Objective Quality Evaluation Method for Noise-Reduced Speech

Noritsugu Egi, Hitoshi Aoki, and Akira Takahashi

NTT Service Integration Laboratories, NTT Corporation

egi.noritsugu@lab.ntt.co.jp, aoki.hitoshi@lab.ntt.co.jp, takahashi.akira@lab.ntt.co.jp

Abstract

We present a method for objective quality evaluation of noise-reduced speech. The experimental results indicate that residual noise and the distortion of speech and noise influence the perceived degradation of speech quality. Therefore, the proposed method is developed based on the relationship among three factors and subjective quality. We verify the validity of the method by comparing subjective quality of noise-reduced speech with its estimation. The verification results indicate that the correlation between subjective quality and the estimation is sufficiently high. Furthermore, the method can be used to perform assessments on a unique scale regardless of the type of noise.

Keywords

objective quality evaluation, subjective quality, noise reduction

1 Introduction

The spread of videoconference or mobile communication services is making hands-free speech communication more important. In such communication, however, speech is apt to be affected by background noise, and the speech quality is degraded. Listening to noisy speech is discomforting for users. Therefore, a noise reduction system is indispensable for comfortable speech communication in hands-free communication services.

Various noise reduction systems have been developed and implemented in various types of terminals for speech communication services (e.g., videoconference terminals and mobile phones). To provide high-quality speech communication services, selecting an appropriate noise reduction system and optimizing parameters for that are important. Therefore, we need to evaluate the speech quality of noise-reduced speech, which is the output of a noise reduction system.

The most reliable quality evaluation method is a subjective quality evaluation method in which users evaluate perceived speech quality. However, that is expensive and time consuming. Therefore, an objective quality evaluation method [1], which estimates subjective quality solely from physical characteristics of speech signals, is desirable for good efficiency.

In this paper, we present an objective quality evaluation method for noise-reduced speech considering subjective quality evaluation characteristics. The subjective quality evaluation test results indicate the need to consider residual noise and the distortion of speech and noise for accurate evaluation. Our method considers all of these so that the perceived quality of noise-reduced speech can be assessed accurately.

The remainder of this paper is organized as follows. In Section 2, we explain quality evaluation methods for noise-reduced speech. In Section 3, we present an objective quality evaluation method. In Section 4, we present a verification of the accuracy of our presented method. In Section 5, we present our conclusions.

2 Quality evaluation methods for noise-reduced speech

To provide high-quality speech communication services, evaluating the perceived speech quality is important. Quality evaluation methods represent the perceived speech quality quantitatively, and they are indispensable for quality design and management. The most reliable quality evaluation method is a subjective quality

evaluation in which subjects judge quality by listening to speech or conversation. The Mean Opinion Score (MOS) is often used as a quality scale in a subjective quality evaluation. The MOS directly represents the perceived speech quality. However, special facilities are required in a subjective quality evaluation experiment to ensure reproducible test results. In addition, subjective quality evaluation is expensive and time consuming. Therefore, an alternative to evaluating subjective quality is desirable. An objective quality evaluation method estimates subjective speech quality by analyzing physical characteristics of speech. The method does not need subjects for the evaluation, and the results are reproducible. Hence, the method can assess speech quality efficiently.

ITU-T P.862 "PESQ" [2] is an objective quality evaluation method. PESQ is the most widely used method for assessing listening speech quality objectively. There have been attempts to verify the accuracy of the perceived quality evaluation of noise-reduced speech obtained by PESQ [3]. In [3], four types of noise are used. For noise-reduced speech with the same noise type, the correlation between the results obtained by PESQ and the perceived speech quality is 0.85 – 0.94. However, when we evaluate speech quality with various noise types on the same scale, the correlation becomes 0.81. This means that the relationship between the results obtained by PESQ and perceived quality depends on the type of noise. Therefore, PESQ is inadequate for quality design and management. The objective quality evaluation method in [4] can assess subjective quality taking into account only the background noise in noise-reduced speech on the same scale regardless of the type of noise. However, the evaluation method cannot assess overall quality of noise-reduced speech.

