# Subjective evaluation of speech privacy at consulting rooms in hospitals: Relationship between feeling evoked by overhearing speech and word intelligibility score

Sato, Hayato

Morimoto, Masayuki

Ohtani, Souichi

Hoshinsa, Yasushi

Sato, Hiroshi

# Subjective evaluation of speech privacy at consulting rooms in hospitals: Relationship between feeling evoked by overhearing speech and word intelligibility score

Hayato Sato[a,*], Masayuki Morimoto[a], Souichi Ohtani[a], Yasushi Hoshino[a,b], Hiroshi Sato[c]

[a]*Environmental Acoustics Laboratory, Department of Architecture, Graduate School of Engineering, Kobe University, Rokko, Nada, Kobe 657-8501, Japan*
[b]*Nippon Sheet Glass Environment Amenity Co., Ltd., Takanawa, Minato, Tokyo 108-0074, Japan*
[c]*Human Informatics Research Institute, National Institute of Advanced Industrial Science and Technology, Higashi, Tsukuba 305-8566, Japan*

## Abstract

Many previous studies and related standards on speech privacy have adopted speech intelligibility as a subjective measure under the assumption of a strong correlation between them. Meanwhile, some studies have attempted to directly evaluate speech privacy. However, the methods used in these studies required evaluators to clearly understand what speech privacy is, while the concept of speech privacy is not widely known to ordinary people using public spaces. In the present study, questionnaire-based surveys and a listening test were conducted to clarify the relationship between speech intelligibility and speech privacy using a more suitable method for directly evaluating speech privacy. The respondents and participants were instructed to imagine a specific situation at a consulting room in a hospital because the required speech privacy depends on the situation. The questionnaire before the listening test showed that a feeling of

*Corresponding author. Tel./fax: +81 78 803 6052.
*Email address:* hayato@kobe-u.ac.jp (Hayato Sato)

dissatisfaction was the strongest feeling evoked by overhearing speech. Therefore, we used this feeling as a term for evaluating speech privacy performance that most people can easily accept and understand. The threshold of the feeling of dissatisfaction obtained from the listening test was compared with word intelligibility scores from our previous study. The comparison showed that the threshold of the feeling of dissatisfaction corresponds to a word intelligibility score of around 10 to 20% for participants with a typical sensitivity to personal information leakage through conversation. In addition, the ratio of participants who felt dissatisfaction increased with increasing background noise level under equal-intelligibility conditions.

## 1. Introduction

It has become necessary to prevent the leakage of personal information through conversation, that is, to ensure speech privacy in public spaces such as hospitals, banks, and offices. One of the important issues of speech privacy is an appropriate subjective measure for the evaluation of speech privacy performance.

Alliance for Telecommunications Industry Solutions (ATIS) defines speech privacy as *techniques using fixed sequence permutations or voice/speech inversion to render speech unintelligible to the casual listener*[1]. This definition clearly relates speech privacy to speech intelligibility. Standards for speech privacy[2, 3] adopt the Articulation Index[4] or the Speech Intelligibility Index[5], which were developed as objective measures for speech intelligibility.

Speech intelligibility is useful as a subjective measure for evaluating speech

privacy since it has high reproducibility when the same test speeches are used. In our previous study[6], we demonstrated a chart by which speech intelligibility can be predicted from the speech level, the background noise level, and the sound insulation performance of a wall. Many papers[7–12] have similarly adopted speech intelligibility as an indirect subjective measure for speech privacy under the assumption of a strong correlation between them.

In some papers, the speech privacy desired for various spaces was directly evaluated as measures different from speech intelligibility[7, 13–15]. Cavanaugh *et al*.[7] defined two types of speech privacy in offices: normal privacy and confidential privacy. Their definition has been a cornerstone of speech privacy studies and is used even today. The former is related to a bothersome feeling in the workplace, and the latter, which is relevant to the present study, is related to the leakage of confidential conversation. Cavanaugh *et al*. used the adjective "*proper*" to instruct participants of laboratory tests which degrees of the leakage to be adjusted. Bradley and Gover[13] directly asked participants to rate the perceived speech privacy in open offices using the query *How would you rate the SPEECH PRIVACY?* on a five category scale of "*None*", "*A little*", "*Acceptable*", "*Moderately Good*", and "*Confidential*". Pop and Rindel[14] investigated the perceived speech privacy in open offices using a computer auralization system. They followed Bradley and Gover's method in their evaluation of the perceived speech privacy.

These methods were based on the premise that participants can clearly understand what speech privacy is. However, the concept of speech privacy is not widely known to ordinary people using public spaces. Furthermore, the definition of speech privacy has not been unified. ATIS narrowly defines speech privacy as

a "*technique*", while the term "*privacy*" is ambiguous: it is a kind of human right of being able to keep personal affairs secret, the freedom to be left alone, the state of being free from unwanted or undue intrusion or disturbance, and so forth. This means that the concept should be carefully explained to participants of subjective tests using expressions that are easy to understand.

It is noteworthy that Cavanaugh *et al*.[7] and Young[16] used a five category scale of "*dissatisfaction*" ("*Apparent satisfaction*", "*Mild dissatisfaction*", "*Moderate dissatisfaction*", "*Strong dissatisfaction*", and "*Extreme dissatisfaction*") to rate the overall degree of speech privacy in field surveys. Using such a general term instead of "*speech privacy*" as an evaluation term may be a reasonable method, although the reason why they used it was not explained in their papers.

Meanwhile, Lee *et al*.[15] evaluated the degree of personal information protection using a method of categorical rating and reported that it has a strong correlation with sentence intelligibility. However, this conclusion is rather self-evident. Because in their subjective test, the participants were asked *Do you think how much personal information of a patient being examined in the next consulting room is protected?*, the participants' responses to this question are expected to have depended on the quantity that they were able to hear, i.e., the speech intelligibility. Therefore, it is still unclear whether or not speech intelligibility and speech privacy are interchangeable.

The purpose of the present paper is to investigate the relationship between speech intelligibility and the subjective evaluation of speech privacy. We pay attention to the feelings evoked by overhearing speech to enable an intuitive and subjective evaluation of speech privacy. The term expressing the feelings evoked

is used as an evaluation term for speech privacy performance that most people can easily accept and understand. In addition, the subject of this study is focused on speech privacy at a room adjacent to a consulting room in a hospital, where the high protection of speech privacy is required, because the feeling depends strongly on the usage of space.

The present paper is divided into three studies. Firstly, we performed a questionnaire-based survey to investigate the feelings evoked by overhearing speech. The respondents were asked to suppose that they were sitting in a waiting room adjacent to a consulting room in a hospital. Secondly, we performed a subjective listening test on speech privacy, focusing on the feeling reported in the questionnaire. Thirdly, we investigated the relationship between the feeling and the word intelligibility obtained from our previous study[6].

## 2. Questionnaire-based survey on feelings evoked by overhearing speech

### 2.1. Methods

A questionnaire on feelings evoked by overhearing speech was given to 245 respondents. The respondents were volunteers and they were university students, their families, university clerical staff, and office workers of private companies. The ages of the respondents ranged from teenage to in their seventies.

