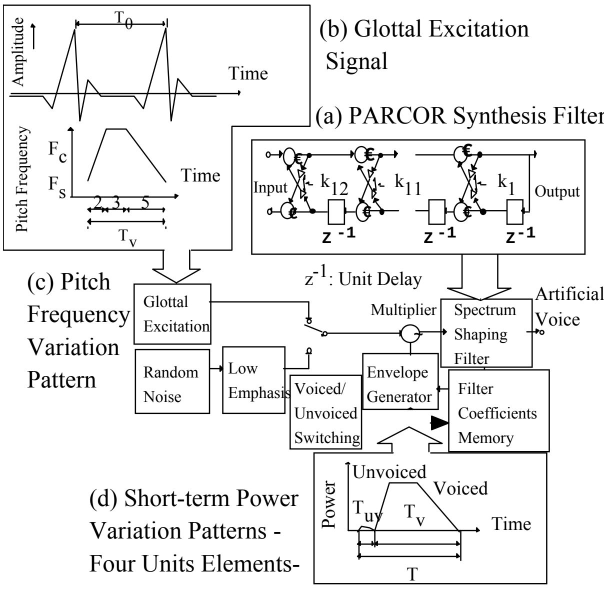


Fig. 2 Generation of the artificial voice (AV).

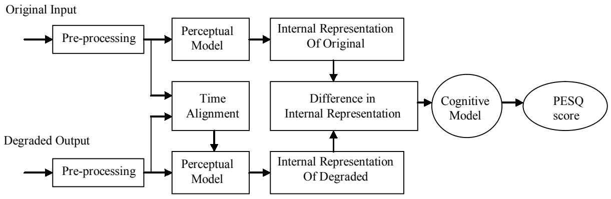


Fig.3 Concept of Objective Quality Assessment using PESQ Measure.

2. Subjective Quality Experiments

2.1 Subjective Assessment

Subjective quality assessment of wideband speech communication was performed in accordance with Recommendations P.800 and P.830. Table 1 shows the experimental conditions for the subjective quality assessment. The listening ACR method with 5 categories, complying with Rec. P.800, was used as the subjective quality assessment method, and the subjective MOS (Mean Opinion Score) was derived from the experimental results.

The test was set up in accordance with Rec. P.830. A Japanese speech database was constructed, which was composed of eight sentences spoken by 4 male and 4 female talkers. The speech material consisted of short, meaningful sentences, chosen at random and easy to understand. These sentences were assembled into a single set, as shown in such a way that there was no obvious connection of meaning between one sentence and the next. The sentences were group in pairs, with the duration of the pause between the first sentence and the second sentence set to be 1 second. Accordingly, 4 sentence pairs were constructed. In this subjective experiment, 2 male and 2 female speakers were selected and 2 sentence pairs from the database were used in each verification test.

Figure 4 shows the frequency characteristics of the test equipment used in the verification test of Wideband-PESQ and artificial voice. This study assumes use of the wideband frequency characteristics, with 7-kHz bandwidth, described in the proposed modification to draft P.862 to allow PESQ to be used for the quality assessment of wideband speech.

Table 1 Subjective assessment conditions.

Condition	Notes
Subjective assessment methodology	Listening ACR method complying to Rec. P.800
Test arrangement	Rec. P.830
Ambient noise in listening room	17.7 dBA
Number of subjects	32 (12 male and 20 female)
Number of talkers	2 male (M1,M2) and 2 female (F1, F2)
Number of sentence pairs	2
Frequency characteristics of test arrangement	7-kHz bandwidth (Fig.4) [3]

Fig.4 Frequency characteristics used in verification test of Wideband-PESQ.

2.2 Signal Processing

Table 2 shows the speech processing conditions used in the verification test. Several bit-rates and three different types of CODECs were used, namely, G.722 SB-ADPCM (Sub-band ADPCM) based on waveform coding, G.722.1 MLT (Modified Lapped Transform) based on transform coding, and G.722.2 AMR-WB (Adaptive Multi Rate-Wideband) based on CELP coding using linear prediction technology. In addition, we used MNRU conforming to Recommendation P.810 as a reference.

Table 2 Wideband-speech processing conditions.