That is, there is no objective quality evaluation method that is effective for various background noise conditions. Therefore, we investigate an objective quality evaluation method that assesses the perceived quality of noise-reduced speech on a single scale regardless of the type of noise.

3 Objective quality evaluation method

We developed an objective quality evaluation method in the following four steps.

(1) Selecting quality factors: We select some quality factors that are necessary for objective quality evaluation according to the result of a subjective listening test.

(2) Investigating effects of quality factors: We investigate the effects of each quality factors based on the relationship between the factors and perceived quality.

(3) Quantifying quality degradations: To define quantitatively the effect obtained in (2), we consider a method that calculates the amount of quality degradation in noise-reduced speech by physically analyzing signals that can be measured.

(4) Determining equation that estimates overall perceived quality: We produce an equation from the relationship between the quality indices obtained in (2) and the perceived quality.

These steps are described in detail below.

3.1 Selecting quality factors

The performance of noise reduction systems is often judged by the amount of residual noise, which influences perceived degradation of noise-reduced speech quality. Therefore, we regard the amount of residual noise as one of the quality factors of noise-reduced speech, but that is not the only factor. Noise reduction systems estimate background noise and subtract the noise from noisy speech. However, the background noise is not stationary, so the systems cannot be perfect, and they adversely affect the speech signal and the background noise signal. The effect distorts speech and noise, which degrades the perceived speech quality. Hence, we chose residual noise, speech distortion, and noise distortion as quality factors. We discuss whether these three quality factors are indispensable in terms of evaluating noise-reduced speech.

We performed a subjective listening test to determine the relationship between the three quality factors and perceived speech quality. In the experiment, subjects evaluated various noise-reduced speech samples. Two types of background noise were used: Hoth noise and noise recorded in an office, "office noise." Hoth noise is used to model indoor background noise when evaluating communication systems such as telephones. Detailed parameters of the test are shown in Table 1.

The noise-reduced speech samples were processed as illustrated in Fig. 1. The samples sound like noise-reduced speech. First, an original speech signal and an original noise signal were distorted by a noise reduction process based on the spectral subtraction algorithm [5]. The noise reduction level, which controls the degree of noise reduction, was changed to vary the degree of speech and noise distortion. When the level of noise reduction was higher, the residual noise was less, but the speech and noise were more distorted. Next, we adjusted the gain of the distorted speech to – 26 dBov.

Table 1 Subjective testing conditions

Subjects	32 (16 males and 16 females)
Original speech sample	•two short Japanese sentences (8 s) (2 males and 2 females)
Original noise sample	• Hoth noise (8 s) • office noise (8 s)
Signal bandwidth	7 kHz
Evaluation rating	ACR (Absolute Category Rating)
Listening equipment	Binaural headphone
Ambient noise at receiving side	Hoth noise at 40 dB(A)
Listening level	– 15 dBPa

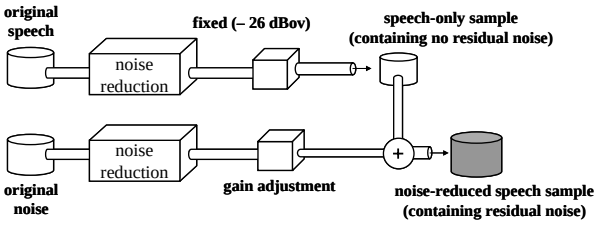

Fig. 1 Construction of noise-reduced speech samples

We call the adjusted distorted speech a "speech-only sample." In addition, we adjusted the gain of the distorted noise to change the amount of residual noise. The gain is between – 76 and – 56 dBov for Hoth noise and between – 66 and – 46 dBov for office noise. Finally, we added the distorted noise to the speech-only sample.

