PAPER Special Section on Communication Quality

Objective Quality Evaluation Method for Noise-Reduced Speech

Noritsugu EGI[†a)], **Hitoshi AOKI**[†], ***and*** **Akira TAKAHASHI**[†], ***Members***

SUMMARY We present a method for the objective quality evaluation of noise-reduced speech in wideband speech communication services, which utilize speech with a wider bandwidth (e.g., 7 kHz) than the usual telephone bandwidth. Experiments indicate that the amount of residual noise and the distortion of speech and noise, which are quality factors, influence the perceived quality degradation of noise-reduced speech. From the results, we observe the principal relationships between these quality factors and perceived speech quality. On the basis of these relationships, we propose a method that quantifies each quality factor in noise-reduced speech by analyzing signals that can be measured and assesses the overall perceived quality of noise-reduced speech using values of these quality factors. To verify the validity of the method, we perform a subjective listening test and compare subjective quality of noise-reduced speech with its estimation. In the test, we use various types of background noise and noise-reduction algorithms. The verification results indicate that the correlation between subjective quality and its objective estimation is sufficiently high regardless of the type of background noise and noise-reduction algorithm.

key words: *objective quality, subjective quality, noise reduction, listening test, speech*

1. Introduction

Videoconference or mobile communication services have spread because hands-free speech communication technologies are available. In such communication, however, speech is apt to be affected by background noise. Consequently, speech quality is degraded because listening to noisy speech is annoying to users. Therefore, a noise-reduction system is indispensable for natural 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 and mobile terminals). To provide high-quality speech communication services, selecting an appropriate noise-reduction system and optimizing its parameters 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 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 is in the range of 1 (Bad) to 5 (Excellent) and represents perceived speech quality. However, performing a subjective quality evaluation is expensive and time consuming. In addition, special facilities are required in a subjective quality evaluation experiment to ensure reproducible test results. Therefore, an objective quality evaluation method [1], which estimates subjective speech quality by analyzing physical characteristics of speech, is desirable. Such a method does not need subjects for the evaluation, and the results are reproducible.

International Telecommunication Union Telecommunication Standardization Sector (ITU-T) Recommendation P.862 "PESQ" [2] is the most widely used method for objective quality assessment of listening speech quality. The PESQ calculates a distance between the original and degraded speech signals. The PESQ score is in the range of −0.5 to 4.5, although for most cases the output range will be between 1.0 and 4.5. There have been attempts to verify the accuracy of the perceived quality evaluation of noise-reduced speech estimated by PESQ [3], [4]. In [3], the results confirmed that PESQ correlates relatively well with the subjective MOS. Speech samples used in [3] are filtered by a function that is defined for the usual frequency range of the telephone bandwidth. However, ITU-T has standardized several wideband speech codecs such as Recommendations G.722 [5] and G.722.2 [6] for use in telecommunications. In addition, wideband coded speech is used more than telephone-band coded speech in hands-free speech communication because wideband coded speech is indispensable to achieve natural and intelligible conversation, which is a main advantage of hands-free speech communication. Therefore, an objective quality evaluation of noise-reduced speech should work well for noise-reduced speech in wideband speech communication. ITU-T Recommendation P.862.2 "Wideband PESQ" [7] is an extension to PESQ for use in measuring wideband speech quality. On the other hand, speech samples used in [4] are sampled at a rate of 16 kHz. In [4], four types of noise are used. When speech quality with various noise types on the same scale is evaluated, the correlation is relatively low. PESQ does not use data of background noise. Therefore, I guess that PESQ cannot measure the artificiality of residual noise caused by noise reduction when the artificiality becomes clearer in wideband speech communication. This means that the relationship between PESQ measurement and perceived quality depends on the type of noise. Therefore, applying PESQ to the assessment of noise-reduction systems with respect to various noise characteristics is not appropriate. Wideband PESQ does not use data of original noise, so it has the same

Manuscript received August 8, 2007.
Manuscript revised December 19, 2007.

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

DOI: 10.1093/ietcom/e91–b.5.1279

problem. The objective quality evaluation method in [8] can assess perceived degree of annoyance due to only the background noise in noise-reduced speech for various types of background noise. However, that method is not an overall quality evaluation method for noise-reduced speech.

