

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Suitable Reverberation Criteria for Distant-talking Speech Recognition

Takanobu Nishiura and Takahiro Fukumori
Ritsumeikan University
Japan

1. Introduction

The recognition of distant-talking speech has rapidly improved in recent years, because many novel speech-recognition techniques have been proposed that are robust against noise and reverberance. The signal to noise ratio (SNR) is generally used as a common criterion in speech-recognition techniques that are robust against noise. SNR is an effective noise criterion for estimating the recognition of speech in noisy environments. As an algorithm based on the perceptual evaluation of speech quality (PESQ) (T. Yamada et al. (2006)) has also been proposed to achieve the same target, we can roughly estimate the recognition of speech in noisy environments. However, no common reverberation criteria have been proposed to attain robust reverberant-speech recognition. It has therefore been difficult to estimate the recognition of reverberant speech. The reverberation time, T_{60} , (M. R. Schroeder (1965)) is currently generally used to recognize distant-talking speech as a reverberation criterion. It is unique and does not depend on the position of the source in a room. However, distant-talking speech recognition greatly depends on the location of the talker relative to that of the microphone and the distance between them. Therefore, T_{60} is unsuitable for measuring the recognition of distant-talking speech. We propose newly reverberation criteria for measuring the recognition of distant-talking speech to overcome this problem. We first investigate suitable reverberation criteria to enable distant-talking speech to be recognized. We calculated automatic speech recognition with early and late reflections based on the impulse response between a talker and the microphone. We then evaluated it based on ISO3382 acoustic parameters (ISO3382 (1997)). Based on above investigation, we finally propose novel reverberation criteria RSR- D_n (Reverberant Speech Recognition criteria with D_n) which utilise ISO3382 acoustic parameters for robustly estimating reverberant speech recognition performance.

2. Conventional reverberation criteria for recognition of distant-talking speech

2.1 Reverberation time, (T_{60})

2.1.1 Reverberation time based on theory of room acoustics

Reverberation time (M. R. Schroeder (1965)) is the most fundamental concept for evaluating indoor acoustical fields and is a parameter that expresses the duration of sound. Reverberation time is the time required for a sound in a room to decay by 60 dB (called T_{60}). As the theory assumes a diffusible sound field in a room, the effect does not change even if sound-absorbing material is placed in any position in the room. The reverberation time is

constant for all positions of the sound source and the microphone in the room. However, it alone is insufficient as the criterion for the recognition of distant-talking speech because this depends on the distance between a talker and the microphone in the same environment.

2.1.2 Method of measuring reverberation time

Schroeder developed a basic method (M. R. Schroeder (1965)) of measuring reverberation by integrating the square of the impulse response. The reverberation time is easily measured with his method. The reverberation curves are derived from Eq. (1) with impulse response $h(\lambda)$.

$$\langle y_d^2(t) \rangle = N \int_t^\infty h^2(\lambda) d\lambda \quad (1)$$

where $\langle \rangle$ is the ensemble average, and N is the power of the unit frequency of random noise. The reverberation time in this reverberation curve is the time it takes to drop 60 dB below the original level.

2.2 Total amplitude of reflection signals (A value)

The A value (H. Kuttruff (2000)) is used as a reverberation criterion as often as reverberation time for the recognition of distant-talking speech. It is derived from Eq. (2).

$$A = \sqrt{\int_\epsilon^n h^2(t) dt / \int_0^\epsilon h^2(t) dt}, \quad (2)$$

where ϵ represents the duration of direct sound within approximately 3 – 5 ms. The A value indicates the energy ratio between direction and reflections on the captured signal, and it depends on the distance between the talker and microphone in the same room. However, it does not distinguish early reflections from late reverberations.

3. Relation between early reflections and distant-talking speech recognition

We define early reflections as high-correlation signals with direct sound, especially those that arrive within a few milliseconds of direct sound in this paper. Late reverberations are defined as low or no correlation signals with direct sound, especially those that arrive over a few milliseconds after direct sound.

3.1 Early reflections in distant-talking speech recognition

Early reflections, especially those that arrive within 50 ms of direct sound, are useful to humans when listening to speech (H. Kuttruff (2000)). However, the higher the reflection energy becomes, the less effectively speech is recognized, subject to clean acoustic phoneme models. However, it was previously unclear whether early reflections were useful for recognizing speech in the recognition of distant-talking speech because the reverberation time and A values were used as reverberation criteria. We evaluated what relation there was between early reflections and the recognition of distant-talking speech on the basis of impulse responses between a talker and the microphone to develop more suitable reverberation criteria for distant-talking speech recognition.

3.2 Evaluation experiment

3.2.1 Recording conditions

We measured impulse responses in actual environments. The impulse responses were measured in $T_{60} = 0.2$ and 0.7 s environments, subject to distances of 0.1 and 0.5 m between

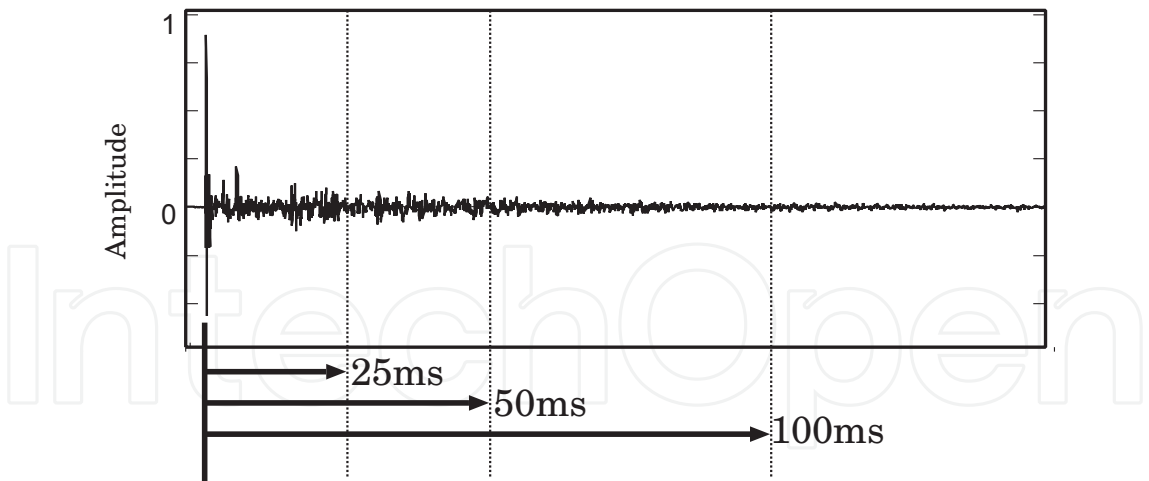


Fig. 1. Example of impulse response extraction for three evaluation periods.

Decoder	Julius (A. Lee et al. (2001))
HMM	IPA monophone model (Gender-dependent)
Feature vectors	12 orders MFCC + 12 orders Δ MFCC + 1 order Δ Power
Frame length	25 ms. (Humming window)
Frame interval	10 ms.