We prepared noise-reduced speech samples to determine the relationship between the quality factors and the perceived quality. In addition, we prepared speech-only samples to determine the relationship between only the speech distortion and the perceived quality. There are 400 noise-reduced speech samples, that is, 4 (speakers) * 2 (noise types) * 4 (speech distortions) * 5 (gains of noise)* 2.5 (2 or 3 noise distortions). There are 16 speech-only samples, that is, 4 (speakers) * 4 (speech distortions). We call both the noise-reduced speech samples and the speech-only samples "listening test samples."

The results of the subjective listening test are shown in Table 2. Only the results for noise-reduced speech samples that have Hoth noise are shown because the results obtained for both added Hoth noise and office noise are similar. In detecting the relationship between the factors and the perceived quality, we exclude the characteristics of the speaker. Therefore, in Table 2, the MOS is the average of 128 subjective data, that is, 32 (subjects) * 4 (speakers). We investigated the influence of the amount of residual noise, speech distortion, and noise distortion upon the subjective data by one-way analysis of variance. The results are shown in Table 3. Each P-value is less than 0.05, so we considered that the

Table 2 Results of subjective listening test (MOS)

speech distortion	residual noise: zero	small				large
zero	4.24	4.14 4.23	4.16 4.15	4.14 3.93	3.32 3.48	2.84 2.85
small	4.13	4.10 4.20 4.16	4.17 4.25 4.22	4.01 3.95 3.91	3.27 3.28 2.97	2.66 2.79 2.46
	3.17	3.09 3.10 3.06	3.13 3.07 2.98	3.03 2.95 2.71	2.70 2.44 2.13	2.28 2.00 1.91
large	2.30	2.20 2.21	2.23 2.19	2.24 2.12	2.03 1.88	1.83 1.64

noise distortion	
small	3.27
	3.28
large	2.97

Table 3 Results of one-way analysis of variance

(a) Factor: amount of residual noise

Objects: noise-reduced speech samples in which speech and noise distortion is zero

Source of variation	SS	df	MS	F	P-value
Between	188	4	47.1	2.39	< 0.001
Within	465	635	0.7		
Total	653	639			

(b) Factor: speech distortion

Objects: speech-only samples

Source of variation	SS	df	MS	F	P-value
Between	320	3	106.7	2.62	< 0.001
Within	363	508	0.7		
Total	683	511			

(c) Factor: noise distortion

Objects: noise-reduce speech samples in which the amount of residual noise and speech distortion is large

Source of variation	SS	df	MS	F	P-value
Between	2.25	1	2.25	3.88	0.020
Within	104	254	0.41		
Total	106	255			

SS: sums of squares , df: degrees of freedom, MS: mean squares

effects of these three factors are statistically significant. Therefore, we decided to use these three factors as quality factors for noise-reduced speech.

3.2 Investigating effects of quality factors

We observed the three principal relationships between these quality factors and the perceived speech quality as follows.

I. Residual noise

When the residual noise is smaller than a certain value, the amount of residual noise has little influence on the perceived speech quality (For each speech distortion, we investigated the influence of residual noise upon subjective data, as shown in the nonshaded area in