Therefore, in this paper, we investigate an objective quality evaluation method that assesses the overall perceived quality of noise-reduced speech in wideband speech communication on a single scale regardless of the type of background noise and noise-reduction algorithm.

The remainder of this paper is organized as follows. In Sect. 2, we present an objective quality evaluation method for noise-reduced speech. Our method is based on results of a subjective listening test. The results indicate the need to consider residual noise and distortion of speech and noise for accurate assessment. Our method considers all of these so that it can assess the perceived quality of noise-reduced speech accurately. In Sect. 3, we present a verification of the accuracy of our method by using practical noise-reduced speech. We conduct extensive subjective quality experiments to investigate the consistency between subjective quality and its objective estimation by the proposed method. In Sect. 4, we present our conclusions.

2. 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. Quality factors are selected on the basis of the result of a subjective listening test.

2. Investigating effects of quality factors

   We investigate the effects of each quality factor obtained in Step 1 based on the qualitative relationship between the factor and perceived quality.

3. Quantifying quality degradations

   To define the effect obtained in Step 2 quantitatively, we consider a method that calculates the amount of quality degradation in noise-reduced speech by analyzing signals that can be measured.

4. Determining equation that estimates perceived quality

   We produce an equation from the relationship between the quality degradation obtained in Step 3 and the perceived quality.

These steps are described in detail below.

2.1 Selecting Quality Factors

One purpose of noise-reduction systems is to reduce background noise, which is annoying to users. Therefore, 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 that noise from noisy speech. However, background noise may be nonstationary, so systems cannot eliminate background noise completely, and they adversely affect speech and noise signals. The effect distorts speech and noise, which degrades perceived speech quality. Therefore, we regard speech distortion and noise distortion as quality factors of noise-reduced speech.

We discuss whether these three quality factors are indispensable in terms of evaluating noise-reduced speech. We performed a subjective listening test to determine relationship among the three quality factors and perceived speech quality.

2.1.1 Experimental Conditions

In the experiment, we prepared various "noise-reduced speech samples" to determine relationship among the three quality factors and perceived quality. Detailed conditions of this test are shown in Table 1. Noise-reduced speech samples, background noise, and the noise-reduction algorithm that we used in this test are described in detail below.

The process of the construction of noise-reduced speech samples is as follows (Fig. 1). Although the processing of practical noise-reduction systems is different from that depicted in Fig. 1, we adopted this configuration so that we can control the amounts of residual-noise level, speech distortion, and noise distortion to simulate various noise-reduced speech. First, an original speech signal [9]

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	8 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

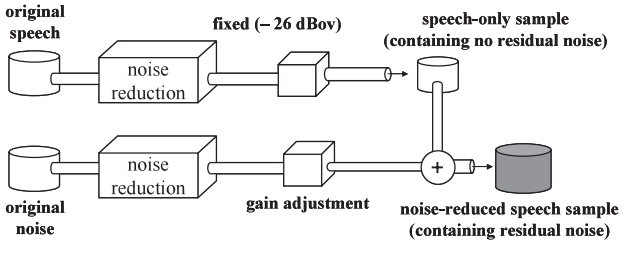

Fig. 1 Construction of noise-reduced speech samples.

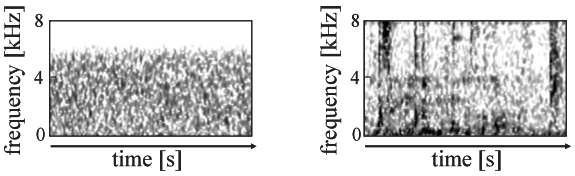

Fig. 2 Sound spectrograms of (a) Hoth noise and (b) office noise.

and an original noise signal [10] were distorted by a noise-reduction algorithm. The noise-reduction level, which controls the degree of noise reduction, was changed to vary the degrees of speech and noise distortions. 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 signal level of the distorted speech to −26 dBov. We call the adjusted distorted speech a "speech-only sample." In addition, we adjusted the signal level of the distorted noise to change the amount of residual noise. The level of the distorted noise was between −76 and −46 dBov. Finally, we added the distorted noise to the speech-only sample. There are 400 noise-reduced speech samples, that is, 4 (speakers) × 2 (noise types) × 5 (levels of noise) × 10 (pairs of speech distortions and noise distortions). In addition, we prepared speech-only samples to determine the relationship between the speech distortion and the perceived quality. There are 16 speech-only samples, that is, 4 (speakers) × 4 (speech distortions).