The questionnaire was divided into two parts. The first half was related to the attributes of the respondents. The items were as follows: gender, age, occupation, frequency of hospital visits, experience of overhearing speech in hospitals, and sensitivity to personal information leakage through conversation. Each item was a forced-choice question with some categories.

In the second half, a probable situation in a consulting room was presented

to the respondents using the drawing in Fig. 1 and the following sentences were visually presented to them; *You are sitting in a waiting room adjacent to a consulting room. You can hear the conversation between the doctor and the patient from the consulting room.* Seven possible feelings evoked by the situation were given to the respondents. Also, a sentence to explain each feeling was also presented to the respondents to avoid different interpretations of these feelings. Table 1 lists the seven feelings and their explanations. These feelings were selected from the section on "feelings" in a Japanese thesaurus[17]. The respondents gave the strength of each feeling as a score from 1 to 7 by circling a number.

[Figure 1 about here.]

[Table 1 about here.]

## 2.2. Results and discussion

A total of 42 KruskalWallis rank sum tests were conducted using R ver. 3.3.0[18] to clarify the effect of each attribute of the respondents on the strength of each feeling. The significance level was set at $p < 0.05$. Table 2 shows $p$-values obtained for each combination of the attribute and the feeling. The $p$-values for gender (two levels), age (seven levels), occupation (four levels), and sensitivity to personal information leakage (four levels, from "*do not mind at all*" to "*extremely concerned*") were significant for most of the feelings, while those for the frequency of hospital visits and the experience of overhearing speech in hospitals were not significant for any of the feelings. Note that the $p$-values for sensitivity were significant for all feelings and much smaller than those for the

6

other attributes. This means that the strength of each feeling was strongly affected by the sensitivity relative to the other attributes.

[Table 2 about here.]

Figure 2 shows relative frequency distributions of the strength of each feeling classified into the four levels of sensitivity. A larger circle represents a higher relative frequency. The sample number for each level (N) is also shown. The sum of N is 244 because one respondent did not answer the sensitivity. The distributions shifted upward with increasing sensitivity, regardless of the feeling. The distributions for unease and dissatisfaction seemed to be higher than the other feelings at all levels of sensitivity. For the lowest sensitivity, curiosity showed relatively high strength.

[Figure 2 about here.]

The differences among the strength of each feeling were statistically tested using the Wilcoxon signed-rank test at each level of sensitivity. In accordance with the Bonferroni correction, the significance level was set at $p < 0.00238 (= 0.05 / 21)$. Table 3 shows $p$-values for the pairs among dissatisfaction and the other feelings. With the exception of all feelings at the lowest sensitivity and unease at the highest sensitivity, the distribution of dissatisfaction was significantly different from those of the other feelings. Furthermore, the $p$-values for unpleasantness, restlessness, embarrassment, and anger at the lowest sensitivity were reasonably small but not statistically significant.

[Table 3 about here.]

7

In summary, dissatisfaction was the strongest feeling evoked by overhearing speech in a consulting room in a hospital, for all types of people in the survey. In other words, the feeling of dissatisfaction can be used as an evaluation term for speech privacy performance that most people can easily accept and understand. This is consistent with that Cavanaugh *et al.*[7] and Young[16] used "dissatisfaction" to rate the overall degree of speech privacy as a result.

## 3. Subjective listening test on speech privacy

In a previous study[6], we obtained word intelligibility scores in sound fields that simulated the situation shown in Fig. 1. In this listening test, the threshold of the feeling of dissatisfaction was measured for the sound fields used in Ref. [6].

### 3.1. Methods

### 3.1.1. Speech stimuli

A speech material simulating the conversation between a doctor and a patient with a diarrheal disease seeing the doctor for the first time was recorded for the listening test. The script for the conversation was written on the basis of a practical training session on medical interviews between a medical school student and a simulated patient in cooperation with Graduate School of Medicine, Kobe University. The recording was performed in an anechoic room using a portable PCM recorder (SONY, PCM-D1). The script was read by a male and a female, both university students, with experience of theatrical performance. The length of the speech material was 272 s.

The recorded speech material had several time frames where the sound pressure level (SPL) was higher than elsewhere. Furthermore, there was a

difference in SPL between the male and female speakers. Therefore, the equivalent continuous A-weighted sound pressure level ($L_{\text{Aeq}}$) for each frame with a length of 1 s and a shift of 200 ms was equalized so that $L_{\text{Aeq}}$ for all frames was equal. As a result, there was no unnaturalness in the listening impression due to this process in the opinion of the authors. Figure 3 shows $L_{\text{Aeq}}$ for the processed speech material in each frame with 200 ms duration.

[Figure 3 about here.]

Two different frequency characteristics of the speech stimulus were used in the present study. One was the material after it had passed through a filter (graphic equalizer; Apple, Logic Pro) with a frequency characteristic of $-5$ dB per octave. The filter was used to simulate speech sound transmitted through a boundary wall. In other words, the frequency characteristic of the filter means that the level difference between the two rooms increased by 5 dB per octave with increasing frequency, and this characteristic was used in Ref. [6]. In the following sections, this characteristic used to simulate sound transmission through a wall is denoted as $Tr$. The other characteristic was the speech material without any filter. This characteristic corresponding to a situation without the boundary wall, is denoted as $Flat$ in the following sections.

### 3.1.2. Background noise

The same noise as that used in Ref. [6] was added as a background noise. A steady-state random noise with $-5$ dB per octave decay in the frequency domain was added to speech stimuli to simulate a general room noise.

### 3.1.3. Presentation of test stimuli

Listening tests were performed in an anechoic room. Figure 4 shows a block diagram of the test equipment and the loudspeaker arrangement used in the listening test. The loudspeakers were placed in a 1.5 m circumference with the center at the center of the participant's head and at a height of 1.2 m above the floor. The frequency characteristics of the loudspeakers were adjusted to be flat within ±3 dB in the range from 100 Hz to 10 kHz using graphic equalizers built in a digital mixer.

[Figure 4 about here.]

The presentation method for the speech stimuli and background noise was the same as in Ref. [6]. The speech stimuli were presented from the loudspeaker in front (0 deg.) of the participant. The speech level was measured at the position corresponding to the center of the participant's head in the absence of the participant.

The background noise was presented from the five loudspeakers shown in Fig. 4. Five different random noises were presented from five different loudspeakers. They were uncorrelated with each other but had the same frequency characteristic. The SPL of the noise presented from the front loudspeaker was set at +3 dB relative to those of the other four loudspeakers so that the frequency characteristic of the degree of interaural cross-correlation (ICC) of the background noise was close to the theoretical value of the correlation coefficient between two different points (distance: 0.3 m) in a diffuse sound field[19]. Our unpublished measurements in a hospital after the close of the outpatient service demonstrated that the frequency characteristics of ICC for environmental noise

in the consulting rooms were close to the theoretical value. As a result, a sound image of background noise was perceived broadly and it might not be localized at a particular direction.

$L_{Aeq}$ for the background noise was measured at the same position as for the speech stimuli. The background noise level was set at 30, 40, and 50 dB referring to that the background noise levels at consulting rooms in hospitals measured by Sato *et al.*[20], which ranged from 33.8 to 41.5 dB and the maximum allowable noise level at offices by Warnock[9] of 48 dB. The background noise of X dB is abbreviated to NX hereafter. For example, N40 represents a background noise of 40 dB.