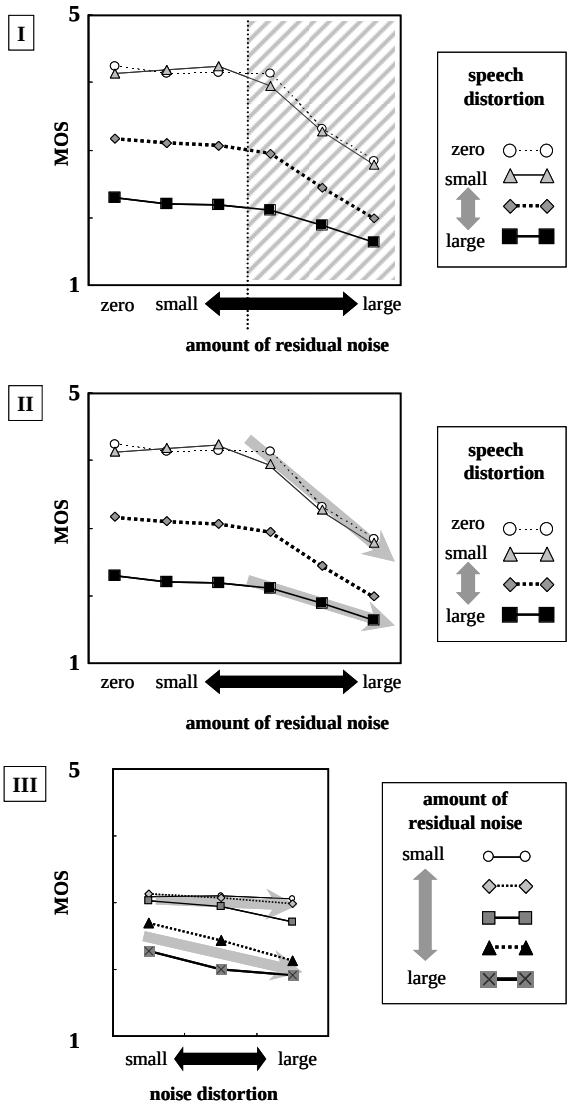

Fig. 2 Relationships between MOS and quality factors

Fig. 2 (I). The P-value is more than 0.4.). When the residual noise is larger than a certain value, the amount of residual noise has a large influence on the perceived speech quality (For each speech distortion, we investigated the influence of residual noise upon subjective data, as shown in the shaded area in Fig. 2 (I). The P-value is less than 0.001.).

II. Residual noise and speech distortion

The perceived speech quality is degraded much more by an increase in residual noise when the speech distortion is small, as illustrated in Fig. 2 (II).

III. Residual noise and noise distortion

Speech quality is degraded much more by an increase in noise distortion when the residual noise is large, as illustrated in Fig. 2 (III).

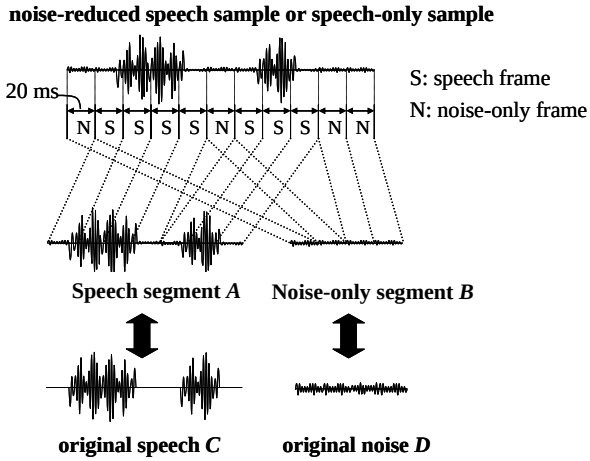

Fig. 3 Signals A, B, C, and D

3.3 Quantifying quality degradations

To represent the three quality factors determined in the previous section, we divided the listening test samples into speech segment *A* and noise-only segment *B* in 20-ms frames. We defined original speech *C* corresponding to speech segment *A* and original noise *D* corresponding to noise-only segment *B*. The relationships among signals *A*, *B*, *C*, and *D* are illustrated in Fig. 3.

We quantified the effect of each quality factor in the listening test samples by using signals *A*, *B*, *C*, and *D* as follows.

3.3.1. Speech distortion

The distortion measured from a comparison of signals *A* and *C* represents the speech distortion and residual noise in signal *A*. The louder residual noise is, the more the distortion exceeds the speech distortion. We calculated the speech distortion of the listening test samples based on these findings. The details are as follows.

