Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons

Sato, Hayato

Morimoto, Masayuki

Wada, Megumi

(Citation)
Journal of the Acoustical Society of America,131(6):4596-4605

(Issue Date)
2012-06

(Resource Type)
journal article

(Version)
Version of Record

Relationship between listening difficulty rating and objective measures in reverberant and noisy sound fields for young adults and elderly persons

Hayato Sato,[a)] Masayuki Morimoto, and Megumi Wada
Environmental Acoustics Laboratory, Graduate School of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501, Japan

(Received 2 September 2011; revised 24 April 2012; accepted 25 April 2012)

Listening difficulty ratings [Morimoto *et al.*, J. Acoust. Soc. Am. **116,** 1607–1613 (2004)] were obtained for 20 young adult listeners and 34 elderly listeners in reverberant and noisy sound fields simulated in an anechoic room. The listening difficulty ratings were compared with acoustical objective measures. The results and analyses showed the following: (i) The correlation between listening difficulty ratings and the revised speech transmission index (STI_r), and that for the useful-detrimental ratio (U_{50}) were high, regardless of the age of the listeners. (ii) STI_r and U_{50} need to be increased by 0.12 and 4.2 dB, respectively, to equalize the listening difficulty ratings for the elderly listeners with those for the young listeners. (iii) The estimation accuracies for STI_r and U_{50} can be improved by calculating them with the L_{eq} of background noise linearly increased by 4 to 10 dB, which depends on the age of the listeners and the objective measures. However, the improvement was not statistically significant for the elderly listeners.
[http://dx.doi.org/10.1121/1.4714790]

PACS number(s): 43.55.Hy, 43.71.Lz [NX] Pages: 4596–4605

I. INTRODUCTION

The evaluation of speech transmission performance is essential for the acoustical design of public spaces such as classrooms, meeting rooms, auditoriums, railway stations, and airports. Thus, the ideal goal of speech transmission is that listeners can understand speech information accurately and effortlessly. This indicates that the performance of the speech transmission system, which includes everything between a speaker and a listener, should be determined by subjective evaluation prior to objective evaluation.

It is thought that reproducibility and practicality are important for the subjective evaluation of speech transmission performance. Considering both reproducibility and practicality, Morimoto *et al.*[1] proposed listening difficulty as a subjective measure for evaluating speech transmission performance. Note that listening difficulty is a proposal of entire procedure in listening tests, which includes test materials, method of stimulus presentation, categories for listener's response, and method of data reduction.

To improve reproducibility, control of test materials is required. Considering practicality, it is desirable that speech materials frequently used in our daily life, which are not nonsense syllables but words or sentences, be used as test materials. In this case, the predictability of words or sentences should also be controlled to decrease the variety of results due to the cognitive process. Familiarity-controlled word lists 2003 (FW03)[2] are the only test materials in Japan that have been successfully used to control predictability. Amano *et al.*[2] reported that word intelligibility score increases with increasing word familiarity. In listening difficulty tests, the lists with the highest word familiarity from FW03 are used because they consist of words that are often used in our daily life. Furthermore, reference signals that are expected to have the lowest and highest ratings are used in listening difficulty tests to improve reproducibility. Sato *et al.*[3] reported that the reproducibility of listening difficulty ratings improves when a reverberation-free and quiet sound field and an extremely reverberant and noisy sound field are used as reference signals.

Listening difficulty ratings are the percentage of responses that indicate some level of difficulty. This definition indicates that speech transmission performance is optimized when no listener experiences listening difficulty. Morimoto *et al.*[1] suggested that the definition applies to reality and matches our normal feelings because people more frequently experience "difficulty" rather than positive feelings such as "easiness" in listening to speeches in their daily life. Sato *et al.*[4] conducted listening difficulty tests using the speech material from the Fairbanks rhyme test modified by Latham,[5] and their result clearly showed the practical advantage of using listening difficulty ratings against using word intelligibility scores. The advantage is that listening difficulty ratings have a range of 90% for conditions with speech-to-noise ratios between -5 and $+15$ dB, which are commonly found in daily life, while word intelligibility scores could not distinguish between conditions because the scores are constantly close to 100%. Lee and Jeon[6] also demonstrated that the advantage of listening difficulty ratings was limited in conditions with lower speech-to-noise ratios between -10 and $+5$ dB.

In public spaces, reverberation sound and background noise are main factors that spoil speech transmission performance. To convert listening difficulty ratings into objective measures that can be used in the acoustical design of public spaces, the relationships between listening difficulty ratings

a)Author to whom correspondence should be addressed. Electronic mail: hayato@kobe-u.ac.jp

 0001-4966/2012/131(6)/4596/10/$30.00

and objective measures from those detrimental sounds have to be clarified. Sato *et al.*[7] reported that speech transmission index (STI),[8] C_{50},[9] and D_{50}[10] are useful objective measures for estimating listening difficulty ratings in reverberant and quiet sound fields. Meanwhile, Sato *et al.*[4] and Kobayashi *et al.*[11] used both reverberation sound and background noise as detrimental sounds in their listening tests. However, in their listening tests, only one of the two detrimental sounds was used as a test parameter, and the other was fixed. Therefore, there is as yet no study using both of them as test parameters.

The effect of aging on listening difficulty is also an important topic for the acoustical design of public spaces. Sato *et al.*[12] reported that listening difficulty ratings for elderly listeners in reverberant and quiet sound fields increase with increasing hearing level. Prodi *et al.*[13] used listening difficulty to evaluate classroom acoustics for pupils and reported that listening difficulty for adults was higher than that for pupils. However, the studies did not consider the reproducibility of the listening difficulty tests, or use both reverberation sound and background noise as test parameters.

The purpose of the present study is to clarify the relationship among listening difficulty, objective measures, and age of listeners from the listening tests using both reverberation sound and background noise as test parameters. Listening tests were performed to obtain listening difficulty ratings for young-adult and elderly listeners in simulated sound fields. Then, the revised speech transmission index (STI_r) based on IEC60268-16 ed. 3.0[14] and the useful-detrimental ratio (U_{50}) proposed by Bradley,[15] which can evaluate the effects of both reverberation sound and background noise on speech transmission performance, were compared with the ratings.

II. METHODS

A. Participants

Two groups of participants, namely, the young and the elderly, took part in the listening test. The young group consisted of 20 young adults (9 males and 11 females, between 19 and 24 years old). They were university students. The elderly group consisted of 34 elderly persons (21 males and 13 females, between 66 and 79 years old). They were employed for the present study without any screening. There were no elderly participants wearing hearing aids in their daily life.