### 3.1.4. Instruction and procedure

Both the threshold of the feeling of dissatisfaction and the masked threshold of the speech stimulus were measured by the method of adjustment. For the threshold, the participants were instructed to adjust the SPL of the speech stimulus to the maximum sound volume without any feeling of dissatisfaction in the situation shown in Fig. 1 by rotating a dial displayed on a touch screen. For the masked threshold, the participants were instructed to adjust the SPL of the speech stimulus to the minimum audible sound volume. The dial was related to the attenuation factor of the speech stimulus, and the factor was recorded throughout each adjustment.

The result of the questionnaire-based survey in section 2 indicated that the sensitivity to personal information leakage through conversation may be an important factor in evaluating the feeling of dissatisfaction. Therefore, not only the instruction regarding the situation shown in Fig. 1 but also the following instruction was presented to the participants to try to raise their sensitivity: *Please*

*imagine that you have a serious disease which you want to keep confidential and that the leakage of speech you hear will also occur when you are consulting your doctor.*

The test parameters were the frequency characteristics of the speech stimuli (*Tr* and *Flat*) and the background noise level (N30, N40, and N50). Six test conditions (two frequency characteristics × three background noise levels) were tested in the present study. Each participant made five adjustments for each test condition and each threshold, in total performing 60 adjustments.

The order of the test conditions was random. The SPL at the beginning of the presentation and the start position of the speech stimulus were different in the five adjustments for each test condition. The SPL was randomly set within one of five speech-to-noise ratio (SNR) ranges of $-25$ to $-15$, $-15$ to $-10$, $-10$ to $-5$, $-5$ to $+5$, and $+5$ to $+15$ dB. The speech stimulus started at one of six positions of 0, 40, 80, 120, 160, and 200 s from the beginning of the 272 s long speech material.

The dial can adjust the SPL of the speech stimulus within an SNR from $-35$ to $+25$ dB. The speech stimulus was repeatedly presented if it ended before the adjustment was completed. There was no time restriction for the adjustment.

The listening test was divided into six sessions, in each of which 10 adjustments were made. The threshold of the feeling of dissatisfaction was measured in the first three sessions, and the masked threshold was measured in the remaining sessions. Before the listening test, each participant completed a pilot trial consisting of four adjustments.

### 3.1.5. Questionnaire-based survey after listening tests

A questionnaire-based survey was conducted after finishing all the sessions because the participants' reports after the listening tests suggested that there

was difference among their sensitivities to personal information leakage through conversation. The purpose was to classify the participants on the basis of their sensitivities.

The questionnaire was divided into two parts. The first half was related to the attributes of the participants and was the same as that in the questionnaire in section 2. In the second half, the participants stated which of the four boundaries regarding the degree of speech understanding shown in Fig. 5 was closest to the threshold of dissatisfaction they had indicated. The boundaries correspond to participants' sensitivity to personal information leakage through conversation.

[Figure 5 about here.]

### 3.1.6. Participants

Fifty-one people participated in the listening test. The participants were volunteers with a normal hearing level. They were university clerical staff and office workers of private companies. The ages of participants ranged from their twenties to their sixties. Most of the participants were also the respondents of the questionnaire survey in section 2. Note that the experience of answering the questionnaire was not expected to affect the results of the listening test for the following three reasons: (1) The listening test was performed about four months after the questionnaire survey. (2) The situation was described in detail to the subjects before the listening test and the instructions were the same as those for the questionnaire survey. (3) The participants were classified according to their sensitivity to personal information leakage through conversation, which strongly affects the feeling of dissatisfaction.

## 3.2. Results and discussion

### 3.2.1. Boundaries of degree of speech understanding

Figure 6 shows the number and percentage of participants indicating each boundary of the degree of speech understanding. Boundary 3, the upper limit of the category "*The conversation is audible, but not even words in the conversation are intelligibile*," has the largest value of 69%. This means that the participants' sensitivities to personal information leakage in the listening test were well controlled. Furthermore, even though the sensitivity was intentionally raised in the present study, it is clear that more than 80% of participants did not require that "*The conversation is completely inaudible*," i.e., perfect speech security[21].

[Figure 6 about here.]

### 3.2.2. Threshold of feeling of dissatisfaction

The threshold of the feeling of dissatisfaction and the masked threshold were obtained by averaging the five adjustments for each participant and each test condition. Figure 7 shows the quartiles and maximum/minimum values for the threshold of the feeling of dissatisfaction and the masked threshold for each test condition. Panels (a)–(d) indicate the results for the groups of participants classified into boundaries 1–4, respectively. Open and shaded boxes indicate the threshold of the feeling of dissatisfaction and the masked threshold, respectively.

[Figure 7 about here.]

Both the threshold of the feeling of dissatisfaction and the masked threshold show almost the same tendency for each of the boundaries, increasing with an increase in the background noise level. Regarding the frequency characteristics

14

of the speech stimulus, *Flat* has lower thresholds than *Tr*. The threshold of the feeling of dissatisfaction decreases as the boundary becomes more sensitive, but the masked threshold is not affected by the boundary. For the most sensitive boundary (boundary 4), it is reasonable that the threshold of the feeling of dissatisfaction and the masked threshold are almost identical. The interquartile range of the threshold of the feeling of dissatisfaction for boundary 2 is clearly larger than those for the other boundaries. This suggests that the judgement for boundary 2 is more difficult than for the other boundaries.

The effects of the boundary (four levels) and the stimulus (six levels = two frequency characteristics of speech stimulus × three background noise levels) on the threshold of the feeling of dissatisfaction and the masked threshold were analyzed statistically. Tables 4 and 5 show the *p*-values obtained from the statistical tests.

[Table 4 about here.]

[Table 5 about here.]

First, their effects on the masked threshold were analyzed. Multiple comparisons using the Wilcoxon rank sum test were performed to determine differences between the medians of the masked threshold over all test conditions for each boundary. In accordance with the Bonferroni correction, the significance level was set at $p < 0.00833 (= 0.05 / 6)$. The *p*-values are shown in Table 4 (*MT*). The results showed no significant difference among all the boundaries. However, the median values were 23.6, 23.9, 19.5, and 20.4 dB for boundaries 1–4, respectively, and the median values of boundaries 1 and 2 were 3–4 dB higher than those of boundaries 3 and 4. Furthermore, the results of the

15

multiple comparisons showed marginal significances between boundaries 1 and 3 ($p = 0.0108$) and between boundaries 2 and 3 ($p = 0.0150$). Thus, it cannot be concluded that the masked thresholds of the four boundaries were equal.

For the medians of the masked threshold over all boundaries for each test condition, multiple comparisons using the Wilcoxon signed-rank test were performed. In accordance with the Bonferroni correction, the significance level was set at $p < 0.00333$ ($= 0.05 / 15$). The $p$-values are shown in Table 5 (*MT*). They ranged from $10^{-10}$ to $10^{-8}$, and were clearly less than the significance level. This means that the frequency characteristic of the speech stimulus and the background noise level affect the masked threshold.