First, we used ITU-T Recommendation P862.2 "Wideband PESQ" [6] for comparison of signals *A* and *C*. We call the value of its output *X*'. Next, we corrected *X*' to exclude the effect of residual noise. We supposed that the RMS level (dBov) of residual noise in signal *A* is nearly equal to that of residual noise in signal *B*. Then, we corrected *X*' based on the RMS level *Y*' (dBov) in signal *B*. The correction equation is Eq. 1, where k_1 and k_2 are constants. We regarded *X* as the effect of the speech distortion in the listening test samples.

$$X = \min\left(k_1, X' \Big/ \left(1 - k_2^{2^{(Y'/10)}}\right)\right) \qquad (1)$$

3.3.2. Amount of residual noise

The human hearing system has different sensitivities at

different frequencies. This means that the perception of noise is not equal at all frequencies. Noise in which energy is mainly in the high or low frequencies will not influence the perceived quality as it would when energy is mainly in the middle frequencies. Therefore we adjust the levels of signal *B* in 10 different frequency bands based on the equal loudness contours and call the adjusted signal *B*'. Then, a majority of the noise-reduced speech samples has a similar relationship between the RMS level of signal *B*' and the perceived degradation. For some noise-reduced samples, however, the RMS level of signal *B*' is bigger than that predicted by the relationship. Residual noise of some samples contains a few impulsive sounds such as the closing of a door and little stationary noise. Therefore, we defined the volume of noise in signal *B* to enable us to measure values smaller than the RMS level of signal *B*' for only such residual noise cases. We divided signal *B*' into several short segments and calculated values N_1, N_2, … , N_m of the RMS level (dBov) of each short segment. We calculated the value *Y* based on N_1, N_2, … , N_m using Eq. 2, where *l* is a natural number. We regarded *Y* as the amount of the residual noise in the noise-reduced speech samples.

$$Y = 10\log_{10}\left(\frac{\sum_{i=1}^{m} 10^{N_i/l}}{m}\right)^{l} \qquad (2)$$

3.3.3 Noise distortion

We measured the distortion of noise by comparing signals *B* and *D* by "Wideband PESQ," and called the output value *Z*. We regarded *Z* as the effect of noise distortion in noise-reduced speech samples.

3.4 Determining equation that estimates overall perceived quality

Based on the indices and effects of each quality factors, we determined Eq. 3 that estimates overall quality of noise reduced speech. In Eq. 3, *Q* is the estimated noise-reduced speech quality. *X*, *Y*, and *Z* are the effect of three quality factors. n_1, n_2, n_3, n_4, and n_5 are constants.

$$Q = \left(\frac{n_1}{1+e^{(n_2-n_3X)}}\right)\left(1-(n_4-n_5Z)^{2^{(-Y/10)}}\right)+1 \qquad (3)$$

The first term of Eq. 3 represents the effect of speech distortion and the second term of Eq. 3 represents the effect of residual noise and noise distortion. We determined the first term based on the relationship between speech distortion *X* and MOS for the noise-only samples as illustrated in Fig. 4.

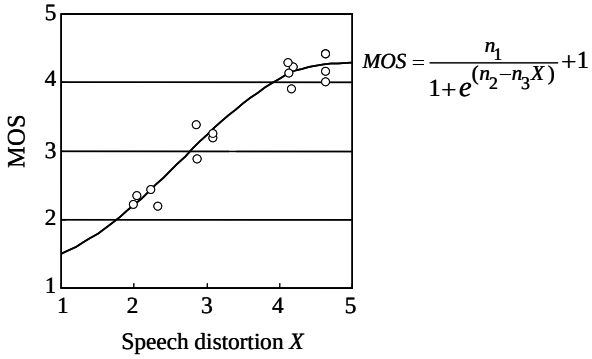

Fig. 4 Relationship between speech distortion X and MOS

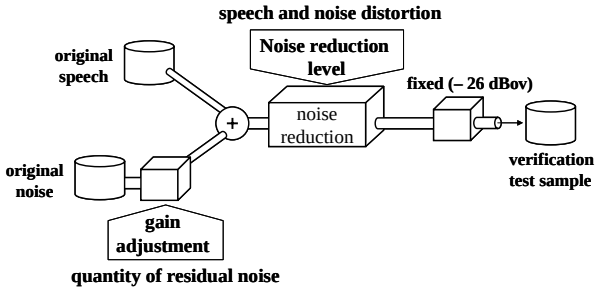

Fig. 5 Construction of verification test samples

In Eq. 3, first term is multiplied by second term to express the effect II in Subsection 3.2 that the effect of degradation by residual noise is much more when the speech distortion is small. In addition, second term expresses the effect I and III in Subsection 3.2.