The hearing levels for both ears of each participant were measured in 5 dB steps using an audiometer in a sound-treated room. Figure 1 shows the mean hearing levels and standard deviations for each group and frequency. The open and closed circles indicate the results for the young and elderly groups, respectively. The mean hearing levels for the young group were less than 15 dB at all frequencies and were within normal hearing. The mean hearing levels for the elderly group showed a rapid increase above 2 kHz, which is a typical tendency of hearing loss due to aging.

B. Stimuli

One hundred Japanese words were used as test words. The test words were the same as those used in the previous study.[7] The test words were selected from familiarity-controlled word lists 2003 (FW03)[2] for maximum familiarity to both the young and elderly people. The length of the test words was constant at four *moras*.[16] The test words were pronounced by female narrators.

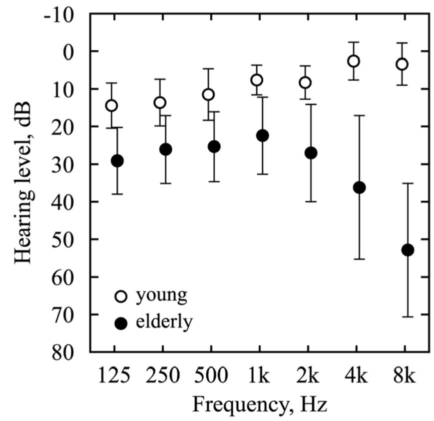

FIG. 1. Hearing levels for each group of participants. The open and closed circles represent the mean hearing levels for the young and elderly groups, respectively. The error bars represent the standard deviations between participants.

The test words were convolved with 13 impulse responses to generate reverberant speech. The impulse responses were selected from 96 impulse responses used in the previous study.[7] The impulse responses were measured in various types of rooms in Japan.[17] STI_r obtained from the impulse responses widely varied from 0.30 to 0.92. The intervals between a STI_r and the nearest one were approximately 0.05. The intervals are close to the just noticeable difference of STI ($=0.03$) reported by Bradley *et al.*[18] Additionally, the test words with no reverberation sound were also used. A total of 1400 reverberant speeches (100 test words $\times$ 14 reverberation conditions) were prepared. The presentation levels of each reverberant speech were aligned to 65 ± 0.3 dB in terms of $L_{AS\max}$ at the position corresponding to the center of participant's head.

A steady-state random noise with a Hoth spectrum[19] was used as the additional noise. The presentation level of the additional noise was set from 35 to 60 dB at 5 dB intervals, in terms of the median level (*A*-weighted, slow response), at the same position as that for the speech stimuli. No additional noise condition was used in the listening test. The ambient noise level (L_{Aeq}) at the listening point was 15 dB.

A reverberant speech and an additional noise were mixed down to a mono track to generate stimuli. The duration of the additional noise was 5 s with a 50 ms rise/fall time. The reverberant speech started 100 ms after the onset of the additional noise. The duration of the reverberant speech varied from 0.8 to 4.0 s depending on the impulse response. A total of 9800 stimuli (1400 reverberant speech $\times$ 7 noise conditions) were prepared.

Figure 2 shows the frequency distribution of STI_r calculated for 98 sound fields, which are combinations of 14 reverberation conditions and seven noise conditions. It can be said that the values of STI_r cover the existent range of sound fields in daily life. Detailed calculation procedures for objective measures are described in Sec. II D.

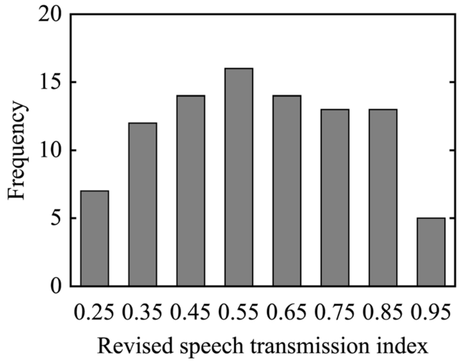

FIG. 2. Frequency distribution of the revised speech transmission index (Ref. 14) for test sound fields. Numerical labels for the abscissa axis are the median of each bin.

C. Procedure

The listening tests were performed in an anechoic room. The stimuli were presented from a loudspeaker (Fujitsu Ten, TD512 Kobe, Japan) at a distance of 1.5 m in front of each participant. The frequency response of the loudspeaker was flat within ± 5 dB from 100 to 10 kHz at the listening position. The stimuli were divided into 100 different sets consisting of 98 stimuli. Furthermore, two test words without both reverberation and noise were added to each set as reference signals to improve reproducibility. Each set consisted of 100 different test words, and simultaneously included two reference signals and 98 different sound fields.

The participants listened to five different sets of stimuli. The listening test was divided into 10 sessions to listen to half of the set of stimuli with a break between sessions. The presentation order of stimuli was random in each session, except that the first and the second stimuli were the reference signal and the stimuli with the lowest STI_r in each session, respectively. The interval between the stimuli was 5 s.

Each participant was asked to write down each test word as they heard it in Japanese and simultaneously rate the listening difficulty into one of four categories as shown in Table I, according to Morimoto *et al.*[1] Before the listening test, each participant listened to 20 stimuli as an exercise. The exercises were repeated until each participant could complete the task within the interval.

Word intelligibility and listening difficulty were obtained from the participants' responses. The word intelligibility score is the percentage of the test words written down correctly. The listening difficulty rating is the percentage of the sum of listening difficulty ratings evaluated from "2" to "4" in Table I. The sample numbers for each sound field were 100 for the young group (20 participants $\times$ 5 sets of stimuli) and 170 for the elderly group (34 participants $\times$ 5 sets of stimuli).

TABLE I. Categories of listening difficulty (Ref. 1).