Next, the threshold of the feeling of dissatisfaction was investigated. The threshold of the feeling of dissatisfaction was normalized by the masked threshold because it may have been affected by differences in the masked threshold, even though there were no significant differences between the masked thresholds. This normalization corresponds to the threshold of the feeling of dissatisfaction expressed as a sensation level. For each participant, the threshold of the feeling of dissatisfaction under each test condition was normalized by the masked threshold for the same test condition. Figure 8 shows the threshold of the feeling of dissatisfaction expressed as a sensation level as functions of the background noise level and the frequency characteristic of the speech stimulus.

[Figure 8 about here.]

A multiple comparison for the effect of the boundary was performed in the same manner as for the masked threshold. The $p$-values shown in Table 4 ($DS_N$) indicated significant differences among all the boundaries except

16

between boundaries 1 and 2. The median values of the threshold of the feeling of dissatisfaction expressed as a sensation level were 13.7, 10.8, 6.3, and 0.8 dB for boundaries 1–4, respectively. The threshold of the feeling of dissatisfaction decreases as the boundary becomes more sensitive. A multiple comparison for the effect of the test condition was also performed. The $p$-values shown in Table 5 ($DS_N$) indicated that only the thresholds for Tr at $Tr$ at N40 and $Tr$ at N50 were significantly different from those for other test conditions. As further evidence of the effect of the test condition, the results of multiple comparisons for the threshold of the feeling of dissatisfaction without normalization by the masked threshold showed a significant difference among all the stimuli (see Table 5 ($DS$)). Therefore, it can be concluded that the feeling of dissatisfaction is higher for Flat than for Tr when the background noise level is identical. Additionally, the threshold of the feeling of dissatisfaction increases as the background noise level increases. Combined with the fact that the thresholds of the feeling of dissatisfaction expressed as a sensation level showed small differences among the test conditions, most of the difference in the threshold of the feeling of dissatisfaction among the test conditions can be explained by the difference in the masked threshold.

If the threshold of the feeling of dissatisfaction can be exactly expressed as a sensation level, it is sufficient to consider only the boundary of speech understanding for the acoustical design of a space. However, it is difficult to adopt the sensation level for actual acoustical design. Therefore, taking its practical use into consideration, we examined whether the threshold of the feeling of dissatisfaction can be expressed by signal-to-noise ratio (SNR), which was obtained as the difference between the threshold and the background noise level

17

in Fig. 7. Figure 9 shows the result. Except for the most sensitive boundary (boundary 4), SNR is almost always higher than −15 dB at the threshold. This means that maintaining SNR at −15 dB can suppress the feeling of dissatisfaction to under the threshold for about 80% of people, who belonged to boundaries 1, 2, and 3.

[Figure 9 about here.]

In summary, the threshold of the feeling of dissatisfaction strongly depends on the participants' sensitivities to personal information leakage when the threshold is expressed as a sensation level. Moreover, most parts of the difference in the threshold between stimuli can be explained by the difference in the masked threshold. The masked threshold depends on the background noise level and the frequency characteristics of the speech stimulus. Maintaining SNR at -15 dB can suppress the feeling of dissatisfaction to under the threshold for about 80% of people, who belonged to boundaries 1, 2, and 3.

## 4. Discussion: Relationship between feeling of dissatisfaction and speech intelligibility

The purpose of this section is to clarify the relationship between the feeling of dissatisfaction obtained in this paper and word intelligibility obtained in Ref. [6].

We set the thresholds of the feeling of dissatisfaction for the participants belonging to boundary 3 as a practical target of analysis. Participants with sensitivity corresponding to boundaries 1 and 2 will feel lower dissatisfaction if the sound insulation performance is designed on the basis of boundary 3. The thresholds of the feeling of dissatisfaction for boundary 4 were almost

equal to the masked threshold. This means that the threshold of the feeling of dissatisfaction is the highest target value for protecting speech privacy. Although such performance might be required in some cases, the ratio of the participants belonging to boundary 4 was low.

To compare the feeling of dissatisfaction with word intelligibility using the same unit, cumulative relative frequency distributions of the thresholds were calculated. The thresholds of the feeling of dissatisfaction for each participant were grouped in 1 dB steps for each test condition, and the frequencies for each of the groups were counted. Figure 10 shows the relative cumulative frequency distribution for $Tr$, which is the condition used in Ref. [6]. Logistic regression curves are also shown in Fig. 10, which were obtained using the data corresponding to relative cumulative frequencies of 10% to 90%. Their fitting was good, and the estimation errors averaged over the data were within 3.6 percentage points. The y-axis indicates the ratio of participants who feel dissatisfaction, denoted as %-Dissatisfaction hereafter.

It can be considered that Boundary 3 in Fig. 5 corresponds to the threshold of *cadence* measured by Gover and Bradley[10]. Gover and Bradley reported that the threshold of *cadence* in the speech-to-noise ratio was $-12.7$ dB. From Fig. 10, 50% of the participants felt dissatisfaction when the speech-to-noise ratio was $-10.9$, $-12.7$, and $-13.5$ dB for N30, N40, and N50, respectively. These speech-to-noise ratios are roughly consistent with that of Gover and Bradley; however, note that the ratio decreases with increasing background noise level in the present study. Meanwhile, the degree of personal information protection defined by Lee *et al.*[15], which is the ratio of the sum of the responses *Secure* and *Very secure*, exceeded 50% when the speech-to-noise ratio was less than

−10 dB. Young[16] suggested that a sound excess of 0 dB, which corresponds to a speech-to-noise ratio of −15 dB, is required to achieve apparent satisfaction in the case of considering confidential speech privacy. The result of the present study is in reasonable agreement with these previous studies on subjective speech privacy ratings.

[Figure 10 about here.]

Table 6 shows the word intelligibility scores for $Tr$ obtained in Ref. [6]. The test words were selected from familiarity-controlled word lists[22] to ensure their high familiarity with both young and elderly people.

Figure 11 shows the relationship between %-Dissatisfaction and the word intelligibility score. The values of %-Dissatisfaction were derived from the regression curves shown in Fig. 10 and the values of $L_S$ shown in Table 6. Roughly speaking, %-Dissatisfaction of 50%, in other words, the threshold of the feeling of dissatisfaction, corresponds to a word intelligibility score of around 10 to 20%. A score in this range can be used as an acceptable limit value for speech privacy in the situation shown in Fig. 1.

Meanwhile, if %-Dissatisfaction and intelligibility were perfectly interchangeable, all of the plots in Fig. 11 would have been located on a single curve. However, the plots for N50 shifted upward relative to those for N30 and N40. This indicates that %-Dissatisfaction depends on not only intelligibility but also other factors.

The upper limits of the background or masking noise level suggested in previous studies were 48 dB by Warnock[9] and 45 dB by Bradley[21]. Therefore, the loudness of background noise might have increased the feeling

of dissatisfaction under the test condition of N50. As another possibility, the gap between the ideal sound insulation performance envisaged by the participants and that presumed from the stimuli is expected to generate a feeling of dissatisfaction. The requirement of a high background noise level to decrease the intelligibility or loudness of speech implies a poor sound insulation performance. Such thinking can also explain the results shown in Fig. 11. The effect of such a gap may have been blurred or canceled by the assumed situation and participants' knowledge or experience, despite these factors being controlled as carefully as possible in the present study. Although Cavanaugh *et al*.[7] and Young[16] used "*dissatisfaction*" to rate the overall degree of speech privacy, they did not report these effects in relation to the background noise level.

