A compensation method for word-familiarity difference with SNR control in intelligibility test

SAKAMOTO Shuichi1, IWAOKA Naoki1, SUZUKI Yōiti1, AMANO Shigeaki2, KONDO Tadahsa†

1 Research Institute of Electrical Communication / Graduate School of Information Sciences, Tohoku University
2 NTT Communication Science Laboratories, NTT Corporation
saka@ais.rieic.tohoku.ac.jp

Abstract
Isolated-monosyllable listening tests are widely used in Japan to assess personal hearing ability. However, word intelligibility tests could be more suitable for measuring hearing ability. We developed new word lists with controlled word familiarity and phonetic balance. Listening test results showed slight differences in intelligibility scores among word lists within the same word familiarity rank. We examined whether such variation can be compensated by changing the signal to noise ratio (SNR) of each word to equalize intelligibility scores of word lists in identical word familiarity ranks. Results of listening tests indicated that a 1.0 difference in word familiarity was approximately equivalent to a 2.0 dB difference in SNR. Based on the results, we developed and tested a method of compensating for the difference caused by word familiarity by controlling SNR. We confirmed that this compensation method was effective in equalizing the intelligibility score of each word list.

1. Introduction
Isolated-monosyllable listening tests have been widely used in Japan to evaluate personal hearing ability. However, it is pointed out that scores by the tests do not always reflect personal hearing ability. This is probably because the isolated monosyllables are rarely used in daily conversation. To assess the hearing ability properly, we should use word-intelligibility and sentence-intelligibility tests. This has been increasingly recognized in Japan, and new lists have been proposed for these tests [1, 2]. However, although control of cognitive difficulty of word is crucial in intelligibility tests, many of the proposed lists do not properly control the cognitive difficulty. We proposed a word-lists for word intelligibility test controlling the cognitive difficulty by word familiarity [3]. The word familiarity indicates how subjectively familiar a specific word is. It is a better index of subjective difficulty than word frequency [4].

The proposed word list fairly well reflect personal hearing ability [5]. However, intelligibility scores show certain difference among the word lists, although the lists have the same averaged word-familiarity. The difference of intelligibility score may be attributable to the variety of word familiarities in each word list.

To equalize intelligibility scores of word lists in the same word familiarity range, we developed a new method of compensating for the word-familiarity difference by controlling SNR. We also performed word intelligibility test with this compensation method to confirm that it is effective.

2. Outline of the word lists [3]
We used a comprehensive word-familiarity database developed by Amano and Kondo [6] for controlling word familiarity. In this database, word familiarity is rated from 1 (low word familiarity) to 7 (high word familiarity) for all entry words (ca. 80,000) and subtitles in the Shinmeikai Japanese Dictionary (Fourth Ed.). First, LHHH-accent-type words (Types 0 & 4) consisting of four moras were selected from the words because these types are the most common ones in Japanese. Next, that population of words was divided into four ranks according to word familiarity: 7.0 to 5.5 (high word-familiarity words), 5.5 to 4.0 (middle-high word-familiarity words), 4.0 to 2.5 (middle-low word-familiarity words), and 2.5 to 1.0 (low word-familiarity words). We selected 1,000 words for each familiarity rank to compose 20 word lists containing 50 words each by maximizing the phoneme entropies to achieve optimum phonetic balance. Table 1 shows the average word familiarity score as well as its standard deviations (S.D.s) of the 20 word lists in each word familiarity rank. As shown in this table, the S.D.s are very small in all word familiarity ranks, indicating that the average word familiarity of the 20 word lists is well controlled.

3. Relationship between word familiarity and SNR

3.1. Experimental Procedure
To realize the method of compensating for the word-familiarity difference by controlling SNR, it is necessary to
Table 1: Averaged word familiarity scores of 20 word lists in each word-familiarity rank.

<table>
<thead>
<tr>
<th>word familiarity rank</th>
<th>average (S.D.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.0 - 5.5</td>
<td>5.81 (0.039)</td>
</tr>
<tr>
<td>5.5 - 4.0</td>
<td>4.84 (0.053)</td>
</tr>
<tr>
<td>4.0 - 2.5</td>
<td>3.15 (0.066)</td>
</tr>
<tr>
<td>2.5 - 1.0</td>
<td>2.16 (0.024)</td>
</tr>
</tbody>
</table>

estimate the amount of SNR needed to compensate for the difference of word familiarity. Therefore we performed word intelligibility test with the proposed word lists under various SNR conditions and obtained an equation for the relationship between SNR, word familiarity, and intelligibility score.

Word intelligibility test was performed with the proposed word lists in a soundproof room of the Research Institute of Electrical Communication, Tohoku University. Seven young male and three young female adult students participated in the experiment. They were all of normal hearing acuity and their ages ranged from 19 to 24. Five word lists were selected from each word familiarity rank and presented with noise to the listeners. The speech signal was uttered by a trained female speaker [6]. The noise signal was filtered random noise with a frequency spectrum simulating a long-term average spectrum of speech [7]. These speech and noise signals were generated by TDT System III and mixed electrically to monaurally present via a headphone (Sennheiser HDA-200) to a listener’s left ear. The speech signal was presented at 60 dBA; the SNR was set to a listener’s left ear. The speech signal was presented at 60 dBA; the SNR was set to the repeated measure. As a result, the difference between the intelligibility scores obtained by the word lists, three-way repeated-measure ANOVA was performed. In this analysis, SNR, list and word familiarity were treated as between-subject variables and the listener was treated as the repeated measure. As a result, the difference between the intelligibility scores was statistically significant in all conditions except following three: at high word familiarity for the SNR of −3 dB and −6 dB, and at middle-high familiarity for the SNR of −6 dB (p < .05).