1	Not	Difficult
2	A little	Difficult
3	Fairly	Difficult
4	Extremely	Difficult

D. Objective measures

STI_r and U_{50}[15] were calculated for 98 sound fields for comparison with listening difficulty ratings. STI_r was calculated according to IEC60268-16 ed. 3.0.[14] The weighting factors for female speech were used. The L_{eq}'s in each 1/1 octave band for the speech stimuli and additional noise are necessary to calculate STI_r. The L_{eq} of the speech stimuli for each reverberation condition was measured as follows—considering utility for estimating listening difficulty ratings from the results of the present study and *in situ* measurements in other spaces. First, the digitized sound files of the test words, which were calibrated to equalize the presentation level described in Sec. II B, were merged into a long sound file without intervals. Second, the long sound file was convolved with the impulse response for each sound field. Finally, the L_{eq} in each 1/1 octave band was measured using a sound level meter at the listening point, while the output was repeatedly radiated from the loudspeaker three times. The L_{eq} measured using this method would be overestimated relative to the speech stimuli that the participants listened to because the reverberation sound of each test word overlaps to the next test word. To quantitatively clarify the overestimation of the L_{eq}, the speech stimuli with STI_r of 0.3, that is the lowest STI_r used in the present study, were trimmed to remove terminal reverberation, then the trimmed speech stimuli were merged into a long sound file without intervals. The L_{eq}'s in each 1/1 octave band measured using in the present study were larger than the L_{eq}'s for the trimmed speech stimuli. However, the differences between them were less than 1.3 dB, and the difference in STI_r was only 0.007 when the additional noise level was 60 dB. Therefore, the overestimation can be ignored. The L_{eq} of the additional noise was measured at the listening point, while repeatedly radiating the additional noise without a rise/fall time.

U_{50} was calculated for 1/1 octave bands from 250 Hz to 8 kHz using

$$U_{50,i} = 10 \log_{10}\left[\frac{c_{50,i}}{1+\left(1+c_{50,i}\right)10^{-\mathrm{SNR}_i/10}}\right], \tag{1}$$

where $c_{50,i}$ is the linear and not the logarithmic early-to-late sound ratio for band i and SNR_i is the speech-to-noise ratio for band i. $c_{50,i}$ was obtained from the impulse response passed through the 1/1 octave band-pass filter for band i. SNR_i was obtained from the L_{eq}'s in each octave band of the speech stimuli and additional noise described above.

III. RESULTS AND DISCUSSION

A. Word intelligibility scores and listening difficulty ratings

Figure 3 represents the relationship between word intelligibility scores and listening difficulty ratings. The open and closed circles represent the results for the young and elderly groups, respectively.

For the young group, the listening difficulty rating varied from 0 to 60% in the sound fields where the word intelligibility score was close to 100%. Furthermore, the listening

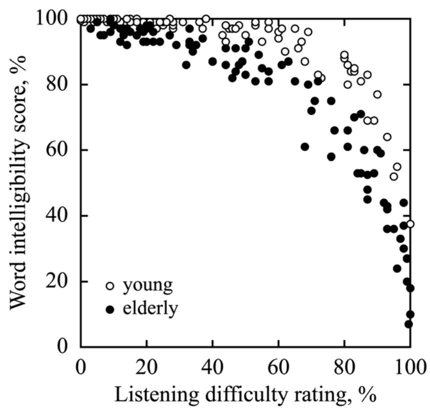

FIG. 3. Relationship between word intelligibility score and listening difficulty rating. The open and closed circles represent the results for the young and elderly groups, respectively.

difficulty rating widely varied from 0 to 100% whereas the word intelligibility score varied from 40% to 100%. This result is clearly consistent with that of previous studies, that is, the listening difficulty rating can be used to evaluate the differences between sound fields where the word intelligibility score cannot be used.

For the elderly group, in contrast, the word intelligibility score was not constant at 100% and varied from 80% to 100%, in the range of the listening difficulty rating for the elderly group from 0 to 60%. However, similarly to that for the young group, the listening difficulty rating widely ranged from 0 to 100%. This means that the listening difficulty rating for the elderly group can also be used to discriminate among a wide range of sound fields.

The relationships between the word intelligibility score and the listening difficulty rating for the young and elderly groups were quantitatively different from each other. More specifically, the listening difficulty ratings for the young group were higher than those for the elderly group when they were compared under the same intelligibility conditions. In other words, young people tend to evaluate listening difficulty more sensitively than elderly people. This result might be related to the difference between young and elderly people in the framework for evaluating listening difficulty. It is possible that, for example, elderly people selected "not difficult" once they thought they caught the test words regardless of the difficulty they experienced, whereas young people selected "not difficult" considering factors other than intelligibility like annoyance of reverberation sound or that of background noise. A similar tendency is also found in the relationship between listening difficulty and intelligibility for pupils reported by Prodi *et al.*,[13] and that for non-native listeners reported by Sato *et al.*[4] This topic is also discussed in Sec. IV A.

It is concluded that the advantage of the listening difficulty rating in evaluating speech transmission performance still holds in reverberant and noisy sound fields, regardless of the age of the listeners. However, note that equal listening difficulty ratings do not always mean equal word intelligibility scores, if speech transmission performances for young and elderly people are compared with each other. The listening difficulty ratings should not be used alone, and the relationship between the rating and the word intelligibility score, as shown in Fig. 3, should be considered at the same time.

B. Estimation of listening difficulty from objective measures

The listening difficulty ratings for both groups were converted into z-scores to analyze the relationship between the ratings and objective measures using a linear regression analysis, as in our previous study.[7] Ratings that were less than 5% and more than 95% were omitted to prevent their excessive effect on the results of linear regression analyses. Furthermore, ratings for the sound fields with no reverberation were also omitted because no c_{50} could be obtained for such sound fields. As a result, for both the young and elderly groups, ratings for 80 of the 98 sound fields were used for the analyses.

1. *Estimation from STI$_r$*

Figure 4 represents the z-score of the listening difficulty rating as a function of STI$_r$ for each group. Panels (a) and (b) represent the results for the young and elderly groups, respectively. Different symbols represent different additional noise levels (L_N). The solid and dashed lines in each panel represent a linear regression line and the 95% prediction

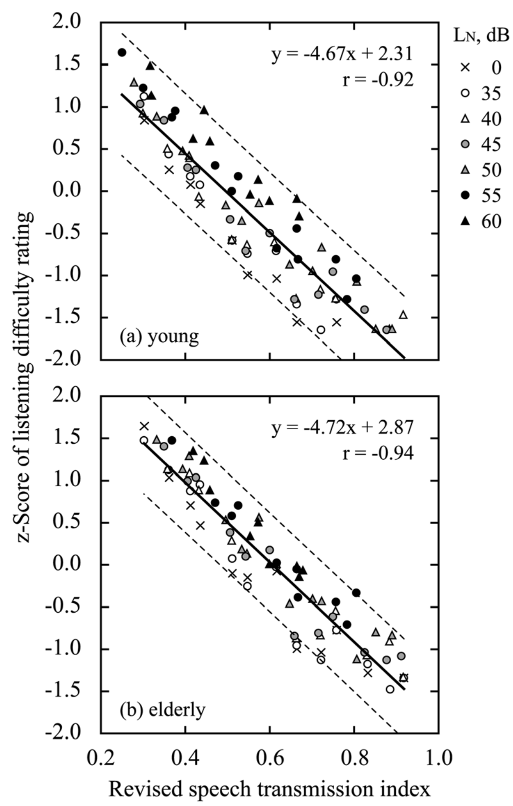

FIG. 4. z-score of listening difficulty rating as function of the revised speech transmission index (Ref. 14). Panels (a) and (b) represent the results for the young and elderly groups, respectively. Different symbols represent different additional noise levels (L_N). The solid and dashed lines in each panel represent a linear regression line and the 95% prediction intervals, respectively.

intervals, respectively. The z-scores for both groups strongly correlated with STI_r ($r = -0.92$ for the young group, $r = -0.94$ for the elderly group). The correlation for the elderly group was stronger than that for the young group.