[Table 6 about here.]

[Figure 11 about here.]

## 5. Conclusions

Questionnaire-based surveys and a listening test were conducted to clarify the relationship between speech intelligibility and speech privacy. The respondents and participants were instructed to imagine that they are sitting in a room adjacent to a consulting room in a hospital because the required speech privacy depends on the situation. In the listening test, furthermore, the participants' sensitivity to personal information leakage through conversation was intentionally raised by instructing them to imagine that they have a serious disease that they wished to remain confidential.

The questionnaire before the listening test showed that

21

(1) the feeling of dissatisfaction was the strongest feeling evoked by overhearing speech, which was common to people with different backgrounds.

Therefore, we used this feeling as an evaluation term for speech privacy performance that most people can easily accept and understand.

The listening test on the feeling of dissatisfaction and the questionnaire after the test showed that

(2) about 80% people did not require that "*The conversation is completely inaudible*," as the speech privacy performance of the consulting room;

(3) the threshold of the feeling of dissatisfaction strongly depended on the participants' sensitivities to personal information leakage when the threshold was expressed as a sensation level;

(4) maintaining the speech-to-noise ratio at $-15$ dB can suppress the feeling of dissatisfaction to under the threshold for about 80% of people.

Finally, the threshold of the feeling of dissatisfaction was compared with word intelligibility scores obtained in our previous study[6]. The comparison showed that

(5) roughly speaking, the threshold of the feeling of dissatisfaction corresponds to a word intelligibility score of around 10 to 20% for participants with a typical sensitivity;

(6) the ratio of participants who felt dissatisfaction increased with increasing background noise level under equal-intelligibility conditions.

## References

[1] ATIS-0100523.2011, 2011.

[2] ASTM E1130-08, Standard test method for objective measurement of speech privacy in open plan spaces using articulation index, 2008.

[3] ISO 3382-3, Acoustics   Measurement of room acoustic parameters   Part 3: Open plan offices, 2012.

[4] ANSI S3.5-1969, American National Standard Methods for the Calculation of the Articulation Index, 1969.

[5] IEC 60268-16 Ed. 4.0, Objective rating of speech intelligibility by speech transmission index, 2011.

[6] H. Sato, M. Morimoto, Y. Hoshino, Y. Odagawa,  Relationship between sound insulation performance of walls and word intelligibility scores, Appl. Acoust. 73 (2012) 43–49.

[7] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, B. G. Watters, Speech privacy in buildings, J. Acoust. Soc. Am. 34 (1962) 475–492.

[8] R. Pirn,  Acoustical variables in open-planning,  J. Acoust. Soc. Am. 49 (1971) 1339–1345.

23

[9] A. C. C. Warnock, Acoustical privacy in the landscaped office, J. Acoust. Soc. Am. 53 (1973) 1535–1543.

[10] B. N. Gover, J. S. Bradley, Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms, J. Acoust. Soc. Am. 116 (2004) 3480–3490.

[11] V. Hongisto, A model predicting the effect of speech of varying intelligibility on work performance, Indoor Air 15 (2005) 458–468.

[12] J. S. Bradley, M. Apfel, B. N. Gover, Some spatial and temporal effects on the speech privacy of meeting rooms, J. Acoust. Soc. Am. 125 (2009) 3038–3051.

[13] J. S. Bradley, B. N. Gover, Describing levels of speech privacy in open-plan offices, NRC Canada Research report IRC-RR-138 (2003).

[14] C. B. Pop, J. H. Rindel, Perceived speech privacy in computer simulated open-plan offices, Proc. Inter-Noise (2005) 7 pages.

[15] H. Lee, K. Ueno, S. Sakamoto, M. Fujiwara, Y. Shimizu, Experimental study on evaluation method of speech privacy, Proc. Inter-Noise (2011) 6 pages.

[16] R. Young, Re-vision of the speech-privacy calculation, J. Acoust. Soc. Am. 38 (1965) 524–530.

[17] T. Yamaguchi (Ed.), Nihongo dai shisourasu (in Japanese), Taishukan, Tokyo, Japan, 2003.

[18] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016.

[19] H. Kuttruff, Room Acoustics (3rd ed.), Elsevier, London, UK, 1991.

[20] H. Sato, M. Morimoto, A case study on speech communication environment in a hospital (in Japanese), Summaries of technical papers of Annual Meeting of Architectural Institute of Japan. D-1, Environmental engineering I (2008) 289–292.

[21] J. S. Bradley, From speech privacy to speech security, Proc. 19th ICA (2007) 15 pages.

[22] S. Amano, S. Sakamoto, T. Kondo, Y. Suzuki, Development of familiarity-controlled word lists 2003 (FW03) to assess spoken-word intelligibility in Japanese, Speech Commun. 51 (2009) 76–82.
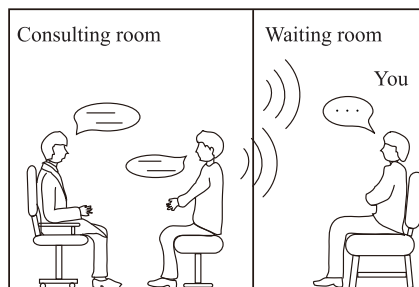
## List of Figures

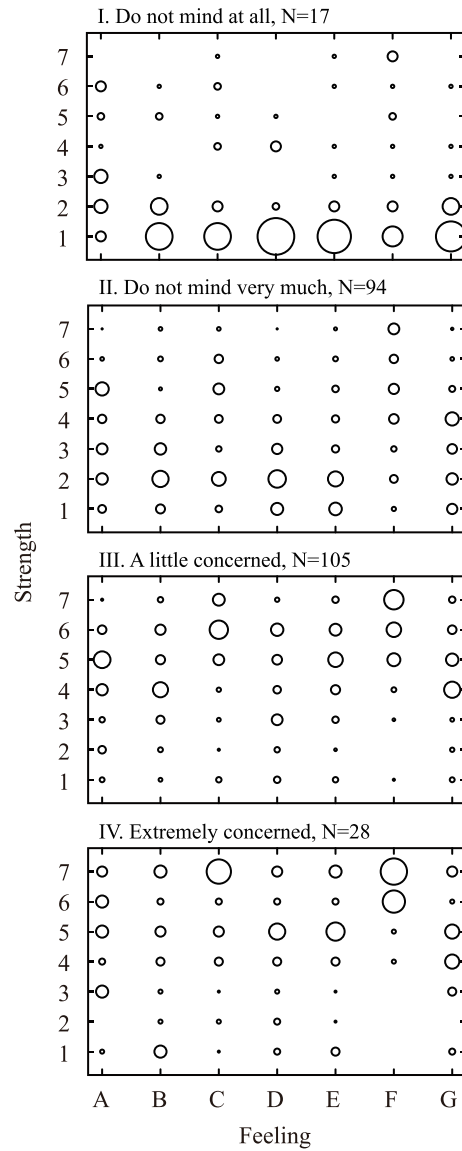Figure 1: The drawing used to explain the situation presented to the respondents.