Table 1. Experimental conditions for speech recognition.

the talker and the microphone. A time stretched pulse (Y. Suzuki et al. (1995)) was used to measure the impulse responses. The recordings were made with 16 kHz sampling and 16 bit quantization.

3.2.2 Experimental conditions

An ATR phoneme-balanced set (K. Takeda et al. (1987)) was employed as the speech samples that were made up of 216 isolated Japanese words that were uttered by 14 speakers (7 females and 7 males). We evaluated the relation between early reflections and the recognition of distant-talking speech by convolving speech samples and impulse responses. Impulse responses were extracted during each period of evaluation as shown in Figure 1 to evaluate the relation between reflections and the recognition of distant-talking speech for all evaluation periods. Table 1 lists the experimental conditions for speech recognition.

3.2.3 Experimental results

Figure 2 plots the experimental results, where T_{60} is the reverberation time, $Dis.$ is the distance between the talker and the microphone, and WRR is the word recognition rate. The A value is the energy ratio between the direction and reflections from the duration of direct sound to each evaluation period. We confirmed that early reflections within about 12.5 ms after direct sound only contributed slightly to the recognition of distant-talking speech in quiet environments on the basis of these results, although early reflections within about 50 ms from the duration of direct sound contributed greatly to human hearing ability. We also confirmed that late

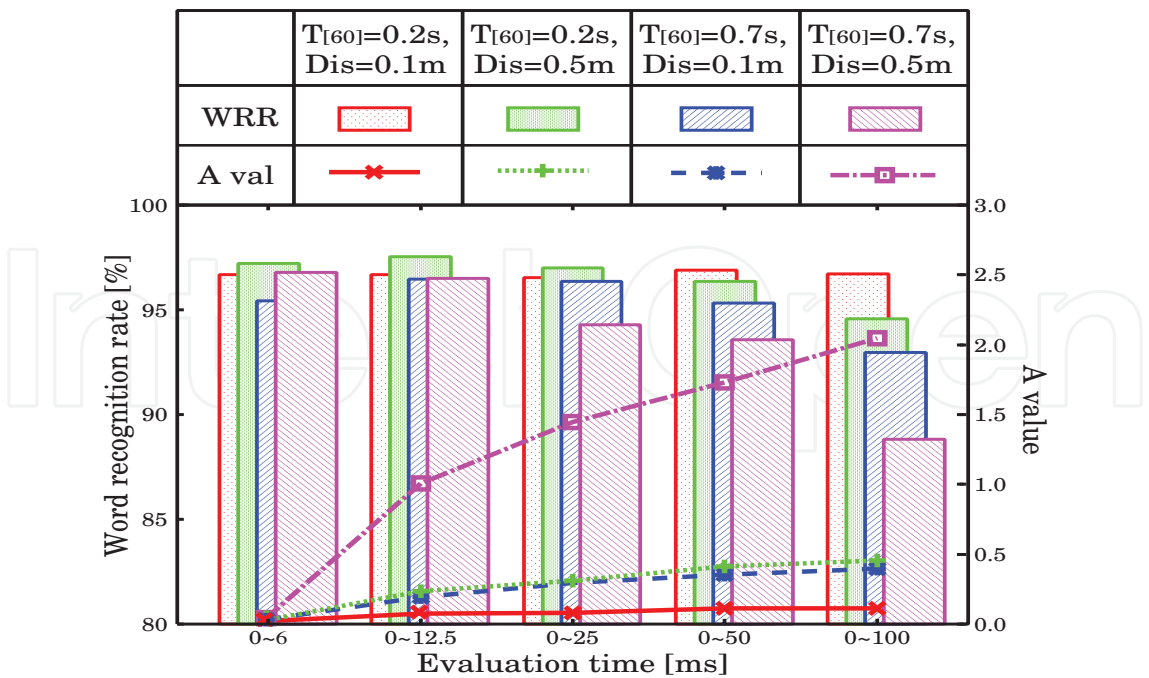


Fig. 2. Effects of early reflections on distant-talking speech recognition.

reflections over about 12.5 ms after direct sound decreased the recognition of distant-talking speech. The higher the *A* value becomes in Figure 2, the greater the number of reflections. However, we confirmed that the ability to recognize speech can be improved despite a higher *A* value. Therefore, we again found that suitable reverberation criteria were necessary for the recognition of distant-talking speech on the basis of our evaluation experiments.

4. Toward suitable reverberation criteria

4.1 ISO3382 acoustic parameters

ISO3382 (ISO3382 (1997)) proposed parameters for measuring room acoustics. The ISO3382 standard defines measurements of reverberation times in rooms with reference to other acoustical parameters. Acoustics parameters are classified into four categories on the basis of this standard:

- 1. Sound level
- 2. Reverberation time
- 3. Balance between early and late arriving energies (Clarity, Definition, and Center time)
- 4. Binaural parameters (IACC, Lateral Fraction)

These parameters are directly calculated based on measured impulse responses. We focused on the third category (balance between early and late arriving energies), because it has a high correlation with clarity and the reverberance of the acoustic sound field.

4.2 Balance between early and late arriving energy

“Clarity,” “Definition,” and “Center time” are defined as the acoustic parameters of balance between early and late arriving energies in the ISO3382 standard. The *C* value expresses the clarity of acoustics and is derived from Eq. (3). The *D* value expresses the definition of

acoustics and is derived from Eq. (4). Center time expresses the center time based on a square impulse response and is derived from Eq. (5).

$$C_n = 10 \log_{10} \left(\int_0^n h^2(t) dt / \int_n^\infty h^2(t) dt \right), \quad (3)$$

$$D_n = \int_0^n h^2(t) dt / \int_0^\infty h^2(t) dt, \quad (4)$$

$$T_s = \int_0^\infty t h^2(t) dt / \int_0^\infty h^2(t) dt, \quad (5)$$

where, n is the border time between early and late arriving energies. The C value measure and the condition of music are highly correlated with $n = 80 \text{ ms}$, and the D value measure and the condition of speech are highly correlated with $n = 50 \text{ ms}$ based on the ISO3382 standard. In addition, the larger T_s becomes, the more late reverberations there are.

4.3 Evaluation experiments

We evaluated the relation of the ISO3382 acoustic parameters and the recognition of distant-talking speech to determine suitable reverberation criteria. We also compared all acoustic parameters with regression analysis based on ordinary least squares.

4.3.1 Recording conditions

We measured impulse responses in six environments, i.e., a "Living room" (LV, $T_{60} = 250 \text{ ms}$), a "Conference room" (CR, $T_{60} = 350 \text{ ms}$), a "Corridor" (CC, $T_{60} = 600 \text{ ms}$), a "Prefabricated bath" (PB, $T_{60} = 700 \text{ ms}$), an "Elevator hall(lobby)" (EV, $T_{60} = 700 \text{ ms}$), and "Standard stairs" (SS, $T_{60} = 800 \text{ ms}$). The distances between the talker and the microphone were between 10 cm and 500 cm in all environments. We measured 307 impulse responses in all. A time-stretched pulse was used to measure the impulse responses as in Section 3.2.1. The recordings were conducted with 16 kHz sampling and 16 bit quantization.