A multiple regression analysis was performed to statistically confirm that the regression equations for the young and elderly group were different from each other, and were parallel to each other. Equation (2) is a multiple regression model tested below:

$$\mathrm{ZLD} = a_0 + a_1 \mathrm{STI}_r + a_2 \mathrm{Group} + a_3(\mathrm{STI}_r \times \mathrm{Group}), \tag{2}$$

where ZLD is the z-score of listening difficulty ratings and "Group" is a dummy variable; the values of 0 and 1 indicate the young and elderly groups, respectively. If the regression coefficient of a_2 is significant, the difference between the groups would be statistically significant. In a similar way, a statistical significance of a_3 means that the interaction between STI_r and Group is significant and the regression lines are not parallel to each other. The result of a multiple regression analysis showed that the regression coefficients of a_1 (-4.67) and a_2 (0.554) were statistically significant ($p < 0.01$). Meanwhile, the regression coefficient of a_3 (-0.058) was not statistically significant ($p = 0.844$). Therefore, it is statistically confirmed that the regression equations for the young and elderly group are different from each other, and are parallel to each other. The regression coefficients for both the young and elderly groups were close to -4.7, while the y-intercept for the young group was 0.56 less than that for the elderly group. This result indicates that STI_r needs to be increased by 0.12 to equalize the listening difficulty ratings for the elderly group with those for the young group.

The widths of the 95% prediction interval were approximately ± 0.7 and ± 0.6 in the z-scores for the young and elderly groups, respectively. The widths correspond to $\pm 26\%$ and $\pm 22\%$ differences in the listening difficulty rating, where the estimated rating is 50%. The widths seem too large for accurate estimation. However, there is a possibility that the estimation errors can be reduced by applying a correction to L_N. Because the errors depend on L_N, in particular, the errors were positive for higher noise levels and negative for lower noise levels, regardless of the group. The reduction in the error is discussed in Sec. IV A.

2. Estimation from U_{50}

As described in Sec. II D, the values of U_{50} in the 1/1 octave bands from 250 Hz to 8 kHz were obtained. Linear regression analyses between the z-scores of the listening difficulty ratings and various types of U_{50} were performed. The values of U_{50} for each octave band, the arithmetical means of several bands, and the weighting sum of the bands from 250 Hz to 8 kHz with the weighting factor of the speech intelligibility index (SII)[20] were used in the analyses.

Table II shows the correlation coefficients between the z-score and the various types of U_{50}. The correlations for 250 and 500 Hz were rather lower than those for the other bands or types, while the correlation coefficients for bands other than 250 and 500 Hz were approximately -0.9, regardless of the group of participants. That is, U_{50} also strongly correlated with the listening difficulty rating, and the degree of correlation was not clearly affected by the difference in frequency band or calculation method. Therefore, in this paper, the arithmetical mean of the bands from 500 Hz to 4 kHz [$U_{50(500\text{-}4k)}$] are shown as an example, considering that the grades of hearing impairment suggested by the World Health Organization[21,22] are based on the mean of hearing levels from 500 Hz to 4 kHz.

Figure 5 represents the z-score of the listening difficulty rating as a function of the $U_{50(500\text{-}4k)}$ for each group. The results were almost the same as those for STI_r. Specifically, the results showed high correlations between the z-score and $U_{50(500\text{-}4k)}$ for both groups; the correlation for the elderly group was higher than that for the young group; the estimation error depended on L_N. However, the correlation between the z-scores and $U_{50(500\text{-}4k)}$ was lower, and the 95% prediction intervals were wider than those for STI_r.

A multiple regression analysis using the same model shown in Eq. (2), except STI_r is replaced with $U_{50(500\text{-}4k)}$, was performed. The result showed that the regression coefficients of a_1 (-0.123) and a_2 (0.509) were statistically significant ($p < 0.01$). Meanwhile, the regression coefficient of a_3 (0.003) was not statistically significant ($p = 0.733$). Therefore, it is statistically confirmed that the regression equations for the young and elderly group are different from each other, and are parallel to each other. The regression coefficients were approximately 0.12 for both the young and elderly groups, and the difference between the y-intercepts for both groups was approximately 0.51. This means that an increase in $U_{50(500\text{-}4k)}$ by 4.2 dB is required to equalize the listening difficulty rating for the elderly group with that for the young group.

IV. FURTHER CONSIDERATION

A. Attempt to correct the difference between the effects of noise and reverberation sound

Figures 4 and 5 clearly show that the estimation error depends on L_N. For a higher L_N, the listening difficulty ratings were higher than those without additional noise. In other

TABLE II. Correlation coefficients between the z-score of listening difficulty rating and various types of the useful-detrimental ratios [see Eq. (1)].

	Center frequency (Hz)						Arithmetical mean						Weighting sum
Group of participants	250	500	1k	2k	4k	8k	250–8k	250–4k	500–8k	250–2k	500–4k	1k–8k	SII
Young	−0.77	−0.84	−0.87	−0.89	−0.87	−0.88	−0.87	−0.87	−0.89	−0.86	−0.88	−0.89	−0.88
Elderly	−0.83	−0.87	−0.89	−0.92	−0.90	−0.89	−0.91	−0.90	−0.91	−0.90	−0.91	−0.91	−0.91