Figure 2: Relative frequency distributions of the strength (1: Not at all, 7: Extremely) of each feeling for each group of respondents. The respondents were divided into four groups according to their sensitivity to personal information leakage. The diameter of the circles is proportional to the relative frequency of the response for each strength. N is the number of respondents for each group. The x-axis represents feelings: A. Curiosity, B. Unpleasantness, C. Unease, D. Restlessness, E. Embarrassment, F. Dissatisfaction, G. Anger.
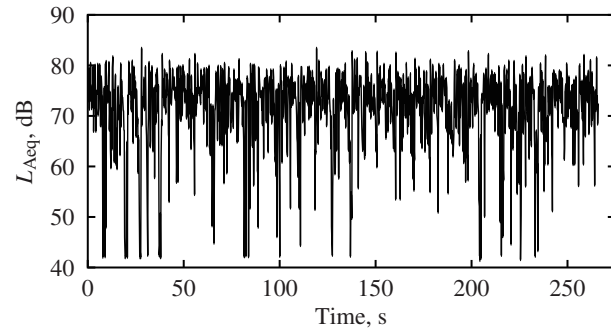
29

Figure 3: $L_{\mathrm{Aeq}}$ for the speech material in each time frame. Each frame has a length of 200 ms and is shifted without overlap along the time direction. The material was processed to ensure its stability over frames without losing its naturalness.
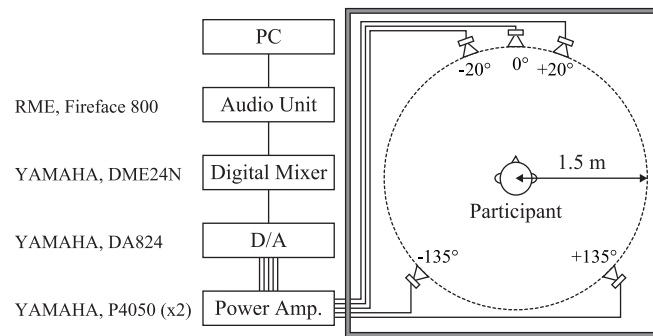
Figure 4: Block diagram of the test equipment and the loudspeaker arrangement used in the listening tests.

The conversation is perfectly intelligible and
understandable.

---------------------------------------------------------------------------- Boundary 1

Most of the conversation is intelligible and
understandable.

---------------------------------------------------------------------------- Boundary 2

Words of the conversation are sometimes intelligible
but it is not understandable as a whole.

---------------------------------------------------------------------------- Boundary 3

The coversation is audible but not even words
in the conversation are intelligible.

---------------------------------------------------------------------------- Boundary 4

The conversation is completely inaudible.

Figure 5: Possible boundaries of the speech understanding corresponding to participants'
sensitivity to personal information leakage through conversation.
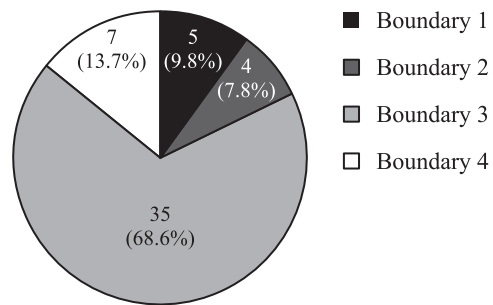
Figure 6: Number and percentage of participants indicating each boundary shown in Fig. 5.
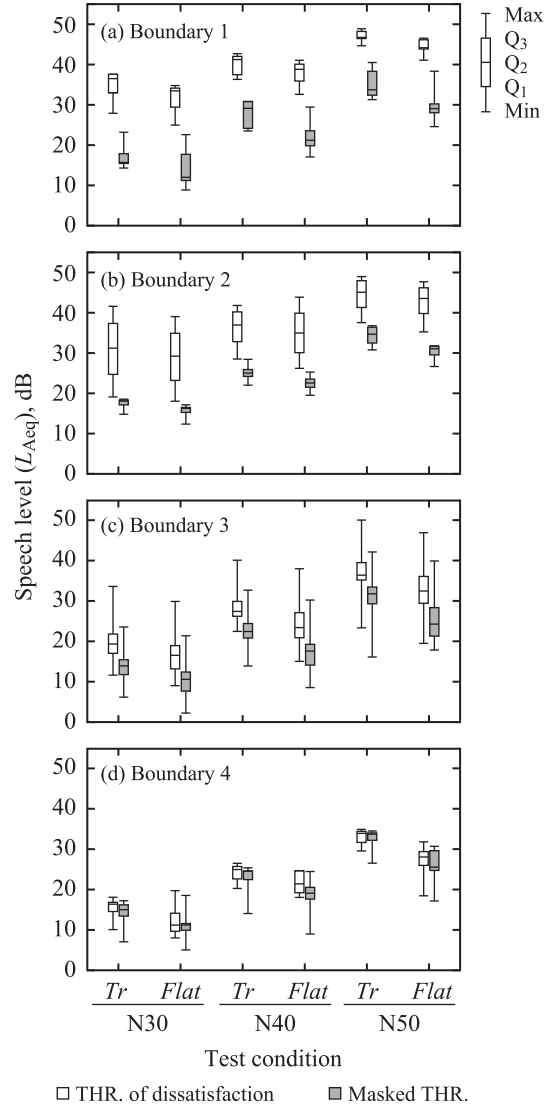
Figure 7: Statistical values of the threshold of the feeling of dissatisfaction and the masked threshold for each test condition. Panels (a) to (d) indicate the results for the groups of participants classified into boundaries 1 to 4, respectively. Open and shaded boxes indicate the threshold of the feeling of dissatisfaction and the masked threshold, respectively.
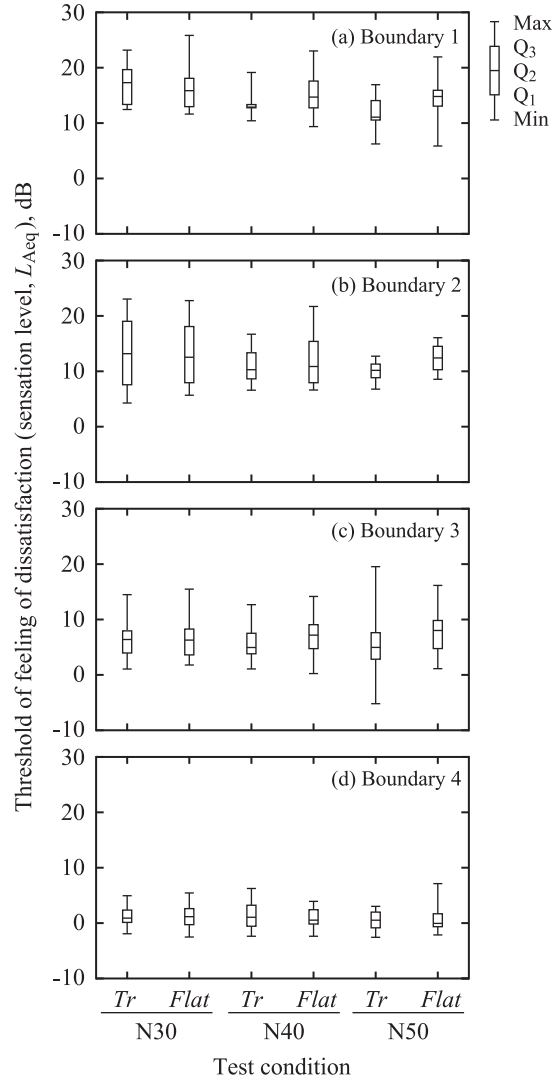
Figure 8: As Fig. 7 but for the threshold of the feeling of dissatisfaction normalized by the masked threshold, i.e., the threshold expressed as a sensation level.
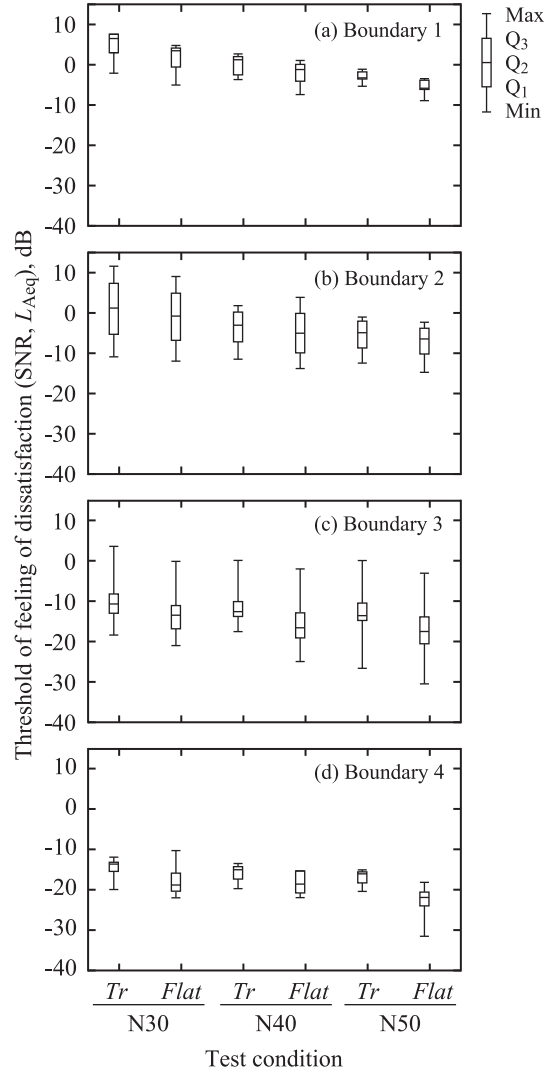
Figure 9: As Fig. 7 but for the threshold of the feeling of dissatisfaction normalized by the background noise level, i.e., the threshold expressed as a speech-to-noise ratio (SNR).
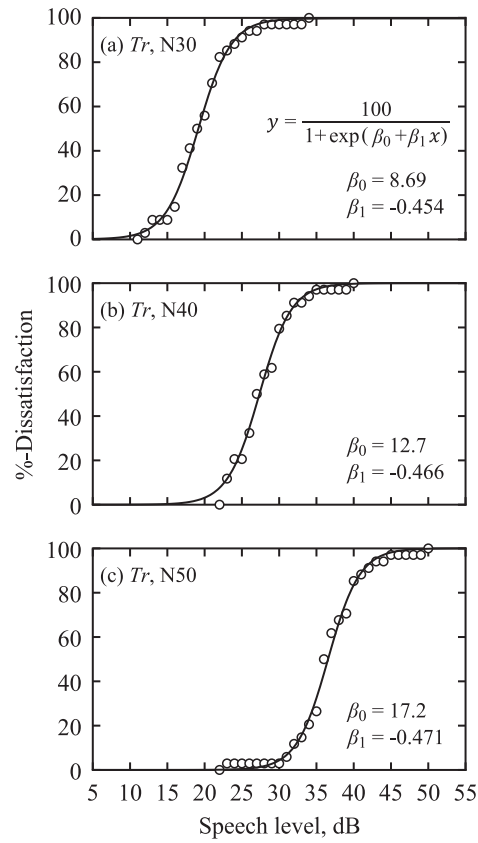
$$y = \frac{100}{1+\exp(\beta_0 + \beta_1 x)}$$