We used two types of background noise in this test. One is a stationary noise, and the other is a nonstationary noise. The stationary noise is Hoth noise that is used to model indoor background noise when evaluating communication systems such as telephones. The nonstationary noise is recorded in an office, which is called "office noise" hereafter. We show the sound spectrograms of Hoth noise and office noise in Fig. 2. Darkness of a spectrogram indicates amplitude of a particular frequency of noise, where darker points indicate larger amplitude.

We used a noise-reduction algorithm based on the spectral-subtraction (SS) method [11], which is very popular due to its relative simplicity and ease of implementation. The algorithm subtracts the averaged noise spectrum from the noisy signal spectrum. The averaged noise spectrum is estimated from the nonspeech part. Therefore, the algorithm does not work efficiently with nonstationary noise, resulting in an increase in speech and noise distortion.

2.1.2 Test Results

Results of the listening test are shown in Table 2. Only results for noise-reduced speech samples using Hoth noise are shown because results obtained for office noise are similar. In detecting the relationship between quality factors and perceived quality, we exclude the characteristic of effects of speaker. Therefore, in Table 2, the MOS is the average of 128 subjective data, that is, 32 (subjects) × 4 (speakers). We

Table 2 Results of listening test (MOS).

speech distortion \ residual noise	none	small				large
none	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 there is no speech and noise distortion

Source of variation	SS	df	MS	F	P-value
Between	188	4	47.1	64.4	< 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	149.2	< 0.001
Within	363	508	0.7		
Total	683	511			

(c) Factor: noise distortion

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

Source of variation	SS	df	MS	F	P-value
Between	9	2	4.72	8.74	0.02
Within	206	381	0.54		
Total	215	383			

SS: sums of squares , df: degrees of freedom, MS: mean squares, F: F test statistic

investigated the influence of the amount of residual noise, speech distortion, and noise distortion upon the subjective data by one-way analysis of variance. Results are shown in Table 3. Each P-value is less than 0.05, so we accept that there are statistically significant effects on perceived speech quality caused by these three factors. Therefore, we decided to use all three factors as quality factors for noise-reduced speech.

2.2 Investigating Effects of Quality Factors

In the experiment described in Sect. 2.1, we observed three principal relationships among these quality factors and perceived speech quality as follows (Fig. 3).

I. Residual noise

When the amount of residual noise is smaller than a certain value, the amount of residual noise has little influence on perceived quality, that is, perceived speech qual-