4.3.2 Experimental conditions

The speech recognition experiments were conducted under the same conditions as in Section 3.2.2. An ATR phoneme-balanced set was employed as the speech samples that were made up of 216 isolated Japanese words that were uttered by 14 speakers (7 females and 7 males). Table 1 lists the experimental conditions for speech recognition.

4.3.3 Experimental results

Figures 3-6 plot the experimental results. The horizontal axes represent the word recognition rate, and the vertical axes represent the A value, C_{80} , D_{50} , and T_s . Table 2 lists the results for all acoustic parameters with regression analysis based on ordinary least squares. We found that the ISO3382 acoustic parameters were strong candidates for the reverberation criteria based on these results because the regression coefficients for the C , D , and T_s values were higher than that for the A value.

4.3.4 Discussion

The results from the evaluation experiments proved the ISO3382 acoustic parameters were strong candidates as the reverberation criteria for the recognition of distant-talking speech. We therefore assumed that the early reflection signal, which is the most important factor in

	LV	CR	CD	PB	EV	SS	AVE.
A	0.81	0.93	0.79	0.89	0.81	0.69	0.82
C_{80}	0.82	0.86	0.85	0.96	0.91	0.89	0.88
D_{50}	0.73	0.91	0.93	0.95	0.92	0.91	0.89
T_s	0.82	0.87	0.95	0.97	0.91	0.77	0.88

Table 2. Regression coefficients for all acoustic parameters.

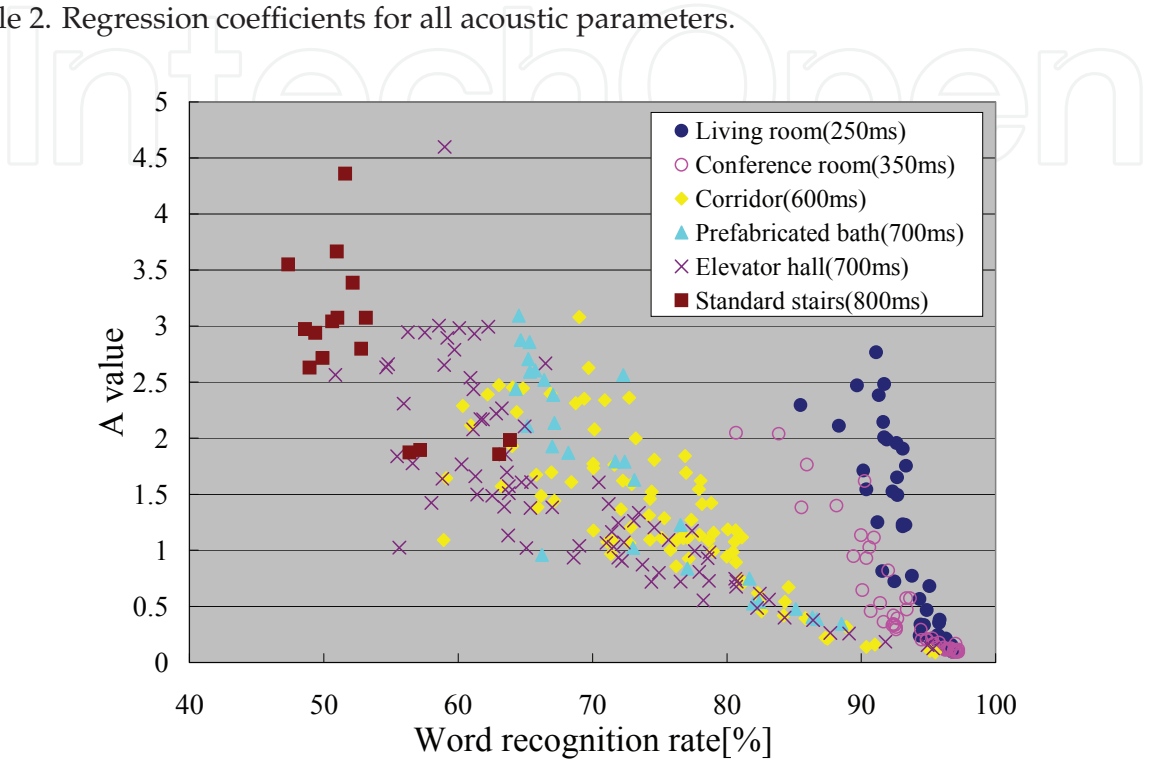


Fig. 3. *A value*.

the recognition of reverberant speech, does not depend on the total amount of reflection, but on the balance between early and late arriving energies. Our next challenge is to examine the use of suitable reverberation criteria based on the C_n and D_n values of ISO3382 acoustic parameters with a suitable border time, n , between early and late arriving energies to prove this hypothesis. If suitable border time n can be estimated, we can easily estimate the recognition of reverberant speech with one impulse response between the talker and microphone.

5. Performance estimation of reverberant speech recognition based on reverberation criteria

5.1 Performance estimation based on reverberation time

Reverberation time is usually used to estimate reverberant speech recognition performance. However, other reverberant features are altered by the difference between assumption of a diffusible sound field in a room and an actual sound field. Thus, it is difficult to estimate speech recognition performance with only reverberation time. In this section, we conducted an evaluation experiment in three reverberant environments shown in Table 3(b) to investigate the relation between reverberation time and speech recognition performance. We first