To analyze the relationship in more detail, the intelligibility of each word was calculated. First, the number of listeners who correctly wrote the word was counted for each word. Next, we averaged the intelligibility scores of words with the same word familiarity value. This averaged intelligibility score was used to represent the intelligibility for a specific word familiarity value. Figure 2 shows the relationship between word familiarity and the intelligibility score. The area of the circles is proportional to the number of words with the same word familiarity. In this figure, a clear floor effect is observed at low word familiarity for the SNR of −9 and −12 dB, while a slight ceiling effect is observed at high word familiarity for the SNR of −3 dB.

A regression plane was calculated to illustrate the complementary relationship between word familiarity and SNR. A logistic regression model was fitted to the data. As a result, word intelligibility was given by the following equation:

\[
\text{Intelligibility}(F, SNR) = \frac{100}{1 + \exp(0.91 - 0.53F - 0.25SNR)}
\]  

(1)

where F and SNR are word familiarity and SNR, respectively. The coefficient of determination was 0.66. Figure 3 shows the word intelligibility estimated by Eq. (1) as a function of word familiarity and SNR.

4. Proposal of compensation method for difference of word familiarity

Equation (1) and Fig. 3 indicate that a difference of 1.0 in the word familiarity value is approximately equivalent to a 2 dB difference of SNR. Using this relationship, we proposed the following new compensation method for equalizing the variability of intelligibility scores of the 20 word lists in the same word-familiarity rank.

1. The median word familiarity value of each familiarity rank is calculated. The calculated values are 6.25, 4.75, 3.25, and 1.75.

2. Each calculated value is subtracted from word familiarity value of each word which is included in each familiarity rank.
5. Evaluation of the compensation method

5.1. Experimental Procedure

To confirm the effect of the proposed compensation method, we performed a word intelligibility test with the control of SNR of each word as described in the previous section.

Word intelligibility test was performed in a soundproof room of the Research Institute of Electrical Communication, Tohoku University. Seven male and three female adult students participated in these experiments. They were all of normal hearing acuity and their ages ranged from 18 to 22. None of them had participated in the previous experiment.

Ten word lists were selected from each word familiarity rank and presented with noise to the listeners. Of these lists, five had been used in the previous experiment (closed test) and the others had not (open test).

This Word intelligibility test was performed under two conditions. In the first condition, a speech signal was presented at 60 dBA, and SNR was fixed at −6 dB. This condition was for reference purposes. In the second condition, a speech signal was also presented at 60 dBA and SNR was changed for each word according to the proposed compensation method.

Experimental equipment, noise signals, and instructions for listeners were the same as in the previous experiments.
5.2. Results

Figure 4 shows the standard deviation of intelligibility scores obtained for the ten word lists. In the closed, the standard deviation of high word-familiarity words decreased with use of this compensation method, while the standard deviation of middle-low word-familiarity words slightly increased. The results of all ten lists showed that the standard deviations were below 4% in all familiarity ranks by using this compensation method.

6. Discussion

Figure 2 shows that word intelligibility is strongly influenced by word familiarity. Therefore, it must be important to control word familiarity of test stimuli when we measure personal hearing ability. When our word lists [3] were developed, word familiarity was divided into four ranks and the average of the word familiarity was equalized in each word list. Nevertheless, the result of the word intelligibility test showed a significant difference between intelligibility scores obtained by the word lists. This suggests that not only the average of word lists but also the distribution of word familiarity in each word list has to be considered. Figure 4 indicates that the influence of the distribution of word familiarity can be equalized by using the relationship between word familiarity and SNR (Eq. (1)).

The difference caused by cognitive difficulty of test stimuli is often compensated by controlling SNR, especially in a clinical situation [1, 2]. However, in such situations, SNR values to be applied are calculated from actual listening tests, so it takes much time to calculate these values. Moreover, when test stimuli are changed, these values have to be re-measured. In this study, the relationship between SNR and word familiarity was investigated. By using this relationship, we can easily calculate compensation scores of various test words based on word familiarity.

7. Conclusion

To equalize intelligibility scores of word lists in the same word-familiarity rank, we investigated the relationship between the differences of SNR equivalent to the differences of word familiarity. The results showed that a difference of 1.0 in the word familiarity value was approximately equivalent to a 2 dB difference of SNR. Thus, we proposed a new compensation method for the difference caused by word familiarity by controlling SNR. Results of a confirmation experiment indicate that this compensation method is effective in equalizing intelligibility scores of word lists with the same word familiarity ranks.

8. Acknowledgements

We are grateful to Prof. Shozo Makino (Graduate School of Engineering, Tohoku University) for his helpful comments and suggestions. This research was partially supported by JSPS, Grant-in-Aid for Scientific Research (B), 14350193, 2002.

9. References