$\beta_0 = 8.69$
$\beta_1 = -0.454$

$\beta_0 = 12.7$
$\beta_1 = -0.466$

$\beta_0 = 17.2$
$\beta_1 = -0.471$

Figure 10: %-Dissatisfaction for $Tr$ as a function of speech level and its logistic regression curves for each background noise level.
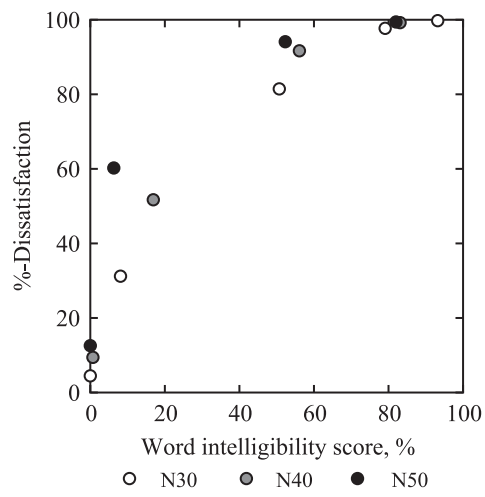
Figure 11: Relationship between %-Dissatisfaction and word intelligibility score for $Tr$. Symbols with different colors indicate different background noise levels.

## List of Tables

Table 1: Possible feelings evoked by overhearing speech from the consulting room

| | |
|---|---|
| A | *Curiosity* about the contents of the conversation in the adjacent room. |
| B | *Unpleasantness* evoked by hearing the consultation of others. |
| C | *Unease* evoked by imagining that others can hear the conversation between a doctor and me. |
| D | *Restlessness* evoked by imagining that others can hear the conversation between a doctor and me. |
| E | *Embarrassment* evoked by imagining that others can hear the conversation between a doctor and me. |
| F | *Dissatisfaction* evoked from the fact that I can hear the conversation of others from the consulting room. |
| G | *Anger* evoked from the fact that I can hear the conversation of others from the consulting room. |

Table 2: *p*-values obtained from Kruskal–Wallis rank sum tests on the strength of feeling. Forty-two pairs of an independent variable (one of six attributes of the respondents: gender, age, occupation, frequency of hospital visits, experience of overhearing speech in hospitals, and sensitivity to personal information leakage through conversation) and a dependent variable (one of seven feelings: A. Curiosity, B. Unpleasantness, C. Unease, D. Restlessness, E. Embarrassment, F. Dissatisfaction, G. Anger) were tested.