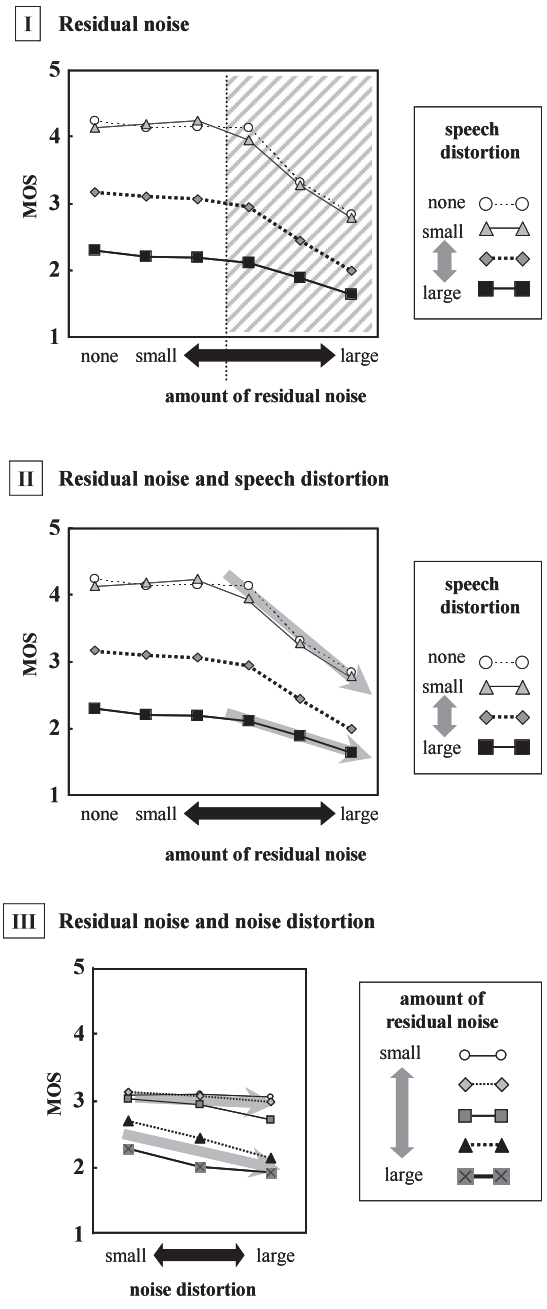

Fig. 3 Relationships between MOS and quality factors.

ity is influenced by speech distortion only, as shown in the nonshaded area in Fig. 3(I). When the amount of residual noise is larger than a certain value, the amount of residual noise has a large influence on perceived quality, as shown in the shaded area in Fig. 3(I).

II. Residual noise and speech distortion

Perceived speech quality is degraded much more by an increase in residual noise when the speech distortion is small, as illustrated in Fig. 3(II).

III. Residual noise and noise distortion

Perceived quality is degraded much more by an increase in noise distortion when residual noise is large, as illustrated in Fig. 3(III).

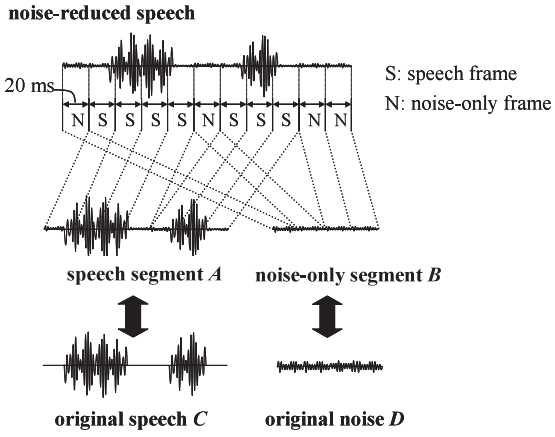

Fig. 4 Signals *A*, *B*, *C*, and *D*.

2.3 Quantifying Quality Degradations

In this section, we show a method that calculates the amount of each quality factor in noise-reduced speech to represent quantitatively the three principal relationships determined in the previous section. To measure the amount of each quality factor related to speech and that related to noise separately, we divided the noise-reduced speech used in the experiment in Sect. 2.1 into speech segment *A* and noise-only segment *B* in 20-ms frames. This division is executed on the basis of the signal level of the original speech. If the signal level of 20-ms frames exceeds a threshold level, the corresponding frame of speech segment *A* is regarded as a speech frame. Otherwise the corresponding frame of speech segment *A* is regarded as a noise-only frame. In the future, we will adopt more sophisticated speech activity detection technologies [12], [13] for avoiding misjudgment. We defined original speech *C* corresponding to speech segment *A* and original noise *D* corresponding to noise-only segment *B*. Relationships among signals *A*, *B*, *C*, and *D* are illustrated in Fig. 4.

We quantified the amount of each quality factor in noise-reduced speech by using signals *A*, *B*, *C*, and *D* as follows.

2.3.1 Speech Distortion

The distortion was measured from a comparison of signals *A* and *C*, where *A* represents three components: speech distortion, residual noise added to signal *C*, and signal *C*. When residual noise is louder, the distortion exceeds the speech distortion. We calculated the speech distortion of noise-reduced speech based on these findings. The details are as follows.

First, we used ITU-T Recommendation P.862.2 "Wideband PESQ" [7] for the 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 Root Mean Square (RMS) level (dBov) of residual noise in signal A is nearly equal to that of residual noise in signal B. Then, we corrected X' on the basis of the RMS level Y' (dBov) in signal B. The equation used for the correction is Eq. (1), where k_1 and k_2 are constants. k_1 represents the Wideband PESQ score of a clear speech sample and k_2 is determined to minimize the Root Mean Square Error (RMSE) of estimated X with respect to the Wideband PESQ score.

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

We regarded X as the effect of speech distortion in noise-reduced speech samples. The Wideband PESQ score is large when the quality distortion is small. That is, when the value of X is larger, speech distortion is smaller.