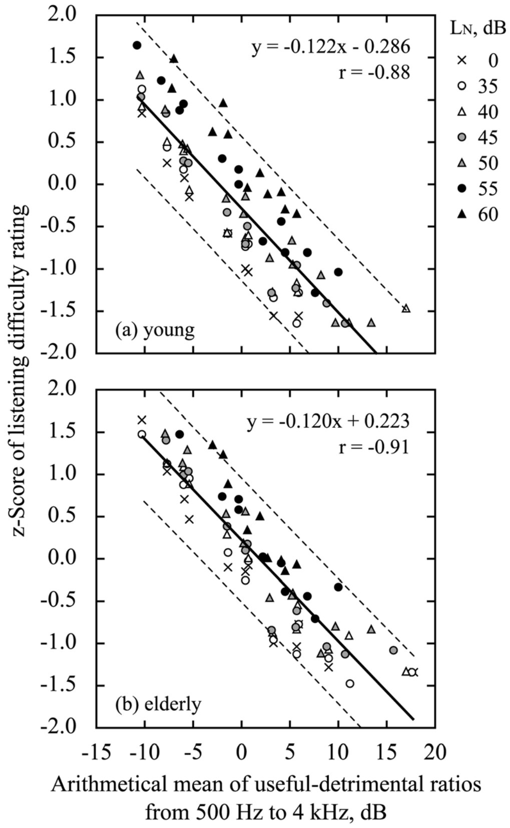

FIG. 5. Same as Fig. 4, but for the arithmetical mean of the useful-detrimental ratios [see Eq. (1)] in the 1/1 octave bands from 500 Hz to 4 kHz.

words, the error should decrease when plots for sound fields with a higher L_N are shifted toward the left direction. When L_{eq} of the additional noise for each octave band linearly increased by X dB are used for calculating the objective measures, each objective measure decreases with increasing X. Note that the decrease in the objective measure is larger for sound fields with a higher L_N than for sound fields without the additional noise because the energy of the additional noise is dominant relative to that of the reverberation sound for sound fields with a higher L_N. In the rest of the paper, the X described above is defined as the increase in noise level (I_N). If I_N was optimized, the plots should be closer to a linear regression line than to those in Figs. 4 and 5. To clarify the optimum values of I_N for STI_r and $U_{50(500\text{-}4k)}$, each objective measure was calculated while increasing I_N by 1 dB step, and the correlation coefficients between the z-score of the listening difficulty rating and the objective measures were also calculated.

Figure 6 represents the correlation coefficient between the z-score of listening difficulty rating and the objective measure as a function of I_N for each group of participants and objective measure. Every relationship represented a quadratic function and had a local minimum.

For the young group, the coefficients were minimized when the I_N values were 7 and 10 dB for STI_r and $U_{50(500\text{-}4k)}$, respectively. Figure 7 represents the z-scores of the listening difficulty ratings for the young group as a function of $\mathrm{STI}_r(I_N = 7)$, that is, the STI_r obtained using the I_N of 7 dB.

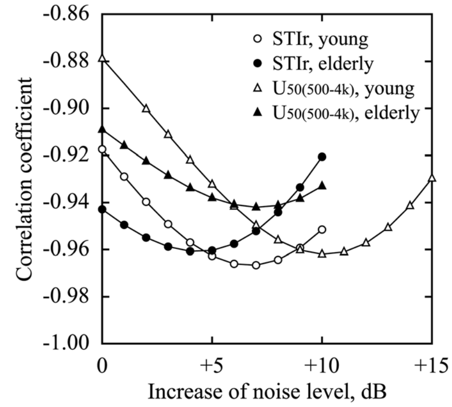

FIG. 6. Correlation coefficients between z-score of listening difficulty rating and objective measure as function of the increase in noise level (I_N). The objective measure was calculated using the L_{eq} of the additional noise linearly increased by I_N dB. The circles and triangles represent the results for the young and elderly groups, respectively. The open and filled symbols represent the results for the revised speech transmission index (STI_r) (Ref. 14) and the arithmetical mean of the useful-detrimental ratios [see Eq. (1)] in the 1/1 octave bands from 500 Hz to 4 kHz [$U_{50(500\text{-}4k)}$], respectively.

The correlation coefficient ($r = -0.97$) improved relative to that for STI_r. A χ^2-test ($p < 0.05$) showed that the correlation coefficient for $\mathrm{STI}_r(I_N = 7)$ was statistically different from that for STI_r. The width of the 95% prediction interval was $\pm 18\%$, which is $\pm 8\%$ narrower than that for STI_r. Figure 8 represents the z-score as a function of $U_{50(500\text{-}4k)}(I_N = 10)$, that is, the $U_{50(500\text{-}4k)}$ obtained using the I_N of 10 dB. The correlation coefficient ($r = -0.96$) also statistically improved ($p < 0.05$) relative to that for $U_{50(500\text{-}4k)}$.

For the elderly group, the correlation coefficients between the z-score of listening difficulty rating and the objective measure were minimized when the I_N values were 4 and 7 dB for STI_r and $U_{50(500\text{-}4k)}$, respectively. The optimum values of I_N were 3 dB lower than those for the young group. Figures 9 and 10 represent the z-score for the elderly group as a function of the objective measures calculated using the optimum I_N for each measure. The correlation coefficients improved relative to those for the objective

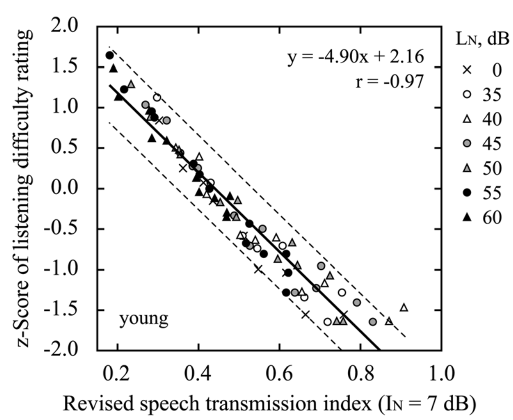

FIG. 7. z-score of listening difficulty rating for young group as function of the revised speech transmission index (Ref. 14) with the increase in noise level (I_N) of 7 dB (see Fig. 6). Different symbols represent different additional noise levels (L_N). The solid and dashed lines in each panel represent a linear regression line and the 95% prediction intervals, respectively.

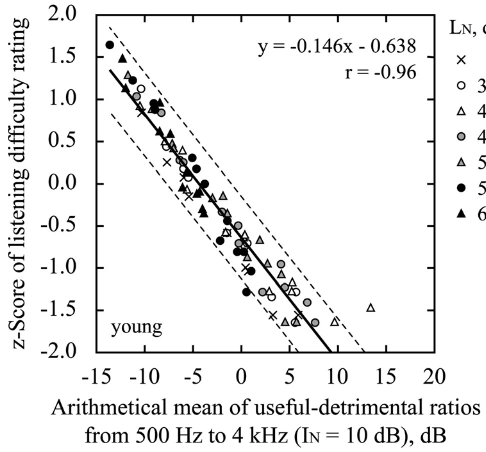

FIG. 8. Same as Fig. 7, but for the arithmetical mean of the useful-detrimental ratios [see Eq. (1)] in the 1/1 octave bands from 500 Hz to 4 kHz with the increase in noise level (I_N) of 10 dB (see Fig. 6).