| Attribute | Feeling | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | A | B | C | D | E | F | G |
| Gender | $3.67 \times 10^{-1}$ | $1.90 \times 10^{-1}$ | $4.94 \times 10^{-3}*$ | $3.34 \times 10^{-2}*$ | $8.60 \times 10^{-3}*$ | $3.57 \times 10^{-4}*$ | $3.68 \times 10^{-2}*$ |
| Age | $1.86 \times 10^{-1}$ | $3.48 \times 10^{-3}*$ | $1.02 \times 10^{-2}*$ | $8.21 \times 10^{-3}*$ | $1.84 \times 10^{-1}$ | $3.65 \times 10^{-4}*$ | $9.10 \times 10^{-5}*$ |
| Occupation | $6.55 \times 10^{-1}$ | $1.53 \times 10^{-4}*$ | $4.23 \times 10^{-3}*$ | $3.70 \times 10^{-4}*$ | $4.73 \times 10^{-2}*$ | $6.62 \times 10^{-4}*$ | $1.50 \times 10^{-3}*$ |
| Frequency | $8.69 \times 10^{-2}$ | $1.12 \times 10^{-1}$ | $3.27 \times 10^{-1}$ | $2.12 \times 10^{-1}$ | $7.90 \times 10^{-2}$ | $7.10 \times 10^{-2}$ | $1.34 \times 10^{-1}$ |
| Experience | $4.22 \times 10^{-1}$ | $7.70 \times 10^{-1}$ | $3.68 \times 10^{-1}$ | $4.04 \times 10^{-1}$ | $1.18 \times 10^{-1}$ | $7.60 \times 10^{-1}$ | $6.81 \times 10^{-1}$ |
| Sensitivity | $1.11 \times 10^{-3}*$ | $4.61 \times 10^{-7}*$ | $2.69 \times 10^{-8}*$ | $5.09 \times 10^{-9}*$ | $1.03 \times 10^{-8}*$ | $2.25 \times 10^{-8}*$ | $5.63 \times 10^{-9}*$ |

$*p < 0.05$

Table 3: *p*-values obtained from the Wilcoxon signed-rank test between the strength of dissatisfaction and those of the other feelings (A. Curiosity, B. Unpleasantness, C. Unease, D. Restlessness, E. Embarrassment, G. Anger) for each group of respondents. The respondents were divided into four groups according to their sensitivity to personal information leakage: I. Do not mind at all, II. Do not mind very much, III. A little concerned, IV. Extremely concerned.

| Feeling | Sensitivity | | | |
|---------|------|------|------|------|
|         | I    | II   | III  | IV   |
| A | $9.49 \times 10^{-1}$ | $2.74 \times 10^{-10}*$ | $8.27 \times 10^{-10}*$ | $3.13 \times 10^{-4}*$ |
| B | $3.00 \times 10^{-2}$ | $1.13 \times 10^{-8}*$ | $9.95 \times 10^{-10}*$ | $1.47 \times 10^{-4}*$ |
| C | $2.28 \times 10^{-1}$ | $6.86 \times 10^{-4}*$ | $5.45 \times 10^{-4}*$ | $7.09 \times 10^{-3}$ |
| D | $5.48 \times 10^{-3}$ | $6.31 \times 10^{-10}*$ | $2.37 \times 10^{-11}*$ | $1.01 \times 10^{-4}*$ |
| E | $1.42 \times 10^{-2}$ | $7.82 \times 10^{-8}*$ | $2.06 \times 10^{-8}*$ | $1.25 \times 10^{-4}*$ |
| G | $1.37 \times 10^{-2}$ | $1.96 \times 10^{-10}*$ | $5.93 \times 10^{-13}*$ | $8.27 \times 10^{-5}*$ |

$*p < 2.38 \times 10^{-3} \ (= 0.05 / 21)$

Table 4: $p$-values obtained from the Wilcoxon rank sum test between thresholds for each boundary. $MT$, $DS$, and $DS_N$ are the masked threshold, dissatisfaction without normalization, and dissatisfaction with normalization, respectively.

| Pair of boundaries | Threshold | | |
|:---:|:---:|:---:|:---:|
| | $MT$ | $DS$ | $DS_N$ |
| 1-2 | $9.66 \times 10^{-1}$ | $3.20 \times 10^{-1}$ | $2.94 \times 10^{-2}$ |
| 1-3 | $1.08 \times 10^{-2}$ | $1.36 \times 10^{-11}*$ | $5.64 \times 10^{-14}*$ |
| 1-4 | $1.92 \times 10^{-1}$ | $2.59 \times 10^{-14}*$ | $2.20 \times 10^{-16}*$ |
| 2-3 | $1.50 \times 10^{-2}$ | $3.44 \times 10^{-6}*$ | $4.35 \times 10^{-7}*$ |
| 2-4 | $1.09 \times 10^{-1}$ | $2.67 \times 10^{-8}*$ | $4.45 \times 10^{-16}*$ |
| 3-4 | $4.80 \times 10^{-1}$ | $2.55 \times 10^{-3}*$ | $2.20 \times 10^{-16}*$ |

$*p < 8.33 \times 10^{-3} \ (= 0.05 \, / \, 6)$

Table 5: As for Table 4 but for the Wilcoxon signed-rank test between thresholds for each test condition.

| Pair of test conditions | | Threshold | | |
|---|---|---|---|---|
| | | $MT$ | $DS$ | $DS_N$ |
| $Tr$, N30 – | $Flat$, N30 | $2.04 \times 10^{-9}*$ | $9.04 \times 10^{-10}*$ | $2.59 \times 10^{-1}$ |
| $Tr$, N30 – | $Tr$, N40 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $1.50 \times 10^{-2}$ |
| $Tr$, N30 – | $Flat$, N40 | $2.04 \times 10^{-9}*$ | $9.58 \times 10^{-10}*$ | $7.42 \times 10^{-2}$ |
| $Tr$, N30 – | $Tr$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $2.66 \times 10^{-3}*$ |
| $Tr$, N30 – | $Flat$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $3.07 \times 10^{-2}$ |
| $Flat$, N30 | $Tr$, N40 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $2.43 \times 10^{-3}*$ |
| $Flat$, N30 | $Flat$, N40 | $5.30 \times 10^{-10}*$ | $7.79 \times 10^{-10}*$ | $2.51 \times 10^{-1}$ |
| $Flat$, N30 | $Tr$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $3.88 \times 10^{-4}*$ |
| $Flat$, N30 | $Flat$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $4.64 \times 10^{-2}$ |
| $Tr$, N40 – | $Flat$, N40 | $5.30 \times 10^{-10}*$ | $9.58 \times 10^{-10}*$ | $9.10 \times 10^{-5}*$ |
| $Tr$, N40 – | $Tr$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $1.76 \times 10^{-1}$ |
| $Tr$, N40 – | $Flat$, N50 | $1.05 \times 10^{-8}*$ | $1.21 \times 10^{-9}*$ | $3.22 \times 10^{-5}*$ |
| $Flat$, N40 | $Tr$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $4.29 \times 10^{-5}*$ |
| $Flat$, N40 | $Flat$, N50 | $5.30 \times 10^{-10}*$ | $5.30 \times 10^{-10}*$ | $3.81 \times 10^{-1}$ |
| $Tr$, N50 – | $Flat$, N50 | $1.02 \times 10^{-9}*$ | $5.30 \times 10^{-10}*$ | $2.16 \times 10^{-6}*$ |

$*p < 3.33 \times 10^{-3}$ (= 0.05 / 15)

Table 6: Word intelligibility scores for $Tr$ obtained in Ref. [6]. $L_S$ is the sound pressure level of the test words.

| $L_S$, dB | N30 | N40 | N50 |
|---|---|---|---|
| 12.4 | 0.0 | - | - |
| 17.4 | 8.1 | - | - |
| 22.4 | 50.7 | 0.7 | - |
| 27.4 | 79.1 | 16.9 | - |
| 32.4 | 93.2 | 56.1 | 0.0 |
| 37.4 | - | 83.1 | 6.3 |
| 42.4 | - | - | 52.3 |
| 47.4 | - | - | 82.0 |
| | | | (%) |