2.3.2 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 high or low frequencies will not influence perceived quality as it would when energy is mainly in middle frequencies. Therefore, we adjust the levels of signal B in 10 different frequency bands based on equal loudness contours and call the adjusted signal B'. Then, a majority of noise-reduced speech samples has a similar relationship between the RMS level of signal B' and the perceived degradation. However, for some noise-reduced samples, which contain a few impulsive sounds such as the closing of a door and little stationary noise, the RMS level of signal B' is bigger than that predicted by the relationship. 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 $N_1, N_2, \ldots, N_m$ of the RMS level (dBov) of each short segment. We calculated the value Y on the basis of $N_1, N_2, \ldots, N_m$ using Eq. (2), where l is a natural number. l is determined in the next section. We regarded Y as the amount of the residual noise in noise-reduced speech samples.

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

2.3.3 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. When the value of Z is larger, noise distortion is smaller.

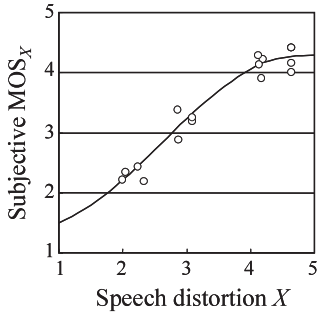

Fig. 5 Relationship between MOS$_X$ and speech distortion X.

2.4 Determining Equation that Estimates Overall Perceived Quality

Based on the degradation of each quality factor and relationships among them, we determined an equation that estimates overall quality of noise-reduced speech.

For noise-reduced speech samples that have speech distortion X, we defined MOS = MOS$_X$ − E, where MOS$_X$ is a subjective MOS of speech-only samples, which have speech distortion X. From Factors I–III in Sect. 2.2, E depends on speech distortion X, amount of residual noise Y, and noise distortion Z. We describe methods that calculate MOS$_X$ and E, as follows.

First, we model the relationship between X and MOS$_X$, as illustrated in Fig. 5. In Fig. 5, each point represents a relationship between speech distortion X and subjective MOS$_X$ for one speech-only sample. Based on the relationship, we determined Eq. (3) that calculates MOS$_X$ from speech distortion X, where n_1, n_2, and n_3 are constants. These constants are determined to minimize the RMSE of estimated MOS$_X$ with respect to subjective MOS$_X$.

$$\mathrm{MOS}_X = \frac{n_1}{1 + e^{(n_2 - n_3 X)}} + 1 \tag{3}$$

The curve in Fig. 5 represents Eq. (3).

Next, we show the method that calculates E. In Factor I (Residual noise) of Sect. 2.2, we note that E nearly equals 0 when Y, which represents the amount of residual noise, is smaller than a certain value and E is increased much more by an increase in Y when Y is larger than a certain value. In Factor II (Residual noise and speech distortion) of Sect. 2.2, we note that E is large when the speech distortion is small, that is, X and MOS$_X$ are large if noise-reduced speech samples have the same Y and Z. In Factor III (Residual noise and noise distortion) of Sect. 2.2, we note that E is increased much more by an increase in noise distortion Z when Y is large. Based on these findings, we determined Eq. (4) that calculates E, where n_4 and n_5 are constants. These constants and l of Eq. (2) are determined to minimize the RMSE of estimated MOS with respect to subjective MOS.

$$E = (\mathrm{MOS}_X - 1)(n_4 - n_5 Z)^{2^{(-Y/10)}} \tag{4}$$

From Eqs. (3) and (4), we show Eq. (5) that calculates overall quality of noise-reduced speech.

$$\begin{aligned} \mathrm{MOS} &= \mathrm{MOS}_X - (\mathrm{MOS}_X - 1)(n_4 - n_5 Z)^{2^{(-Y/10)}} \\ &= (\mathrm{MOS}_X - 1)(1 - (n_4 - n_5 Z)^{2^{(-Y/10)}})) + 1 \\ &= \left(\frac{n_1}{1 + e^{(n_2 - n_3 X)}}\right)\left(1 - (n_4 - n_5 Z)^{2^{(-Y/10)}}\right) + 1 \quad (5) \end{aligned}$$

Equation (5) takes into account the relationships among three quality factors and perceived quality. However, there could be other quality factors. In addition, verification of this method is insufficient for practical noise-reduced speech. Therefore, in the next chapter, we will verify our method from these standpoints.