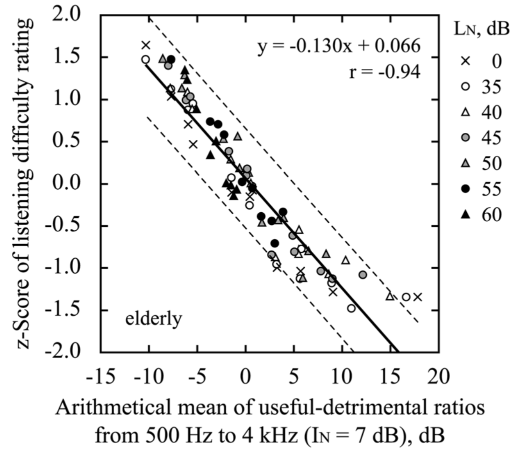

FIG. 10. Same as Fig. 8, but for elderly group. The increase in noise level (I_N) is 7 dB (see Fig. 6).

measures without applying I_N. However, there was no statistically significant difference between the correlation coefficients for the objective measures with I_N application and for those without I_N application [p=0.23 for STI$_r$, p = 0.15 for $U_{50(500\text{-}4k)}$].

Roughly speaking, STI$_r$ and U_{50} assume that the degrees of interruption from the reverberation sound and noise are equal if their energies are equal. However, the result showing that the optimum values of I_N in all cases were positive indicates that the effect of noise on the listening difficulty ratings was larger than that of reverberation sound, regardless of the age of the listeners. I_N is a useful variable for correcting the difference between the effects of reverberation sound and noise on the subjective measures of speech transmission performance. Note that the optimum I_N is a characteristic value for the listening difficulty rating obtained here and depends on many factors such as the subjective measure to be estimated, the presentation method of stimuli, the familiarity of speech materials, and the duration of speech materials. For example, the effects of reverberation sound on intelligibility are different between speeches with/without a carrier phrase.[23]

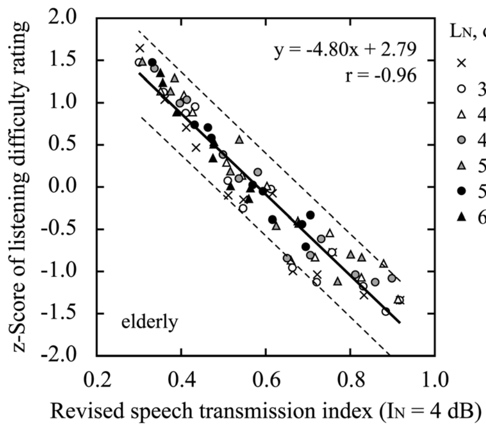

FIG. 9. Same as Fig. 7, but for the elderly group. The increase in noise level (I_N) is 4 dB (see Fig. 6).

Meanwhile, the age of the listeners also affected the optimum I_N. The optimum values of I_N for the elderly group were 3 dB lower than those for the young groups. Furthermore, the correlation coefficients for the elderly group did not statistically improve with I_N application. These results indicate that the effect of the reverberation sound relative to that of the additional noise on the listening difficulty rating is larger for the elderly group than for the young group.

It can be thought that it is not energetic masking but informational masking[24] that causes the difference between the effects of reverberation sound and noise on the subjective measures of speech transmission performance. Reverberation sound resembles speech signals more closely than random noise. Therefore, listeners would have a larger difficulty in the cognitive process to focus attention on speech signals and suppress maskers in reverberant sound fields than noisy sound fields, if the amounts of energetic masking of reverberation sound and random noise are equal. However, studies of the relationship between aging or hearing impairment and informational masking[25–27] suggested that the amount of informational masking does not increase with aging. In other words, the difference in optimum I_N is not related to informational masking.

As described in Sec. III A, an alternative possibility is that the young group is more sensitive to annoyance or factors other than the intelligibility in rating listening difficulty than the elderly group. Although annoyance due to noise depends on noise sources, listening conditions, listeners' tasks, and other factors, Gjestland and Oftedal[28] and Fields and Walker[29] show that annoyance increases with increasing noise level where the L_{Aeq}[28] or L_{eq}[29] of noise exceeds approximately 40 dB. This means that the listeners in the present study could feel annoyance due to noise in the sound fields with an L_N of more than 40 dB, at least for the young group. For the elderly group, an increase in listening difficulty rating related to the annoyance caused by noise disappeared because of the lack of attention to the annoyance of noise or a decrease in the loudness of noise caused by hearing loss. Accordingly, the optimum I_N of this group decreased relative to that of the young group.

B. Effect of hearing level on listening difficulty rating

Sato *et al.*[12] reported that listening difficulty ratings increased with increasing hearing level. To study how hearing level affects the results obtained in the present study, the grades of hearing impairment[21,22] for each elderly listener were obtained on the basis of the mean of hearing levels at 0.5, 1, 2, and 4 kHz for the better ear. The elderly listeners ($N = 34$) were classified into "normal" ($N = 19$) or "mild" ($N = 15$) grades. No listener was classified into grades above "mild." The mean hearing levels for the normal and mild groups were 17.0 and 33.0 dB, respectively.

In the following analyses, the effect of hearing level was studied using the listening difficulty rating for the young group as a reference. Figure 11 shows the relationship between the z-score of listening difficulty rating for the young group and those for the elderly group. The ordinates in panels (a) to (c) represent the z-scores of the listening difficulty ratings for all the elderly listeners, and the normal and mild groups, respectively. Ratings that were less than 5% and more than 95% were omitted from the analyses. Almost all the symbols were plotted above the dashed lines for each panel. This means that the z-scores for the elderly group were higher than those for the young group, even for the normal group. This result is consistent with that of Sato *et al.*[12]

Multiple regression analyses using a model shown in Eq. (3) were performed to statistically test the effect of hearing level on listening difficulty ratings for the elderly listeners.

$$\mathrm{ZLDE} = a_0 + a_1\mathrm{ZLDY} + a_2\mathrm{Group} + a_3(\mathrm{ZLDY} \times \mathrm{Group}), \tag{3}$$

where ZLDY and ZLDE are the z-scores of listening difficulty ratings for the young and elderly groups, respectively, and "Group" is a dummy variable. In the rest of paper, three groups of listeners, which are all elderly listeners, the normal group, and the mild group, are abbreviated as ALL, NORM, and MILD. The differences among the three groups were tested independently. Table III represents the summary of multiple regression analyses. The regression coefficients of a_1 and a_2 were statistically significant ($p < 0.01$) for all analyses. This means that the differences in listening difficulty ratings among ALL, NORM and MILD are statistically significant. The regression coefficient of a_2 indicates the difference in the z-score between them. For the difference between NORM and MILD, an increase in the z-score by 0.384 corresponds to an increase in the listening difficultly rating by around 15% when the listening difficulty rating for NORM is 50%. In a similar way, when the listening difficulty rating for ALL is 50%, the ratings for NORM and MILD are estimated to be 44.3% and 59.5%, respectively. The differences are less than 10%. This means that the results for all the elderly listeners are practically acceptable for estimating the listening difficulty ratings for elderly listeners with "normal" or "mild" grades of hearing impairment from the objective measures.