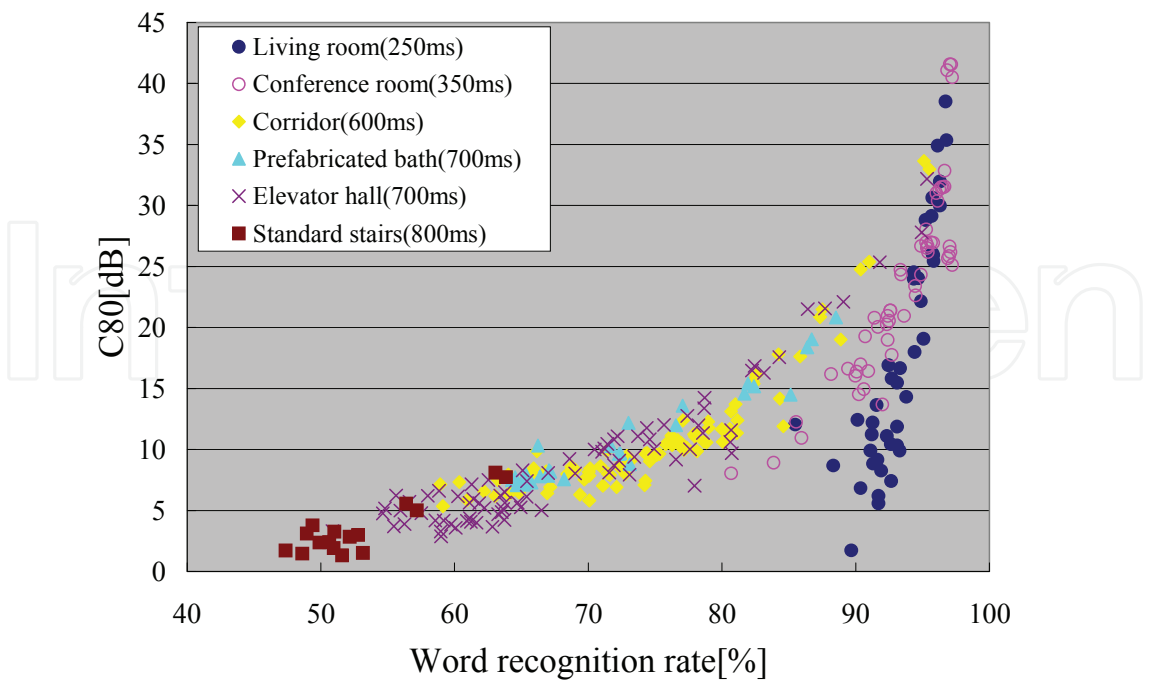


Fig. 4. Clarity (C_{80}).

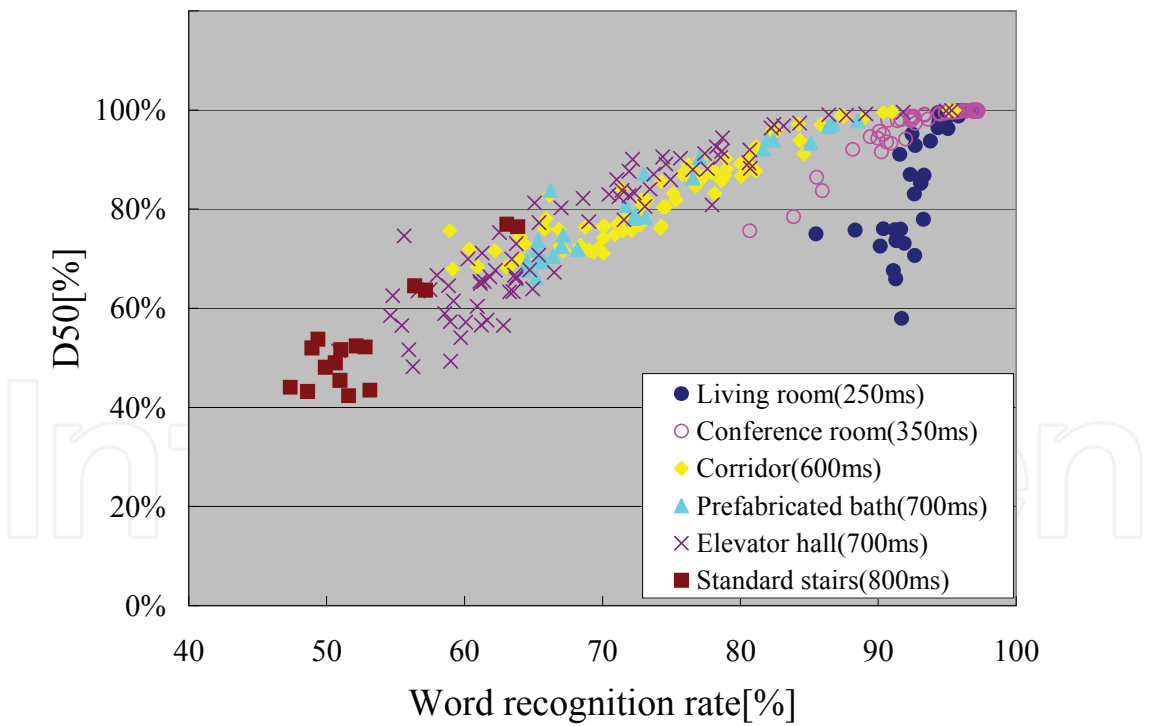


Fig. 5. Definition (D_{50}).

measured several impulse responses in each environment. After that, we acquired speech recognition performance with a speech recognition engine (A. Lee et al. (2001)) by using the training data convolved speech sample and each measured impulse response. Figure 7

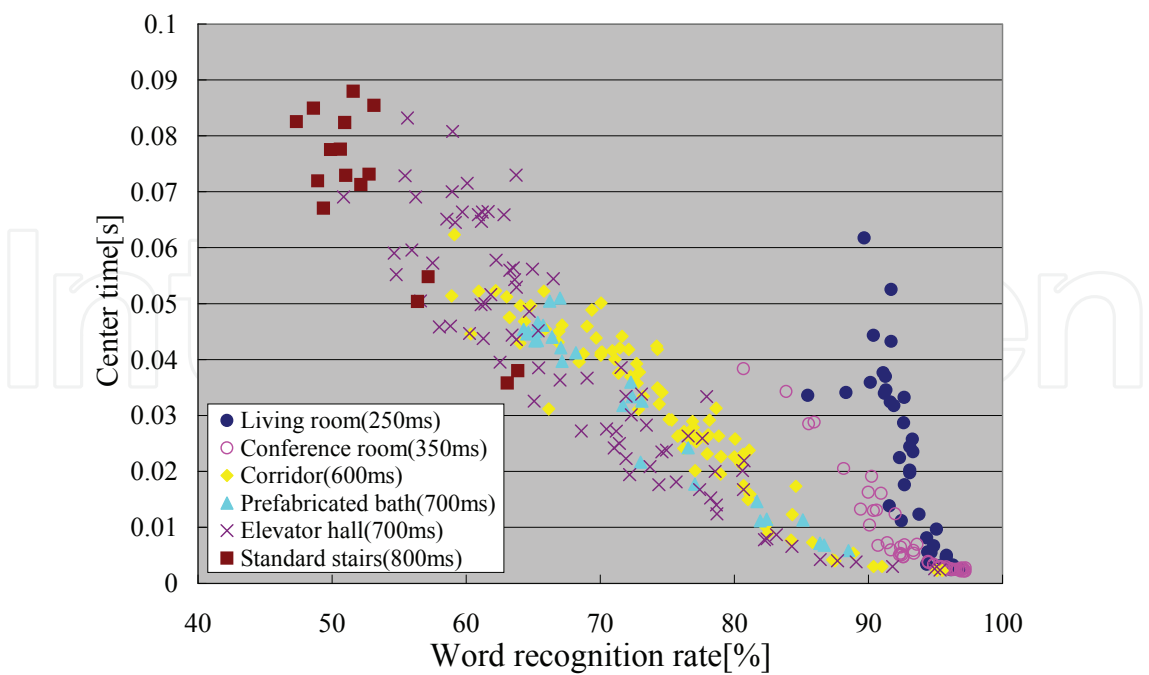


Fig. 6. Center time (T_s).

shows the obtained result. The line in Figure 7 represents the average of speech recognition performance in each reverberant environment. We confirmed the speech recognition performance degradation and the variance increase in heavy reverberant environment. As a result, we could confirm that it is significantly difficult to estimate speech recognition performance in heavy reverberation environment in comparison with light reverberant environment.