3. Verification of Proposed Method

We performed subjective listening tests to verify our method by using unknown data. Samples of noise-reduced speech were produced by passing noisy speech through a noise-reduction system. That is, these samples are practical noise-reduced speech. We prepared various verification test samples using the scheme illustrated in Fig. 6. In this test, we used different original speech samples in [9] from what we used in the experiment in Sect. 2.1. First, we changed the level of the original noise between −86 and −36 dBov. Next, we constructed noisy speech by adding the noise to the original speech, and we distorted the noisy speech using a noise-reduction algorithm. In doing this, we changed the noise-reduction level. Finally, we normalized the signal level of the noise-reduced speech to −26 dBov.

We performed two verification tests, as follows.

Verification test 1: For several types of background noise

The noise-reduction algorithm is the same as that in the experiment in Sect. 2.1. The types of background noise are Hoth noise, office noise, and two other types of noise.

Verification test 2: For different noise-reduction algorithm and several types of background noise

The noise-reduction algorithm is different from that in the experiment in Sect. 2.1. The types of background noise are Hoth noise, office noise, and two other types of noise, which are different from those in test 1.

These verification tests are described in detail below.

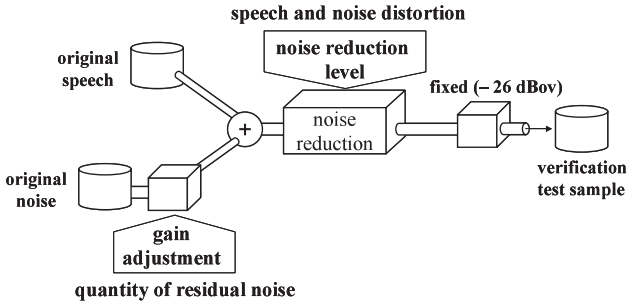

Fig. 6 Construction of verification test samples.

3.1 Verification Test 1

In this test, we verify that the proposed method can assess a perceived quality of noise-reduced speech accurately on a single scale regardless of the type of noise. We used Hoth noise, office noise, and noise recorded in a restaurant and an airport in [10] as the background noise, which we call "restaurant noise," and "airport noise," respectively. We selected the restaurant noise and airport noise so that each noise has different characteristics of impulsive sounds to investigate whether the estimation by the proposed method depends on the characteristics. We show the sound spectrograms of restaurant noise and airport noise in Fig. 7. Restaurant noise contains many impulsive sounds and airport noise contains few impulsive sounds.

The noise-reduction algorithm is the SS algorithm used in the experiment in Sect. 2.1. The number of verification test samples is 256, that is, 4 (speakers) × 4 (noise types) × 4 (noise-reduction levels) × 4 (levels of noise). Other conditions of this test were the same as those in Table 1.

The relationship between subjective MOS and its objective estimation is shown in Fig. 8. In Fig. 8, each subjective MOS is the average of 128 subjective votes, that is, 32 (subjects) × 4 (speakers). Good correlation (0.96) was achieved between the perceived speech quality and the estimation on the same scale for various background noise types. Hence, we concluded that the proposed method works for other types of noise than those used in optimizing the equations.

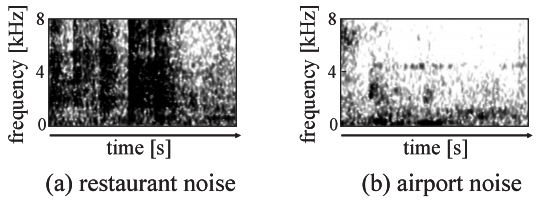

Fig. 7 Sound spectrograms of (a) restaurant noise and (b) airport noise.

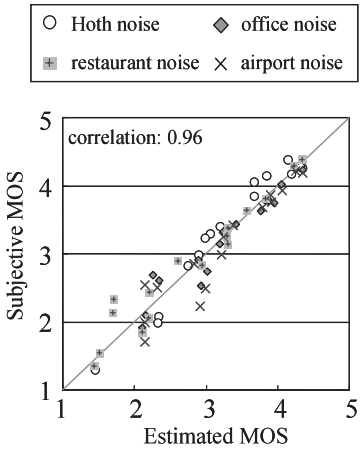

Fig. 8 Results of verification test 1.