CODEC	Bit-Rate/Q	Packet Length	Packet Loss Rate
G.722 SB-ADPCM	64 kb/s	20 ms	0, 1, 5, 10 %
	56 kb/s	20 ms	0, 1, 5, 10 %
	48 kb/s	20 ms	0, 1, 5, 10 %
G.722.1 MLT	24 kb/s	20 ms	0, 1, 5, 10 %
	32 kb/s	20 ms	0, 1, 5, 10 %
G.722.2 AMR-WB	15.85 kb/s	20 ms	0, 1, 5, 10 %
	23.85 kb/s	20 ms	0, 1, 5, 10 %
MNRU	Q=10 dB		
	Q=17 dB		
	Q=24 dB		
	Q=31 dB		
	Q=38 dB		
	Q=45 dB		

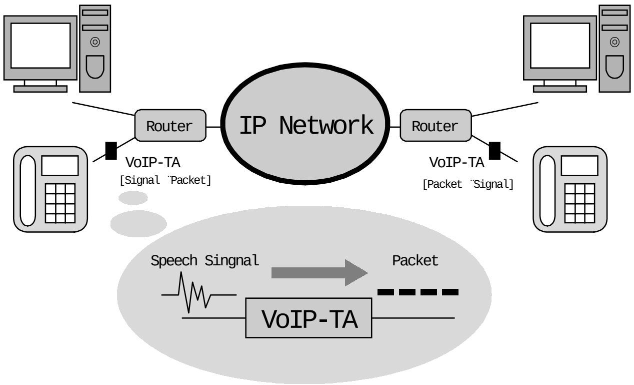


Fig.5 Test Configuration of IP network

Figure 5 shows the test configuration using an IP network. A 20 ms packet length was used for each CODEC. For each set of conditions, we used three different packet-loss rates, 1, 5, and 10%. The packet-loss pattern was controlled by using the Discrete Gilbert Elliot Channel Model that is specified in Recommendation G.191 “Software tools for speech and audio coding standardization.”

The total number of speech samples was 272 (that is, 8 speech pattern *((7 CODEC configurations * 4 packet loss rates + 6 MNRU settings)).

3. Objective Quality Measure

3.1 Mapping Function from Wideband-PESQ to Japanese MOS

This section examines the consistency between objective MOS and subjective MOS. Figure 6 shows the subjective MOS for each CODEC. However, the MOS obtained by the ACR method depends on the certain factors, such as a difference in instructions due to language translation, and mother tongue. For example, even under identical conditions there is a slight different between European MOS and Japanese MOS. Therefore, we tentatively used a function for mapping from Telephone-band PESQ to Japanese MOS proposed by NTT and represented by the following equation [4]. Note that the mapping function has not been verified for Wideband-PESQ.

MOS
= (0.804-4.690)/[1+ exp(PESQ-2.848)/0.680] + 4.690

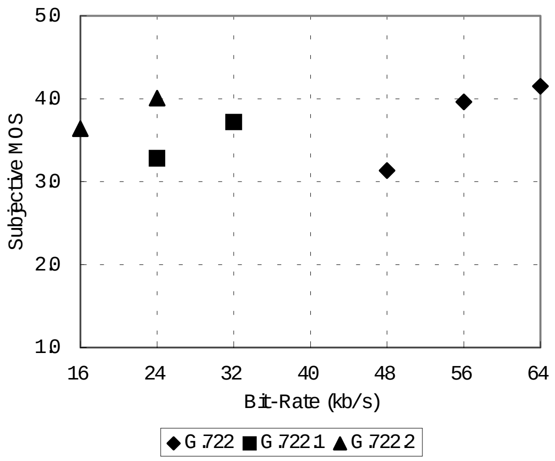


Fig.6 Subjective MOS for each CODEC.

3.2 Subjective MOS and Objective MOS

The Wideband-PESQ value was evaluated for each set of conditions (see Table 2) by applying real speech samples used in the subjective quality experiment. Figure 7 shows the relation between subjective MOS and objective MOS. Table 3 shows the correlation coefficients and the RMSE (Root Mean Square Error) to indicate the consistency between subjective MOS and objective MOS. This was done by first calculating the Wideband-PESQ value for each speech sample, and then averaging them for each test condition shown in Table 2.