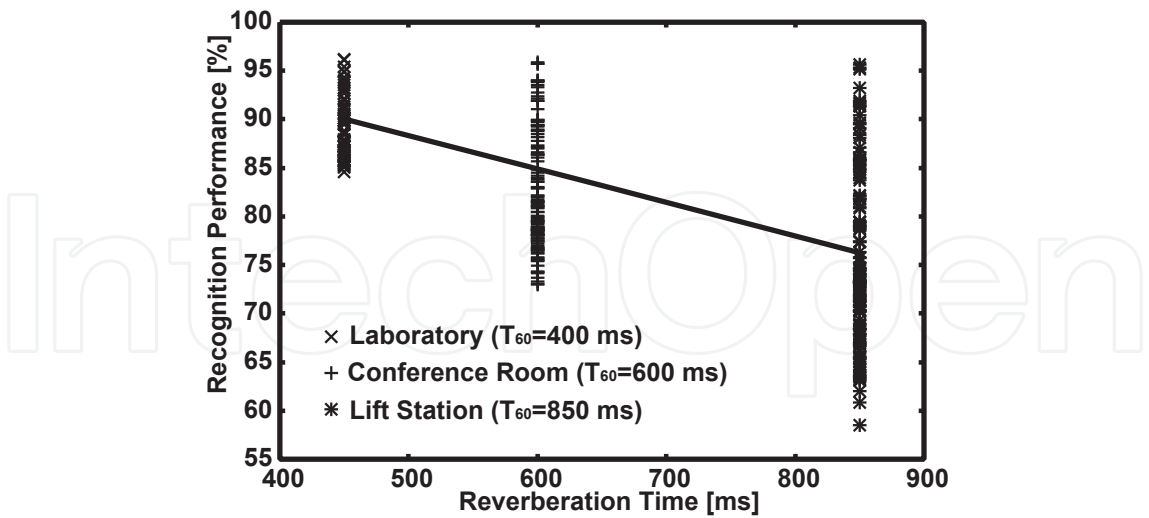


Fig. 7. Reverberant speech recognition performance in three reverberant environments.

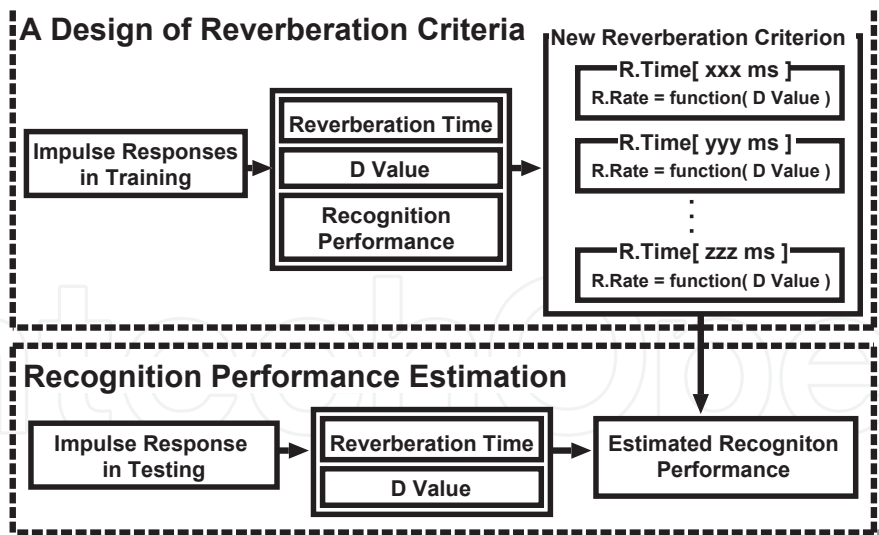


Fig. 8. Overview of the proposed method.

5.2 Performance estimation based on new reverberation criteria $RSR-D_n$

5.2.1 Early reflections in reverberant speech recognition

In previous section 3.2.3, we confirmed two facts about reverberant speech recognition. One is that early reflections within about 12.5 ms after direct sound contributed slightly to the recognition of reverberant speech in quiet environments, although early reflections within about 50 ms from the duration of direct sound contributed greatly to human hearing ability. The other is that late reflections over about 12.5 ms after direct sound decreased the recognition of reverberant speech. Based on these results, we confirmed that it is difficult to estimate the reverberant speech recognition performance using only reverberation time, since it does not take these factors into consideration. Therefore, we concluded that we would need to use the experimental results we had previously obtained to determine suitable reverberation criteria for recognizing reverberant speech.

5.2.2 New reverberation criteria with $RSR-D_n$

We attempted to design the new reverberation criteria $RSR-D_n$ to estimate reverberant speech recognition performance as shown at the top of Figure 8. First, we investigated the relation between the D value and reverberant speech recognition performance. We then used regression analysis based on the correlation coefficients for these to design the $RSR-D_n$ to cover each reverberation time. We used four steps in our approach, explained in detail as follows.

- Step.1:** We measured many impulse responses in a number of environments to obtain training data. Using the measured impulse responses as a basis, we used Eq. (1) to calculate reverberation times.
- Step.2:** We next calculated the D value with Eq. (4) after performing Step 1. In Eq. (4), the border time n is essential for determining the maximum value of the relation between D value and speech recognition performance. Thus, we determined the suitable border time n as described in Section 5.3.1 and then used the value to calculate D_n .
- Step.3:** We then acquired speech recognition performance with a speech recognition engine (A. Lee et al. (2001)) by using the training data obtained using dry data and measured impulse responses as described in Step 1.

(a) Training environments	
Soundproof room	(T_{60} = 100 ms, 72 RIRs)
Japanese style room	(T_{60} = 400 ms, 72 RIRs)
Laboratory	(T_{60} = 450 ms 72 RIRs)
Conference room	(T_{60} = 600 ms, 120 RIRs)
Living room	(T_{60} = 600 ms, 72 RIRs)
Corridor	(T_{60} = 600 ms, 120 RIRs)
Bath room	(T_{60} = 650 ms, 28 RIRs)
Lift station	(T_{60} = 850 ms, 120 RIRs)
Standard stairs	(T_{60} = 850 ms, 56 RIRs)
(b) Environments to calculate speech recognition performance	
Laboratory	(T_{60} = 450 ms, 72 RIRs)
Conference room	(T_{60} = 600 ms, 120 RIRs)
Lift station	(T_{60} = 850 ms, 120 RIRs)
(c) Environments to calculate a suitable n	
Japanese style room	(T_{60} = 400 ms, 72 RIRs)
Conference room	(T_{60} = 600 ms, 120 RIRs)
Standard stairs	(T_{60} = 850 ms, 56 RIRs)
(d) Environments to design $RSR-D_n$	
Japanese style room	(T_{60} = 400 ms, 72 RIRs)
Conference room	(T_{60} = 600 ms, 120 RIRs)
Standard stairs	(T_{60} = 850 ms, 56 RIRs)
(e) Test environments	
Laboratory	(T_{60} = 450 ms, 72 RIRs)
Bath room	(T_{60} = 650 ms, 28 RIRs)
Lift station	(T_{60} = 850 ms, 120 RIRs)

