Available online at www.sciencedirect.com
ScienceDirect
Journal of Sound and Vibration 297 (2006) 1088–1096
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Short communication

Study on achieving speech privacy using masking noise

Takahiro Tamesue[a,*], Shizuma Yamaguchi[b], Tetsuro Saeki[b]

[a] *Nagano National College of Technology, 716 Tokuma, Nagano 381-8550, Japan*
[b] *Faculty of Engineering, Yamaguchi University, 2-6-11 Tokiwadai, Ube 755-8611, Japan*

Received 1 March 2003; received in revised form 8 May 2006; accepted 8 May 2006
Available online 30 June 2006

Abstract

This study focuses on achieving speech privacy using a meaningless steady masking noise. The most effective index for achieving a satisfactory level of speech privacy was selected, choosing between spectral distance and the articulation index. From a result, spectral distance was selected as the best and most practical index for achieving speech privacy. Next, speech along with a masking noise with a sound pressure level value corresponding to various speech privacy levels were presented to subjects who judged the psychological impression of the particular speech privacy level. Theoretical calculations were in good agreement with the experimental results.

1. Introduction

In recent years, importance has been attached to achieving speech privacy [1] in open spaces such as for oral consultations near waiting rooms of small-scale clinics, tax-payment consultations at a taxation office window, course consultations in school classrooms, and legal aid services in temporary booths. Generally, although measures such as the use of sound partition are instituted in many cases, measures that use other sounds to mask speech by emitting sound other than speech have also been considered. A way of masking meaningful speech with meaningless noise would have great benefits. Another way is to mask speech is with background music (BGM); however, this method can itself lead to further disturbances because music is a meaningful sound and is sometimes difficult to choose in various business/living environments in which many people are working/interacting together [2].

From the above view point, previous studies focused on masking meaningful speech noise with meaningless steady noise; this noise at the minimum sound pressure level for robustly masking male or female meaningful speech that has various power spectra forms was considered from a practical viewpoint. As a result, band-limited pink noise with [180, 5630] [Hz] frequency band-width can now be selected as the most effective noise for masking speech [3].

In this paper, we evaluate speech privacy and listening score when speech is masked with the above noted band-limited pink noise. First, taking notice of two indices, the articulation index [4] and spectrum distance [5] which is newly introduced, the relationships between these indices and the speech privacy/listening score are

*Corresponding author. Tel./fax: +81 26 295 7090.
E-mail address: tamesue@ei.nagano-nct.ac.jp (T. Tamesue).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2006.05.012

first discussed using regression models. Next, to verify the stability of these set-up regression functions, we conduct a psychological experiment in which both speech and a masking noise, with a sound pressure level corresponding to the degree of speech privacy, are presented to subjects for assessment of their evaluation of the speech privacy and listening score. The predicted values for evaluation for speech privacy and listening score from regression functions are in good agreement with the values observed in the psychological experiment.

2. Outline of Experiment I

To examine which of articulation index and spectral distance is an effective index to decide a sound pressure level for a masking noise corresponding to the degree of speech privacy determined beforehand, Experiment I was conducted. In outline, Experiment I was as follows.

2.1. Location

The experiment was conducted in a simple soundproof room on campus having the following dimensions: length 5.1 m, width 3.3 m, and height 2.2 m. The sound pressure level of the background noise was about 37 dB. (A weighted sound pressure level was about 21 dB. The sound pressure level in this paper is the value measured by a sound level meter with FLAT response.)

2.2. Subjects

Subjects were eight male students with normal hearing.

2.3. Presented sound

2.3.1. Experiment I-1

(a) *Audio signal*: Male speech, made by deleting handclaps, sound effects, and music, etc. from commercially available speech tapes, was used in the experiment. Maximum band levels of speech measured with a sound level meter were adopted as the band levels of speech peaks. The overall sound pressure level of the speech peaks was about 65 dB. The power spectrum of the speech peaks is shown in Fig. 1.