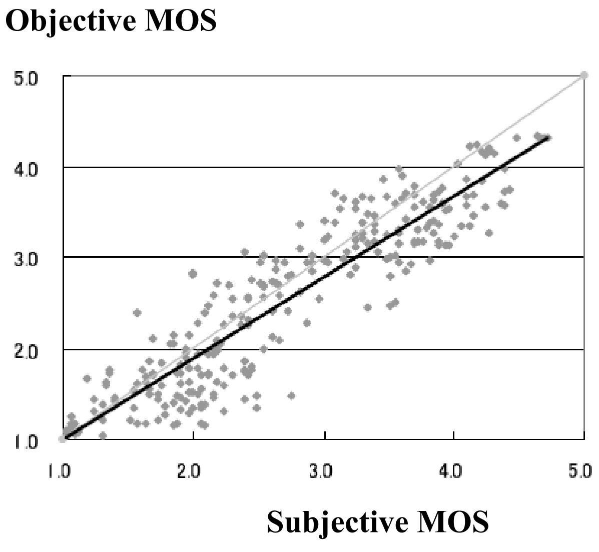


Fig.7 Relationship between objective MOS and subjective MOS.

In Fig.6 and Table 3, the subjective MOS is seen to correlate well with the objective MOS, although the absolute value for objective MOS is a little lower than that for subjective MOS. This may be caused by the use of the mapping function of telephone-band PESQ. This requires further study.

Table 3 Correlation coefficients and RMSE.

Talker	Coefficient	RMSE
Average	0.913	0.442
Talker M1	0.938	0.366
Talker M2	0.918	0.379
Talker F1	0.932	0.412
Talker F2	0.903	0.571
Average for males	0.927	0.373
Average for females	0.913	0.502

4. Artificial Voice

4.1 Processing of Real Voice

In this experiment for the verification of artificial voice, 2 sentences pairs were used. The number of talkers was 4 (2 male and 2 female). The frequency characteristics used in the verification test were the same as in the verification test of Wideband-PESQ, as shown in Fig.4.

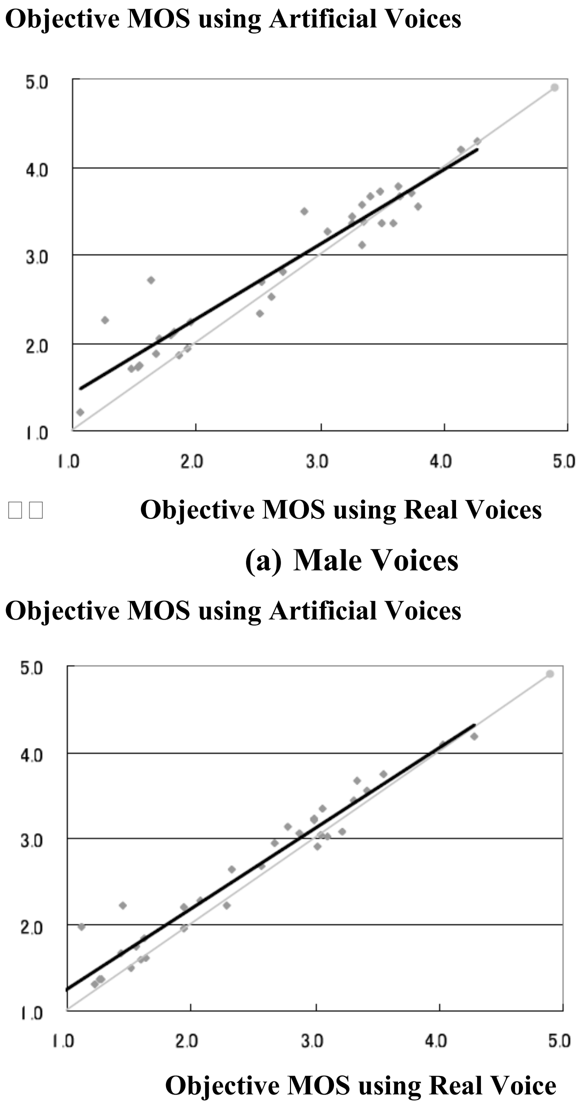
4.2 Signal Processing

The speech processing conditions used in the verification test were the same as in the verification test of Wideband-PESQ, as shown in Table 2. Several bit-rates and three different types of CODECs were used, namely G.722 SB-ADPCM (Sub-band ADPCM) based on the waveform coding, G.722.1 MLT (Modified Lapped Transform) based on the transform coding, and G.722.2 AMR-WB (Adaptive Multi Rate-Wideband) based on the CELP coding by linear prediction technology. In addition, we used MNRU conforming to Recommendation P.810 as a reference.