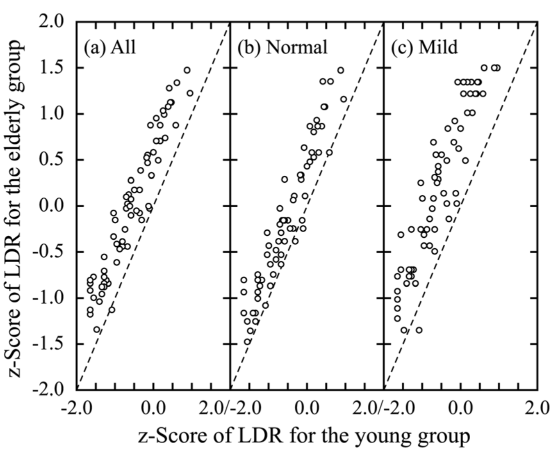

FIG. 11. Relationships between z-score of listening difficulty rating (LDR) for the elderly groups and that for the young group. Panels (a) to (c) show the results for all the elderly listeners ($N = 34$), elderly listeners with normal hearing level ($N = 19$), and elderly listeners with mild hearing loss ($N = 15$), respectively. The dashed line indicates that the z-scores for the young and elderly groups are equal.

C. Application to public spaces

Temporal, spectral, and spatial characteristics of reverberation sound and background noise in public spaces were more complex than those in the present study. In this subsection, the effects of the three physical characteristics on listening difficulty ratings are discussed to clarify applicability of the results of the present study in public spaces.

Temporal characteristics of the reverberation sound varied widely while the additional noise was limited to steady-state in the present study. Festen and Plomp[30] reported that the speech reception threshold for steady-state noise was higher by around 5 dB than that for noise with the envelope of speech. Lee and Jeon[6] reported that listening difficulty ratings showed a wider variance than intelligibility scores when background noise was impulsive. These studies indicate that listening difficulty ratings will be affected by temporal characteristics of background noise. However, there is no objective measure of speech transmission performance which can consider time-varying noises. The results of the present study are applicable only for sound fields with steady-state noise, primarily because both STI_r and U_{50} assume that background noise is steady-state.

TABLE III. Intercepts and regression coefficients of multiple regression equations based on a regression model shown in Eq. 3. "Group" is a dummy variable, and the values of 0 and 1 are respectively assigned to each two of three groups of elderly listeners: all the elderly listener (ALL), the elderly listeners with normal hearing (NORM), and the elderly listeners with mild hearing loss (MILD) as shown in the first column. The asterisk represents statistical significance ($p < 0.01$).

Group			Regression coefficient		
0	1	Intercept (a_0)	a_1	a_2	a_3
NORM	MILD	0.417*	1.04*	0.384*	0.067
ALL	NORM	0.560*	1.03*	−0.143*	0.009
ALL	MILD	0.560*	1.03*	0.241*	0.076

*$p < 0.01$.

Spectral characteristics of the reverberation sound also varied widely in the present study because the impulse responses measured in various sound fields were used to add reverberation sound to the test words. Meanwhile, only Hoth spectrum was used for the additional noise. Some studies[31–33] showed that long-term frequency spectra of background noise in public spaces is broad and decreases with increasing frequency. Therefore, Hoth spectrum can be a representative frequency spectrum of background noise in public spaces. If frequency spectrum of background noise were extremely different from Hoth spectrum, the relationship between listening difficulty ratings and objective measures would be different from that obtained here because of factors other than intelligibility, such as annoyance.

Spatial characteristics of the additional noise and the reverberation sound in the present study were also different from those in public spaces. For example, Soeta and Simokura[33] reported that interaural cross-correlation coefficients of environmental sound in railway stations were less than 0.35, while the interaural cross-correlation coefficients of the additional noise and the reverberation sound were close to 1. Sato *et al.*[34,35] reported that interaural cross-correlations of speech and noise made a significant difference in listening difficulty ratings only for a speech-to-noise ratio of -10 dB, and did not for higher speech-to-noise ratios such as $+10$ dB. A speech-to-noise ratio of -10 dB is approximately corresponding to STI_r of 0.17. Therefore, in sound fields that often appear in public spaces, interaural cross-correlations would not affect the relationship between listening difficulty ratings and objective measures.

V. CONCLUSIONS

The listening tests were performed to clarify the relationship among listening difficulty rating, objective measures, and age of listeners. The results and the analyses lead to conclusions as follows:

(1) Listening difficulty ratings can be used to evaluate speech transmission performance sensitively in reverberant and noisy sound fields regardless of the age of the listeners. However, note that equal listening difficulty ratings do not always mean equal word intelligibility scores if speech transmission performances for young and elderly listeners were compared with each other.
(2) The correlation between listening difficulty ratings and STI_r, and that for U_{50} were high, regardless of the age of the listeners. The correlation coefficients were within the range from -0.88 to -0.94.
(3) The difference in listening difficulty ratings between the young and elderly listeners was statistically significant. STI_r and U_{50} need to be increased by 0.12 and 4.2 dB, respectively, to equalize the listening difficulty ratings for the elderly listeners with those for the young listeners.

As a further consideration, the correction of the difference between the effects of noise and reverberation sound on listening difficulty ratings was attempted. The effect of hearing level on listening difficulty rating was also analyzed. A summary of this consideration is as follows:

(i) The estimation accuracies for STI_r and U_{50} can be improved by calculating them with the L_{eq} of background noise linearly increased by 4 to 10 dB, which depends on the age of the listeners and the objective measures. However, the improvement was not statistically significant for the elderly listeners.
(ii) The effect of hearing level on the estimation of the listening difficulty rating for the elderly listeners can be practically ignored when the grades of hearing impairment for the elderly listeners are normal or mild.