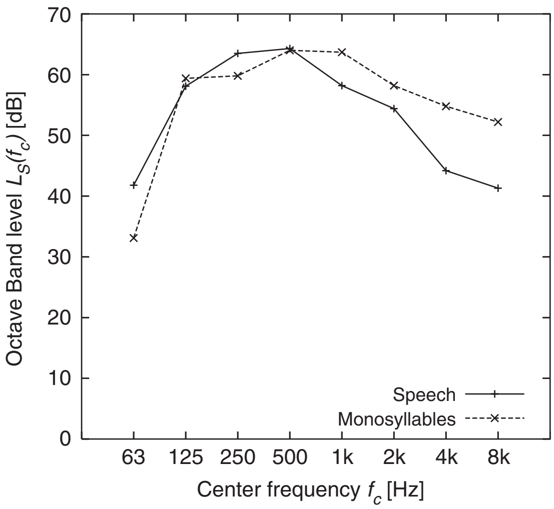

Fig. 1. Octave band level of speech peaks.

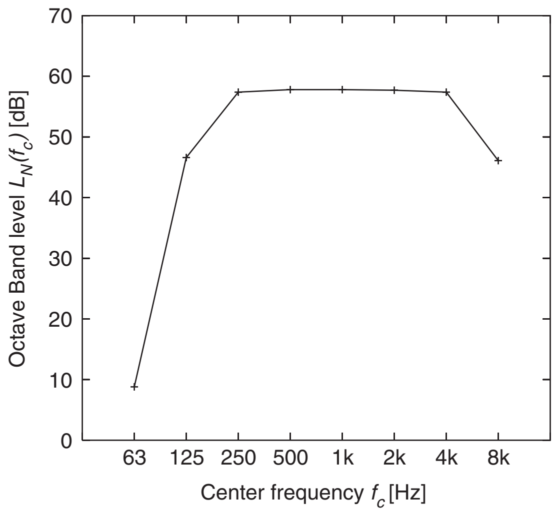

Fig. 2. Octave band level of masking noise.

(b) *Masking noise*: Band-limited pink noise with [180, 5630] Hz frequency band-width was selected as the best and most practical meaningless steady noise for robustly masking speech, which has various forms of power spectral levels, at a minimum sound pressure level [3], was used. The sound pressure levels of the masking noise were set at 50, 53, 56, 59, 62, 65, 68, 71, 74, 77 dB. The power spectrum of the masking noise is shown in Fig. 2.

2.3.2. Experiment I-2

(a) *Audio signal*: A monosyllable list (a list contained 50 monosyllables) from a CD for the evaluation and fitting condition of hearing aids (TY-89) [6], was used. The band levels of speech peaks was also shown in Fig. 1.
(b) *Masking noise*: The same masking noise as presented in Experiment I-1.

2.4. Measurement method

Eight subjects simultaneously participated in the psychological experiment. Both the audio signal and the masking noise were transmitted from the speakers to all subjects. It was confirmed ahead of time that there was no difference in the sound pressure levels at the subjects' ears. The above operations were executed twice in each of the cases.

2.4.1. Experiment I-1

After listening to the speech for 30 s, an evaluation of the speech privacy was made using the following evaluation scales [2]:

F_1: Did not notice talking.
F_2: Possible to recognize the sound as a voice, though the contents of speech were not understood.
F_3: When listened to carefully, the contents of the speech were understood to some extent.
F_4: Even if not intensively listened to, all the contents of the speech were understood.

2.4.2. Experiment I-2

Subjects noted the monosyllables exactly as they heard them. Then the number of correct answers were assessed. The listening score was defined as the percentage of correct monosyllables from the total (50).

3. Psychological intervals of the evaluation scale for speech privacy

It was considered whether the evaluation scale for speech privacy can be treated as interval scale. Using the observed data from Experiment I, psychological intervals [7] among the four-point scales were obtained (A results are omitted). Since the correlation coefficient was 0.999, a value near 1, it was satisfactory to treat the ordinal scale $F_i(i = 1, 2, \ldots, 4)$ as an interval scale $F_i = i(i = 1, 2, \ldots, 4)$.

4. Selection of index to set up the degree of speech privacy

In this paper, two indices, articulation index and spectral distance, were introduced. We considered what was an effective index for deciding the sound pressure level of masking noise necessary to achieve speech privacy.