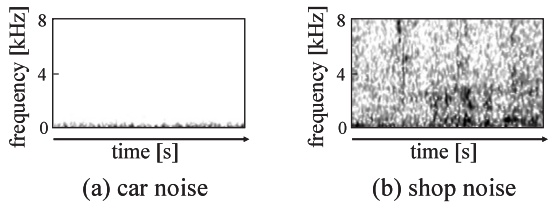

Fig. 9 Sound spectrograms of (a) car noise and (b) shop noise.

3.2 Verification Test 2

In this test, we verify that the proposed method can assess a perceived quality of noise-reduced speech accurately on a single scale regardless of the noise-reduction algorithm.

The noise-reduction algorithm used in this test is different from the SS algorithm. An outline of the noise-reduction algorithm is as follows. First, nonstationary noise is detected based on the amplitude of the nonstationary noise. Next, the nonstationary noise is reduced by the subband limiter at high frequencies, and the harmonic structure that is a characteristic of speech is enhanced at low frequencies [14]. That is, the algorithm enhances the speech signal and reduces other signals. Therefore, the noise-reduction algorithm, which is denoted as SE hereafter, effectively works for noise that contains impulsive sounds.

In this test, we used Hoth noise, office noise, and noise recorded in a car and a shop in [10] as the background noise, which we call "car noise," and "shop noise," respectively. We selected those two noise types so that each noise has a different frequency characteristic to investigate whether the estimation of the proposed method depends on the characteristic. We show the sound spectrograms of car noise and shop noise in Fig. 9. Car noise is dominated by low-frequency energy. Shop noise has low- and high-frequency energy.

In this test, we used both the SE algorithm and SS algorithm. The number of verification test samples using the SE algorithm is 256, that is, 4 (speakers) × 4 (noise types) × 4 (levels of noise) × 4 (noise-reduction levels). The number of verification test samples using the SS algorithm is 64, that is, 4 (speakers) × 2 (noise types, which are Hoth noise and office noise) × 4 (gains of noise) × 2 (noise-reduction levels). Other conditions were the same as those in verification test 1.

The relationship between subjective MOS and its objective estimation is shown in Figs. 10 and 11. In Figs. 10 and 11, each subjective MOS is the average of 128 subjective data, that is, 32 (subjects) × 4 (speakers).

In Fig. 10, for various noise types on the same scale, a good correlation (0.96) was obtained between the perceived quality and its estimation. That result confirmed that the finding in Sect. 3.1 is correct, that is, the proposed method does not depend on the frequency characteristics of background noise.

In Fig. 11, the correlation between the perceived quality and its estimation was high (0.96) for two algorithms

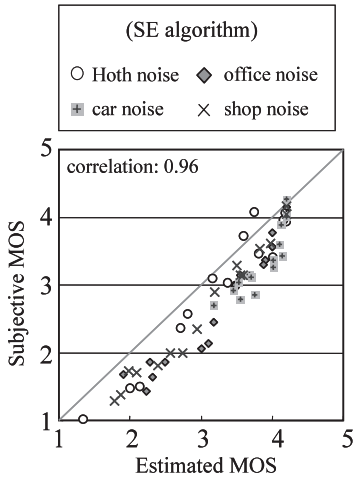

Fig. 10 Results for each background noise of verification test 2.

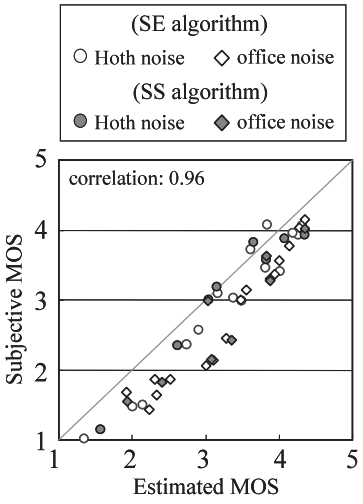

Fig. 11 Results for each noise-reduction algorithm of verification test 2.

on the same scale. That is, the relationship between the perceived quality and its estimation for two noise-reduction algorithms is almost the same. Therefore, the relationship between the perceived quality and its estimation does not depend on the noise-reduction algorithm.

In this verification test, we used some samples used in verification test 1. The result indicates that subjects in this verification test assign lower MOS values. I guess that was caused by a difference in the quality range of samples. Therefore, the estimated MOS tends to be larger than the subjective MOS.