Thus, Eq. 3 is determined based on the relationships among three quality factors and the perceived quality. So it enables to assess the perceived quality of noise-reduced speech accurately.

4 Verification of proposed method

We performed a subjective listening test to verify our method. The samples of noise-reduced speech were produced by passing noisy speech through a noise reduction system. We called the samples "verification test samples." We prepared various verification test samples, as illustrated in Fig. 5.

First, we changed the gain of the original noise to prepare verification test samples that have different quantities of residual noise. The gain is between – 86 and – 36 dBov. Next, we added the noise to the original speech and we distorted the combination by a noise reduction process. The noise reduction is the same as that used in Section 3. Then, we changed the noise reduction level to prepare verification test samples that have different speech and noise distortions. Finally, we

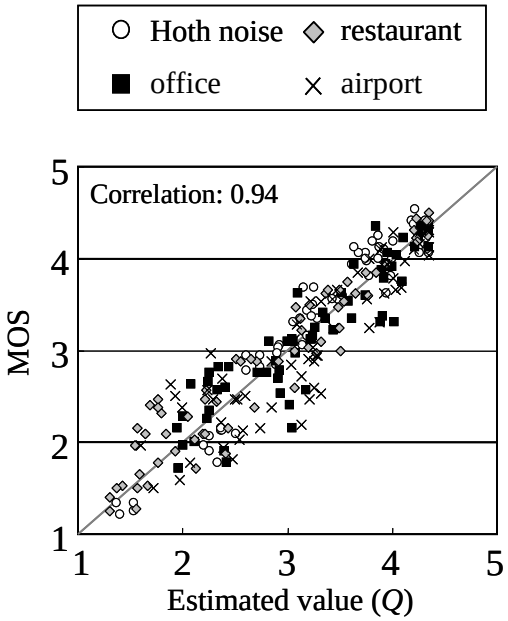

Fig. 6 Results of verification test

adjusted the level of the noise-reduced speech to – 26 dBov. We used Hoth noise, office noise, and noise recorded in a restaurant and an airport as the original noise, which contains more impulsive noise. There are 256 verification test samples, that is, 4 (speakers) * 4 (noise types) * 4 (noise reduction levels) * 4 (gains of noise). Other conditions of the test were the same as those in Table 1.

The relationship between subjective MOS and its objective estimated value *Q* is shown in Fig. 6. In Fig. 6, each MOS is the average of 32 subjective data. Good correlation (0.94) was obtained between the perceived speech quality and the estimated value *Q* regardless of the type of noise. Hence, we concluded that the proposed method works for Hoth noise, office noise, and more generic noise on a single scale.

5 Conclusion

We proposed an objective quality evaluation method for noise-reduced speech considering the effects of the quality factors that influence the perceived speech quality. Based on the subjective quality assessment, we verified that our method estimates the perceived quality of noise-reduced speech accurately on a single scale regardless of the type of noise. This means that the proposed method enables appropriate selection and parameter optimization of noise reduction systems.

Future work includes verifying that the method can evaluate the quality of noise-reduced speech in which the noise has been reduced by noise reduction algorithms other than the spectral subtraction method.

6 References

[1] A. Takahashi, H. Yoshino, and N. Kitawaki, "Perceptual QoS assessment technologies for VOIP," IEEE Communications Magazine, 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," February 2001.

[3] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective perceptual quality measures for the evaluation of noise reduction schemes," IWAENC2005, pp 169-172, September 2005

[4] V. Turbin and N. Faucheur, "A perceptual objective measure for noise reduction systems," Proc. MESAQIN2005, pp. 81-84, June 2005.

[5] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoustic Speech Signal Processing Vol. 27, No. 2, pp. 113-120, April 1979.

[6] ITU-T Recommendation P.862.2, "Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs," November 2005.