4.1. Relationships between indices and evaluations of speech privacy

Based on the results of Section 3, treating ordinal scale F_i $(i = 1, 2, \ldots, 4)$ as an interval scale $F_i = i$ $(i = 1, 2, \ldots, 4)$, the relationships between each index and the averaged value of the evaluations for every noise condition were investigated using the observed data from Experiment I. Fig. 3 shows the relationship between articulation index and the evaluation of speech privacy. Points in this figure represent averaged values of the evaluations. Dotted line represents regression function selected by Akaike's information criterion (AIC) [8] from the following types of model describing regression between them:

Logistic function:

$$y = \frac{k - c}{1 + a\exp(-bx)} + c. \tag{1}$$

Modified exponential function:

$$y = (k - c)\left(1 - \exp\left(-\frac{x - b}{a}\right)\right) + c. \tag{2}$$

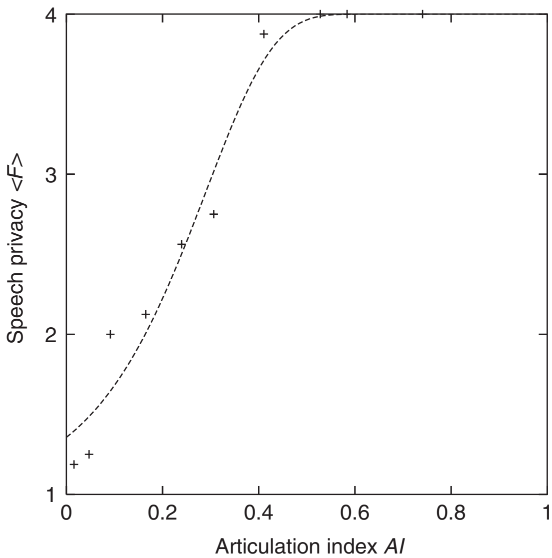

Fig. 3. Relationships between articulation index and evaluation of speech privacy.

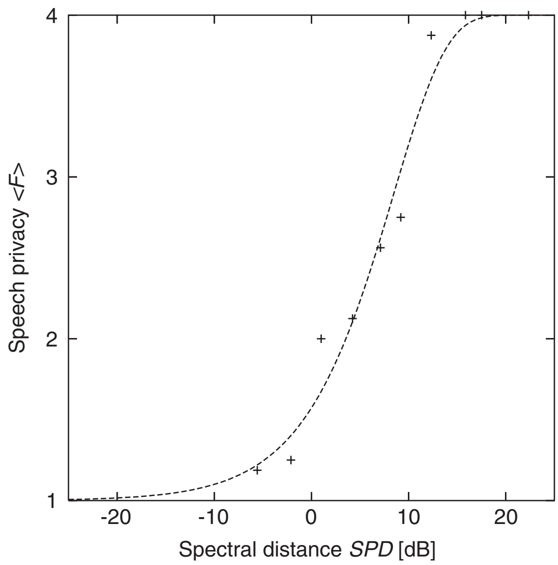

Fig. 4. Relationships between spectral distance and evaluation of speech privacy.

Gonperz function:

$$y = (k - c)\exp(-\exp(-a(x - b))) + c. \tag{3}$$

This figure reveals that it is difficult to set up the degree of speech privacy using the articulation index since the vertical deviations from observed data to regression curve are large. Fig. 4 shows the relationship between spectral distance and the evaluation. When spectral distance is used, the vertical deviations are relatively small, and it becomes easy to set up the degree of speech privacy, compared to when the articulation index is used. Fig. 4 also shows that when spectral distance is almost 3 dB, evaluation of speech privacy was 2, which means "it was possible to recognize the sound as a voice, though the contents of the speech were not understood".

4.2. Relationships between indices and listening score

The relationship between the evaluation index and listening score were obtained from Experiment I(2). The relationship between the articulation index and the listening score is shown in Fig. 5. As the value of the articulation index approaches 0, a minute change in the value of the articulation index brings a big change in the listening score. Adjustments of the listening score are comparatively difficult using the articulation index. The result utilizing spectral distance is shown in Fig. 6. The listening score does not change sharply as the spectral distance value becomes smaller. Adjustments of the listening score are comparatively easy utilizing spectral distance. From Figs. 4 and 6, when the listening score drops to 60% or less, speech privacy evaluation becomes 2 or less.

4.3. Selection of most useful index

AIC [8] is introduced as an evaluation standard for selecting the most effective index. AIC values for each of the cases in Figs. 3–6 are shown in Table 1. The AIC value in the relationship between the spectral distance and speech privacy evaluation (Fig. 4), is smaller than in the relationship between the articulation index and speech privacy evaluation (Fig. 3). Further, a similar trend is found even with the listening score (Figs. 5 and 6). From the above results, it can be seen that it is logical to select spectral distance as the target index.