We concluded that the three quality factors are sufficient to explain the perceived quality of noise-reduced speech and the presented method works for generic noise-reduction algorithms and background noise, enabling the comparison of different noise-reduction systems under various noise conditions. I guess that the reason for the absolute error between these two MOS is the difference in the quality range of speech samples. That is, the quality of the verification test samples was higher on average than that of the

speech samples we used in the experiment in Sect. 2.1. This resulted in the systematic difference between subjective and estimated MOS.

4. Conclusion

To select an appropriate noise-reduction system and optimize parameters effectively, we presented an objective quality evaluation method that estimates subjective speech quality of noise-reduced speech in wideband speech communication.

First, we selected quality factors that influence perceived speech quality. From the results of the subjective listening test, we found that there are statistically significant effects of the amount of residual noise, speech distortion, and noise distortion on perceived quality of noise-reduced speech.

Next, we proposed a method that estimates the perceived quality of noise-reduced speech from these factors. The method is based on characteristics of the human hearing system and the relationship between quality degradation and perceived quality. Therefore, the method enables accurate estimation of perceived quality.

Finally, we executed other subjective listening tests to investigate whether our method can measure perceived quality accurately. Then, we used various background noise and noise-reduction algorithms to investigate their properties. Results of the tests suggested that the correlation between the perceived quality and its estimation is high for various noise-reduction algorithms and types of background noise on the same scale. Hence, our method can assess and compare various types of noise-reduced speech, and it enables appropriate and effective selection of noise-reduction systems and parameter optimization of those systems.

Our investigation validated the proposed method in simulation environments. Future work will develop quality evaluation methodology not utilizing original speech and original noise and incorporate this method for appropriate quality planning and management of hands-free speech communication.

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] T. Yamada, M. Kumakura, and N. Kitawaki, "Sucjective and objective quality assessment of noise reduced speech signals," Processing IEEE-EURASIP International Workshop on Nonlinear Signal and Image Processing, NSIP2005, pp.328–331, May 2005.
[4] T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Objective perceptual quality measures for the evaluation of noise reduction schemes," IWAENC2005, pp.169–172, Sept. 2005.
[5] ITU-T Recommendation G.722, "7 kHz audio-coding within 64 kbit/s," Jan. 1998.
[6] ITU-T Recommendation G.722.2, "Wideband coding of speech at around 16 kbit/s using adaptive multi-rate wideband (AMR-WB)," July 2003.
[7] ITU-T Recommendation P.862.2, "Wideband extension to recommendation P.862 for the assessment of wideband telephone networks and speech codecs," Nov. 2005.
[8] V. Turbin and N. Faucheur, "A perceptual objective measure for noise reduction systems," Proc. MESAQIN2005, pp.81–84, June 2005.
[9] NTT-AT Corp, "Multi-lingual speech database for telephonometry," 1994.
[10] NTT-AT Corp, "Ambient noise database for telephonometry," 1996.
[11] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust. Speech Signal Process., vol.27, no.2, pp.113–120, April 1979.
[12] D. Hoyt and H. Wechlser, "RBD models for detection of human speech in structured noise," Proc. 1994 IEEE International Conference on Neural Networks, pp.4493–4496, 1994.
[13] H. Puder and O. Soffke, "An approach to an optimized voice-activity detector for noisy speech signals," EUSIPCO'2002, vol.I, pp.243–246, 2002.
[14] K. Noguchi, S. Sakauchi, Y. Haneda, and A. Kataoka, "Single-channel non-stationary noise reducetion based on subband limiter and harmonic enhancement," IWAENC2005, pp.41–44, 2005.

Noritsugu Egi received B.E. and M.E. degrees in Electrical Communication Engineering from Tohoku University in 2003 and 2005. He joined NTT Service Integration Laboratories of the NTT R&D Center, Tokyo, Japan, in 2005. Currently, he is researching speech and audio quality assessment.

Hitoshi Aoki received B.E. and M.E. degrees in Electrical and Electronic Engineering from Utsunomiya University in 1992 and 1994. He joined NTT Laboratories in 1994 and has been engaged in the quality assessment of speech and video telecommunications. Currently, he is primarily working on the quality assessment of speech over IP networks. He is a member of the Acoustical Society of Japan.

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 Systems and Information 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 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.