Table 3. Experimental conditions(RIRs : Room Impulse Responses)

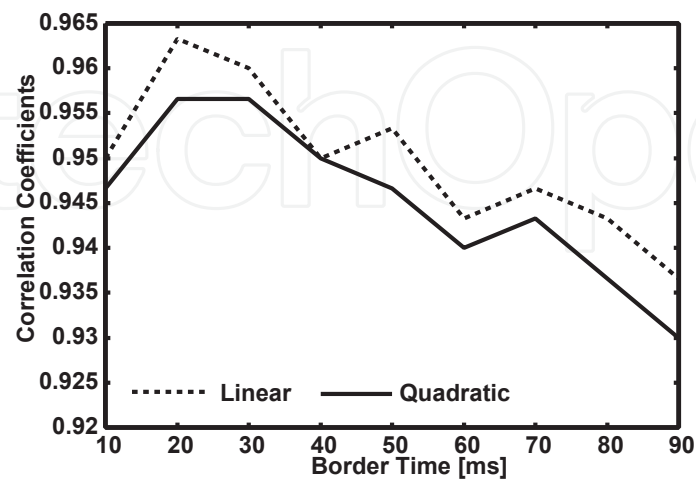


Fig. 9. Relation between correlation coefficient in each regression curve and border time n

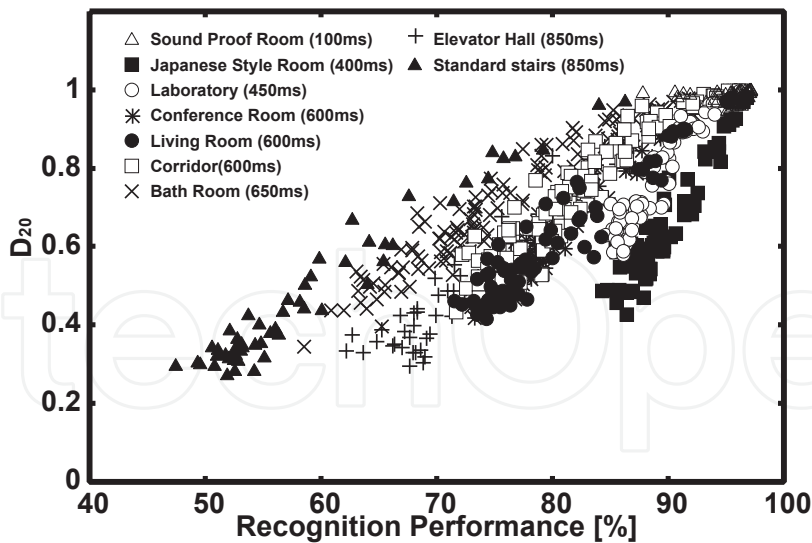


Fig. 10. Relation between D_{20} and speech recognition performance (Overall)

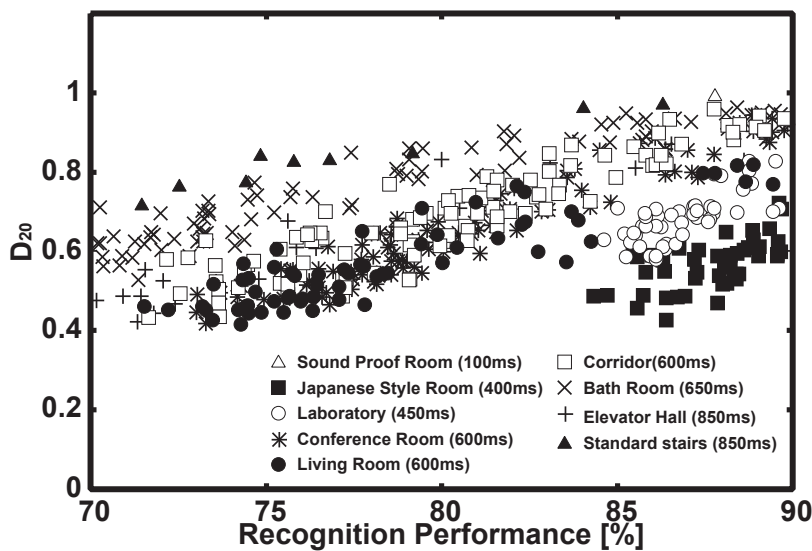


Fig. 11. Relation between D_{20} and speech recognition performance (Close-up)

Step.4: Finally, we conducted regression analysis based on the D value calculated from Steps 1 and 2 and the speech recognition performance calculated in Step 3. We used linear and quadratic functions as regression curves calculated with regression analysis based on ordinary least squares.

5.2.3 Performance estimation of reverberant speech recognition with $RSR-D_n$

As shown at the bottom of Figure 8, we will try to estimate the speech recognition performance with the $RSR-D_n$. We first calculate the reverberation time and the D value based on impulse responses in test environments. Based on them, we try to estimate the speech recognition performance with the $RSR-D_n$ in same reverberation time.

5.3 Evaluation experiments

We used the proposed criteria to estimate the reverberant speech recognition performance. Initially, we measured 732 impulse responses to design the reverberant criteria $RSR-D_n$ in the nine training environments shown in Table 3(a). A time-stretched pulse was used to measure the impulse responses. The recordings were conducted with 16 kHz sampling and 16 bit quantization. All impulse responses were measured for distances ranging between 100 ~ 5,000 mm. For estimation of speech recognition performance, we used an ATR phoneme-balanced set as the speech samples that were made up of 216 isolated Japanese words that were uttered by 14 speakers (7 females and 7 males). In addition, the recognition performance varies largely depending on the recognition task. Thus, $RSR-D_{20}$ design and performance estimation should be conducted in the same recognition task.