A packet length of 20 ms was used for each CODEC. For each condition, we used three different packet-loss-rate conditions, namely 1, 5, and 10%. The packet-loss pattern was controlled by using the Discrete Gilbert Elliot Channel Model that is used in Recommendation G.191 "Software tools for speech and audio coding standardization."

4.3 Objectively Estimated MOS by Real Voice and Artificial Voice

We calculated the Wideband-PESQ values for each male and female talker by applying the real speech samples used in the subjective quality experiment, and the artificial voices composed of a male and a female voice. Figure 8 shows the relation between real voice and artificial voice. Here, three artificial voices, generated by different initial condition, were used for each male and female voice [5]. The total number of real speech samples was 272 {namely, 4 talkers * 2 speech samples *(7 CODEC configurations * 4 packet loss rates + 6 MNRU settings)}, the equivalent figure for artificial voice was also 272 {namely, 2 talkers * 3 speech pattern *(7 CODEC pattern * 4 packet loss pattern + 6 MNRU pattern)}, and the total number of conditions was 34 for each male and female utterance.



(a) Male Voices

(b) Female Voices

Fig.8 Objective MOS using Real Voices and Artificial Voice.

Table 4 shows the consistency of objectively estimated MOS between real voice and artificial voice, expressed in terms of the correlation coefficients and the RMSE (Root Mean Square Error).

In Fig.8 and Table 4, the objective MOS using real voices correlates well with the objective MOS using artificial voice and the absolute value of RMSE is very small.

Table 4 Correlation coefficients and RMSE.

Talker	Coefficient	RMSE
Male	0.953	0.327
Female	0.972	0.264

5 Conclusion

This paper has described a verification of Wideband-PESQ drafted to allow PESQ to be used for the quality assessment of wideband speech. Three CODECs, conforming to such as G.722, G.722.1, and G.722.2, were used. It is concluded that Wideband-PESQ is a promising measure for the objective quality assessment of wideband-speech coding. However, the mapping function from Wideband-PESQ values to subjective MOS requires further study.

This contribution also describes the verification of artificial voice conforming to Recommendation P.50, as the input test signal, in terms of the consistency of objectively estimated MOS using a real voice and an artificial voice. Wideband-PESQ was used as an objective quality measure for three wideband CODECs, namely, G.722, G.722.1, and G.722.2. Experimental results show that the correlation between the MOS for real and artificial voices is very strong, and the RMSE is very small. Therefore, it is concluded that an artificial voice can be used in place of real voices in the objective quality assessment of wideband-speech coding.

References

[1] Nobuhiko Kitawaki, Kenzo Itoh, Masaaki Honda, and Kazuhiko Kakehi: “Comparison of Objective Speech Quality Measures for Voiceband CODECs,” IEEE Int. Conf. Acoust. Speech, Signal Processing, ICASSP’82, pp.S9.5.1-S9.5.4, 1982.

[2] Kenzo Itoh, Nobuhiko Kitawaki, Hiromi Nagabuchi, and Hiroshi Irii: “A New Artificial Speech Signal for Objective Quality Evaluation of Speech Coding Systems,” IEEE Trans. Commun., Vol.42, No.2/3/4, pp.664-672, 1994.

[3] Antony Rix, Andries P. Hekstra, and Mike P. Hollier: “Proposed modification to draft P.862 to allow PESQ to be used for quality assessment of wideband speech,” ITU-T COM12-D7, Study Period 2001-2004, February 2001.

[4] Hitoshi Aoki: “A method of determining mapping function from PESQ to MOS,” ITU-T COM12-D143, Study Period 2001-2004, September 2003.

[5] Nobuhiko Kitawaki: “Artificial voice for objective quality measurement of CELP coding and packet loss using PESQ measure,” ITU-T COM12-D191, Study Period 2001-2004, March 2004.