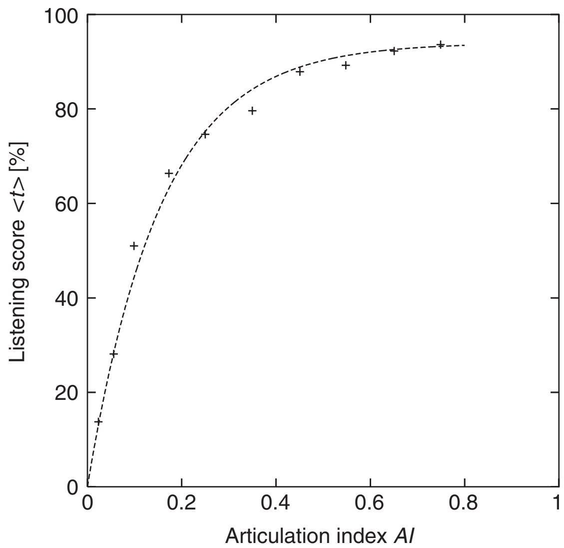

Fig. 5. Relationships between articulation index and listening score.

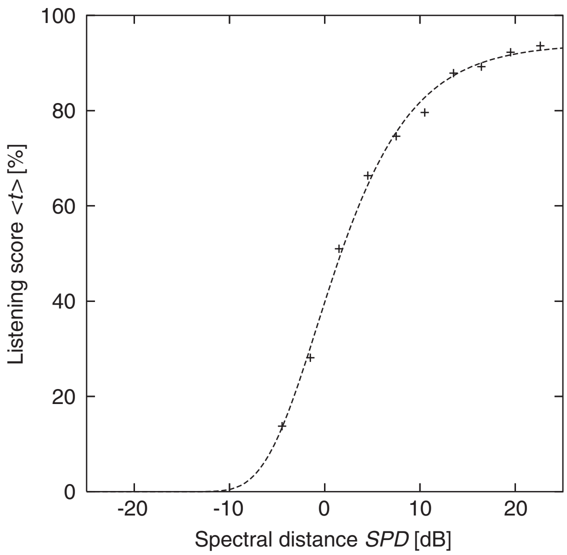

Fig. 6. Relationships between spectral distance and listening score.

Table 1
Values of AIC

Index	Speech privacy	Listening score
AI	−29.79	21.68
SPD	−31.14	13.54

5. Outline of Experiment II

To examine what psychological evaluation of speech privacy was obtained when a masking noise corresponding to some degree of the evaluation was transmitted, Experiment II was conducted. In outline, Experiment II was as follows.

5.1. Location

The same location as in Experiment I.

5.2. Subjects

The subjects were seven male and one female students with normal hearing, and were different from the subjects in Experiment I.

5.3. Presented sound

(a) *Audio signal*: The same audio signal as presented in Experiment I.
(b) *Masking noise*: The same masking noise as presented Experiment I. From the relationship between spectral distance and evaluation (Fig. 4), the sound pressure level of the masking noise was set up so that the spectral distance could be set to −3, 0, 3, 6, 9, 12 dB which corresponded to the cases of evaluations of speech privacy of 1.0, 1.5, 2.0, 2.5, 3.0, 3.5.

5.4. Measurement method

The same method as in Experiment I.

6. Comparison between predicted and observed values

Predicted values of evaluations of speech privacy from Fig. 4 were compared with the observed values in Experiment II. The results are shown in Fig. 7. Fig. 8 shows the comparison between predicted and observed values of listening score. A high level consistency is seen between the predicted and observed values.

7. Conclusions

In this study, we focused on achieving speech privacy using a meaningless steady masking noise, a method for setting up masking noise was considered based on the results of psychological experiments. From the

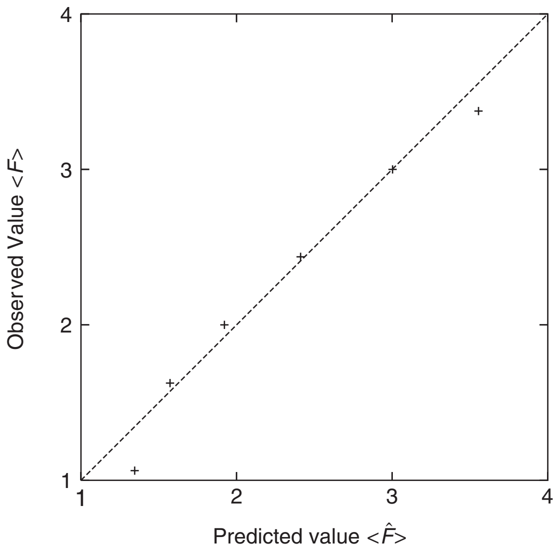

Fig. 7. Comparison between predicted and observed values of speech privacy.