5.3.1 Suitable border time n for reverberant criteria $RSR-D_n$

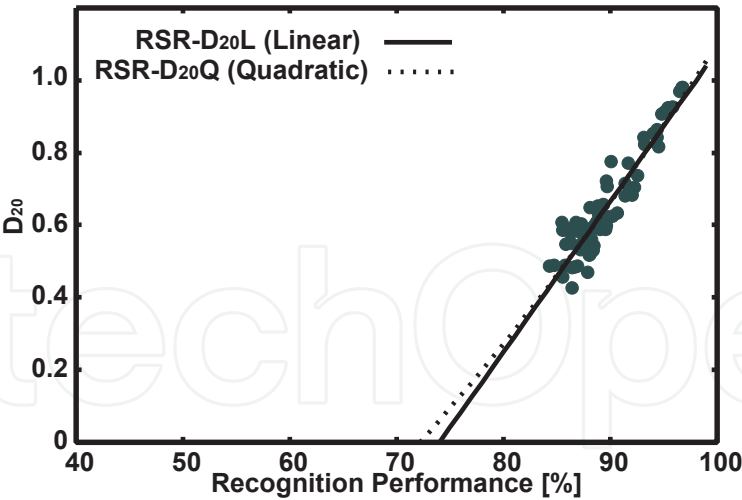
In Eq. (4), the border time n is essential for determining the maximum value of the relation between D value and speech recognition performance. Thus, we conducted evaluation experiments in the three environments shown in Table 3(c), using the D value and two regression functions (linear and quadratic) to determine the most suitable border time n . Figure 9 shows the results we obtained. From linear and quadratic regression analysis, it was determined that 20 msec was the most suitable border time value. We therefore used 20 msec as the border time for calculating D_n and designing $RSR-D_{20}$.

5.3.2 Suitable $RSR-D_{20}$ design

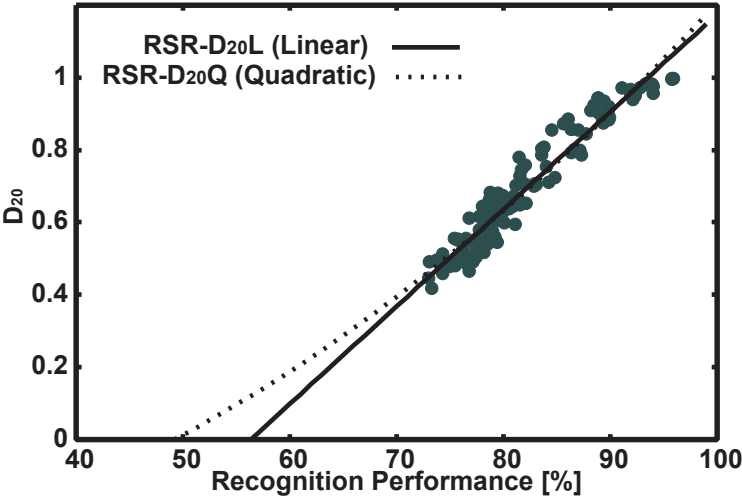
Figure 10 and 11 show the relation between speech recognition performance and D_{20} for the nine training environments shown in Table 3(a). These figures also show the regression analysis results for the three environments shown in Table 3(d). Figure 12 shows the relation between $RSR-D_{20}$ and speech recognition performance based on the regression analysis results in three environments (Japanese room, Conference room and Standard stairs). Table 4 shows correlation coefficients with their respective regression functions for these three environments. We defined that $RSR-D_{20}L$ represents $RSR-D_{20}$ with a linear regression function, and $RSR-D_{20}Q$ represents $RSR-D_{20}$ with a quadratic regression function. As a result of Table 4, we confirmed that both $RSR-D_{20}L$ and $RSR-D_{20}Q$ are much the most suitable criteria for estimation of reverberant speech recognition.

5.3.3 Performance estimation with $RSR-D_{20}Q$

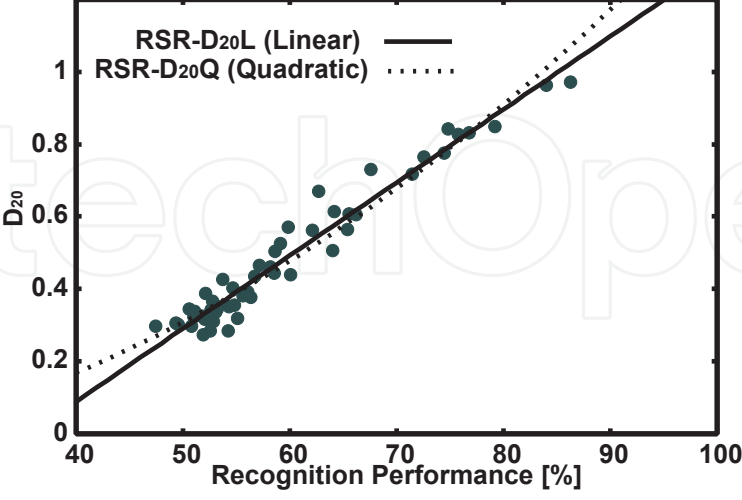
Finally, we attempted to estimate the reverberant speech recognition performance for the three test environments shown in Table 3(e). Both closed and open tests were carried out for this purpose. In closed test, we estimated speech recognition performance on known condition with $RSR-D_n$ designed in the same environment. On the other hand, in open test, we estimated recognition performance on unknown condition with $RSR-D_n$ designed in the other environments including same reverberation time. Figure 13 shows the obtained results. Standard deviations are given in Table 5. The results showed that average estimation error of less than 5% was achieved with $RSR-D_{20}Q$ in all environments. Table 4 shows the correlation coefficients obtained with $RSR-D_{20}L$ and $RSR-D_{20}Q$. As the table shows, both the $RSR-D_{20}L$ and $RSR-D_{20}Q$ coefficients are higher than 0.93 in all environments. Thus, it can be concluded that the $RSR-D_{20}$ criteria provides much better estimation performance than conventional reverberation criteria and that it is a particular strong candidate for suitably recognizing reverberant speech.