ACKNOWLEDGMENTS

This research project was partially supported by the research and development grant of the Japan Institute of Construction and Engineering, 03011, 2004, and by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (B), 16360292, 2004–2006. A part of this study was carried out as a project of the Sub Working Group on Research in Speech Transmission Quality at the Architectural Institute of Japan.

[1]M. Morimoto, Hi. Sato, and M. Kobayashi, "Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces," J. Acoust. Soc. Am. **116**, 1607–1613 (2004).
[2]S. Amano, S. Sakamoto, T. Kondo, and Y. Suzuki, "Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese," Speech Commun. **51**, 76–82 (2009).
[3]Ha. Sato, M. Morimoto, A. Hakamada, M. Kobayashi, and Hi. Sato, "Learning effect and reduction of context effect on listening difficulty for words," Mem. Grad. Sch. Sci. Technol., Kobe Univ. **22B**, 47–57 (2004) (in Japanese).
[4]Hi. Sato, J. Bradley, and M. Morimoto, "Using listening difficulty ratings of conditions for speech communication in rooms," J. Acoust. Soc. Am. **117**, 1157–1167 (2005).
[5]H. G. Latham, "The signal-to-noise ratio for speech intelligibility—An auditorium acoustics design index," Appl. Acoust. **12**, 253–320 (1979).
[6]P. J. Lee and J. Y. Jeon, "Evaluation of speech transmission in open public spaces affected by combined noises," J. Acoust. Soc. Am. **130**, 219–227 (2011).
[7]Ha. Sato, M. Morimoto, Hi. Sato, and M. Wada, "Relationship between listening difficulty and acoustical objective measures in reverberant sound fields," J. Acoust. Soc. Am. **123**, 2087–2093 (2008).
[8]T. Houtgast and H. J. M. Steeneken, "The modulation transfer function in room acoustics as a predictor of speech intelligibility," Acustica **28**, 66–73 (1973).
[9]J. S. Bradley, "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80**, 837–845 (1986).
[10]R. Thiele, "Richtungsverteilung und zeitfolge der Schallrückwürfe in Räumen (Directional distribution and time sequence of sound reflections in rooms)," Acustica **3**, 291–302 (1953).
[11]M. Kobayashi, M. Morimoto, Hi. Sato, and Ha. Sato, "Optimum speech level to minimize listening difficulty in public spaces," J. Acoust. Soc. Am. **121**, 251–256 (2007).
[12]Ha. Sato, Hi. Sato, and M. Morimoto, "Effects of aging on word intelligibility and listening difficulty in various reverberant fields," J. Acoust. Soc. Am. **121**, 2915–2922 (2007).
[13]N. Prodi, C. Visentin, and A. Farnetani, "Intelligibility, listening difficulty and listening efficiency in auralized classrooms," J. Acoust. Soc. Am. **128**, 172–181 (2010).
[14]IEC 60268-16 Ed. 3.0, *Sound System Equipment-Part16: Objective Rating of Speech Intelligibility by Speech Transmission Index* (International Electrotechnical Commission, 2003), pp. 1–34.

[15]J. S. Bradley, "Relationships among measures of speech intelligibility in rooms," J. Audio Eng. Soc. **46**, 396–405 (1998).
[16]T. J. Vance, *The Sounds of Japanese* (Cambridge University Press, Cambridge, 2009), pp. 117–120.
[17]Y. Nishikawa, Hi. Sato, S. Inoue, and Y. Kobayashi, "Evaluation of speech transmission performance with impulse response in rooms (in Japanese)," J. Environ. Eng. **605**, 9–14 (2006).
[18]J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in C_{50} for speech," Appl. Acoust. **58**, 99–108 (1999).
[19]D. H. Hoth, "Room noise spectra at subscribers' telephone locations," J. Acoust. Soc. Am. **12**, 499–504 (1941).
[20]ANSI S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index* (American National Standards Institute, New York, 1997), pp. 1–22.
[21]WHO/PDH/91.1, *Report of the Informal Working Group on Prevention of Deafness and Hearing Impairment Program Planning* (World Health Organization, Geneva, 1991), pp. 1–2.
[22]WHO/PDH/97.3, *Report of the First Informal Consultation on Prevention of Deafness and Hearing Impairment Program Planning* (World Health Organization, Geneva. 1997), pp. 1–16.
[23]T. Houtgast and H. J. M. Steeneken, "A multi-language evaluation of the Rasti-method for estimating speech-intelligibility in auditoria," Acustica **54**, 185–199 (1984).
[24]I. Pollack, "Auditory informational masking," J. Acoust. Soc. Am. **57**, S5 (1975).
[25]L. Li, M. Daneman, J. G. Qi, and B. A. Schneider, "Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults?" J. Exp. Psychol. Hum. Percept. Perform. **30**, 1077–1091 (2004).
[26]T. L. Arbogast, C. R. Mason, and G. Kidd, Jr., "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180 (2005).
[27]T. R. Agus, M. A. Akeroyd, and S. Gatehouse, "Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise," J. Acoust. Soc. Am. **126**, 1926–1940 (2009).
[28]T. Gjestland and G. Oftedal, "Assessment of noise annoyance: The introduction of a threshold level in L_{eq} calculations," J. Sound Vib. **69**, 603–610 (1980).
[29]J. M. Fields and J. G. Walker, "Comparing the relationships between noise level and annoyance in different surveys: A railway noise vs. aircraft and road traffic comparison," J. Sound Vib. **69**, 51–80 (1982).
[30]J. M. Festen and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736 (1990).
[31]V. Mohanan, O. Sharma, and S. P. Singal, "A noise and vibration survey in an underground railway system," Appl. Acoust. **28**, 263–275 (1989).
[32]S. Yokoyama and H. Tachibana, "Study on the acoustical environment in public spaces," CD-ROM in *Proceedings of Inter-Noise 2008*, Shanghai, China (2008), pp. 1–8.
[33]Y. Soeta and R. Shimokura, "Change of acoustic characteristics caused by platform screen doors in train stations," Appl. Acoust. **73**, 535–542 (2012).
[34]Ha. Sato, M. Morimoto, and T. Takaoka, "Effects of spatial factors of speech, noise, and reverberation sounds on listening difficulty," J. Acoust. Soc. Am. **123**, 3608 (2008).
[35]T. Takaoka, M. Morimoto, Ha. Sato, and Y. Semba, "Effects of interaural cross-correlation of speech and background noise on listening difficulty," J. Acoust. Soc. Jpn. **63**, 520–528 (2007) (in Japanese).