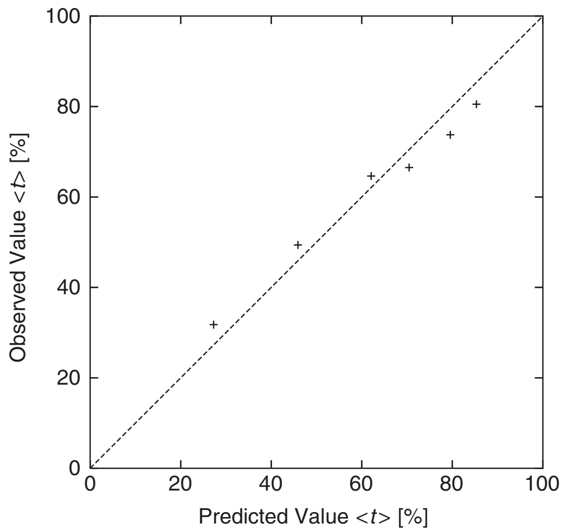

Fig. 8. Comparison between predicted and observed values of listening score.

results of Experiment I, the spectral distance was shown to be more effective than the articulation index in setting up the degree of speech privacy. Speech privacy can be protected in the range of listening scores of about 60% or less, and in the range of spectral distances of about 3 dB or less. From the results of Experiment II, setting up the degree of speech privacy by spectral distance was shown to be effective because of the good agreement between predicted and observed values for speech privacy evaluation.

Acknowledgments

The authors would like to express their cordial thanks to Shin'ya Fujinaka and Takuya Yamane for their helpful assistance and to their colleagues who provided valuable comments at the 2003 Fall Meeting of Acoustical Society of Japan [9], the Technical Group on Engineering Acoustics of the Institute of Electronics Information and Communication Engineers [10], and the 18th International Congress on Acoustics 2004 [11]. This study was supported in part by the Japan Society for the Promotion of Science, Grant-in-Aid for Scientific Research (C), no.15500138.

References

[1] W.J. Cavanauch, W.R. Farrell, Speech privacy in buildings, *Journal of the Acoustical Society of America* 34 (4) (1962) 475–492.
[2] Y. Tokita, *Sound Environment and Control Technology* II, Fuji Techno-System Co., Tokyo, 1999 (in Japanese).
[3] T. Saeki, T. Tamesue, S. Yamaguchi, K. Sunada, Selection of meaningless noise for masking of speech, *Applied Acoustics* 65 (29) (2004) 203–210.
[4] K.D. Kryter, Methods for the calculation and use of the articulation index, *Journal of the Acoustical Society of America* 34 (11) (1962) 1689–1697.
[5] T. Tamesue, S. Yamaguchi, T. Saeki, Psychological impressions and listening score when listening to audio signals composed of monosyllables and words, while subject to meaningless steady noise—introduction of weighted-mean spectral distance, *Journal of Sound and Vibration* 271 (3–5) (2004) 519–533.
[6] K. Yonemoto, Characteristics of CD for the evaluation of fitting condition with hearing aids (TY-89), *Journal of Otolaryngology, Head and Neck Surgery* 11 (9) (1995) 1395–1401 (in Japanese).
[7] J.P. Guilford, *Psychometric Method*, McGraw-Hill, New York, 1954.
[8] H. Akaike, A new look at the statistical model identification, *IEEE Transaction on Automatic Control* AC-19 (6) (1974) 716–723.
[9] T. Saeki, S. Yamaguchi, T. Tamesue, Study on evaluation indices of speech privacy, *Proceedings of the 2003 Fall Meeting of Acoustical Society of Japan*, 2003, pp. 717–718 (in Japanese).

[10] T. Saeki, S. Yamaguchi, T. Tamesue, Study on achieving speech privacy using masking noise, *Technical Report of Institute of Electronics Information and Communication Engineers* EA-2003-93, 2003, pp. 43–48 (in Japanese).
[11] T. Saeki, T. Tamesue, S. Yamaguchi, Evaluation indices for speech privacy, *Proceedings of the 18th International Congress on Acoustics*, vol. 4, 2004, pp. 3191–3194.