(a) Japanese room

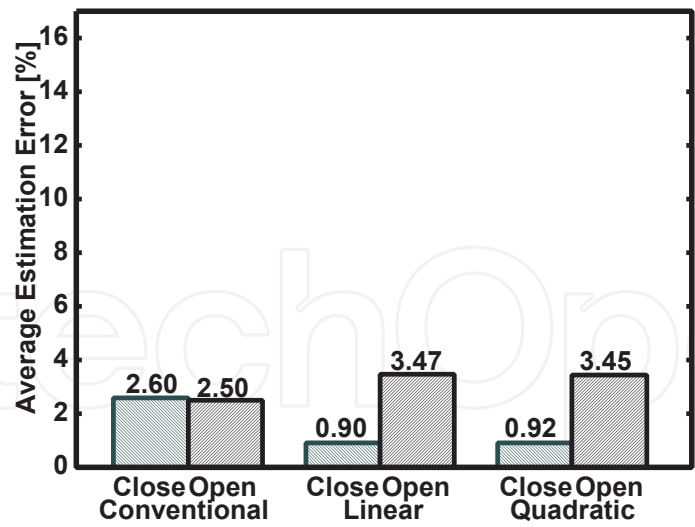


(b) Conference room

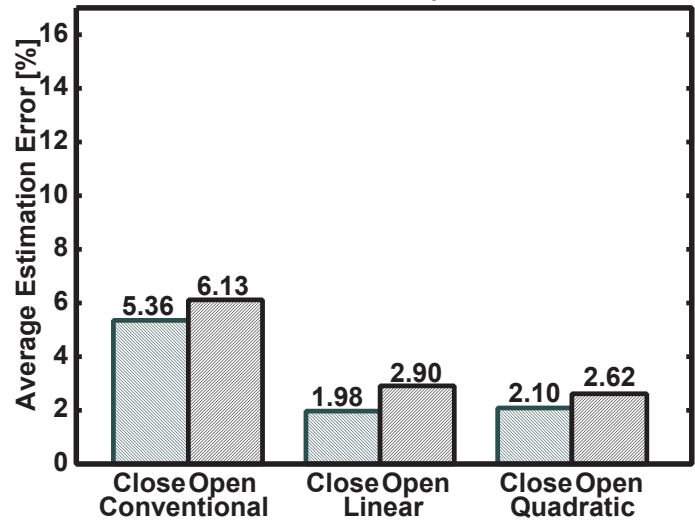


(c) Standard stairs

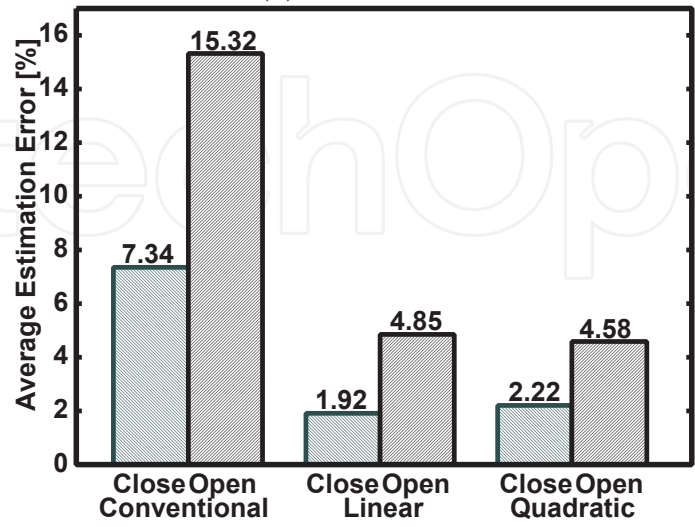
Fig. 12. Relation between $RSR-D_{20}$ and speech recognition performance



(a) Laboratory



(b) Bath room



(c) Lift station

Fig. 13. Average estimation error

Env.	RSR- D_{20} L (Linear)	RSR- D_{20} Q (Quadratic)
$T_{60} = 400$ ms	0.937	0.939
$T_{60} = 600$ ms	0.966	0.963
$T_{60} = 850$ ms	0.977	0.972

Table 4. Correlation coefficients

Env.	Conventinal Method Close / Open	RSR- D_{20} L (Linear) Close / Open	RSR- D_{20} Q (Quadratic) Close / Open
$T_{60} = 450$ ms	3.10 / 3.26	1.10 / 3.62	1.13 / 3.60
$T_{60} = 650$ ms	6.92 / 7.18	2.46 / 3.49	2.59 / 3.14
$T_{60} = 850$ ms	8.80 / 17.64	2.41 / 5.35	2.81 / 5.23

Table 5. Standard deviations

6. Conclusions

We first evaluated the relation between early reflections and the recognition of distant-talking speech toward suitable reverberation criteria to enable distant-talking speech to be recognized. As a result, we found that early reflections within about 12.5 ms from the duration of direct sound contributed slightly to the recognition of distant-talking speech in non-noisy environments. We also confirmed that the *C* and *D* values of ISO3382 were strong candidates for the reverberation criteria of distant-talking speech recognition as a result of evaluation experiments with ISO3382 acoustic parameters. Therefore, to facilitate the recognition of reverberant speech, we then proposed new reverberation criteria RSR- D_{20} (Reverberant Speech Recognition criteria with D_{20}), which calculates recognition performance based on D_{20} for ISO3382 acoustic parameters. Experiments conducted in actual environments confirmed that the proposed criteria (particularly RSR- D_{20} Q) provide much better estimation performance than conventional reverberation criteria. We also intend to investigate suitable reverberation criteria in the frequency domain for distant-talking speech recognition with the Modulation Transfer Function (MTF) (T. Houtgast et al. (1980)) in future work.

7. Acknowledgments

This work was partly supported by Global COE and Grants-in-Aid for Scientific Research funded by Japan’s Ministry of Education, Culture, Sports, Science, and Technology of Japan. This work was also partly supported by the "Ono Acoustics Research Fund".

8. References

T. Yamada, M. Kumakura, and N. Kitawaki, "Performance estimation of speech recognition system under noise conditions using objective quality measures and artificial voice," IEEE Trans. on ASLP, Vol. 14, No. 6, pp. 2006-2013, Nov. 2006.

M. R. Schroeder, "New method of measuring reverberation time," J. Acoust. Soc. Am., Vol. 37, pp. 409-412, 1965.

ISO3382: Acoustics- Measurement of the reverberation time of rooms with reference to other acoustical parameters. International Organization for Standardization, 1997.

- H. Kuttruff, "Room Acoustics," Spon Press, 2000.
- Y. Suzuki, F. Asano, H.-Y. Kim, and Toshio Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses", J. Acoust. Soc. Am. Vol. 97 (2), pp. 1119-1123, 1995.
- K. Takeda, Y. Sagisaka, and S. Katagiri, "Acoustic-Phonetic Labels in a Japanese Speech Database," Proc. European Conference on Speech Technology, Vol. 2, pp. 13-16, Oct. 1987.
- A. Lee, T. Kawahara, and K. Shikano, "Julius - an open source real-time large vocabulary recognition engine," In Proc. European Conf. on Speech Communication and Technology, pp. 1691-1694, 2001.
- T. Houtgast, H.J.M. Steeneken, and R. Plomp, "Predicting speech intelligibility in room acoustics," Acustica, Vol. 46, pp. 60-72, 1980.

IntechOpen



Speech Technologies

Edited by Prof. Ivo Ipsic

ISBN 978-953-307-996-7

Hard cover, 432 pages

Publisher InTech

Published online 23, June, 2011

Published in print edition June, 2011

This book addresses different aspects of the research field and a wide range of topics in speech signal processing, speech recognition and language processing. The chapters are divided in three different sections: Speech Signal Modeling, Speech Recognition and Applications. The chapters in the first section cover some essential topics in speech signal processing used for building speech recognition as well as for speech synthesis systems: speech feature enhancement, speech feature vector dimensionality reduction, segmentation of speech frames into phonetic segments. The chapters of the second part cover speech recognition methods and techniques used to read speech from various speech databases and broadcast news recognition for English and non-English languages. The third section of the book presents various speech technology applications used for body conducted speech recognition, hearing impairment, multimodal interfaces and facial expression recognition.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Takanobu Nishiura and Takahiro Fukumori (2011). Suitable Reverberation Criteria for Distant-talking Speech Recognition, *Speech Technologies*, Prof. Ivo Ipsic (Ed.), ISBN: 978-953-307-996-7, InTech, Available from: <http://www.intechopen.com/books/speech-technologies/suitable-reverberation-criteria-for-distant-talking-speech-recognition>

INTech
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2011 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen