SIMILARITIES OF WORDS IN NOISE IN JAPANESE

Kiyoko Yoneyama*
National Institute for Japanese Language
yoneyama@kokken.go.jp

ABSTRACT
This paper investigated how Japanese listeners perceived words in a noisy environment. The data of a word identification in noise experiment reported in Yoneyama (2002) were analyzed. There are three main findings. First, more than 86% of the total responses correctly reproduced the pitch accent patterns of the stimulus words, suggesting Japanese listeners hardly misperceived pitch accent patterns of words in a noisy condition. Second, vowels were perceived more correctly than consonants. Third, MDS analyses revealed that similarities of sounds were mapped in terms of phonologival features ([±high] and [±back] for vowels; [±voice] and [±sonorant] for consonants). These findings suggest that similarities of sounds that are highly related to the sonority scale of sounds in Japanese and pitch accent patterns contribute to word recognition in a noisy condition in Japanese.

1. INTRODUCTION
“Phonological neighborhood” in English is calculated based on sound similarities of word forms. Four different neighborhood calculations have been used. Among them, the only one calculation considers phonetic similarity of sounds, which is based on experimentally derived sound confusability [1][2]. This rule is based on R. D. Luce’s general biased choice rule [3]. Sound confusability matrices were calculated for CVC words in order to understand similarities of sounds in English. A basic assumption here is that if two sounds are similar, their confusability must be higher. This aspect was implemented in a neighborhood density calculation in [1] and [2].

This paper aims to investigate similarities of words from the actual experimental data in Japanese not from confusion matrices. In a word identification in noise experiment, words were presented to the participants in a noisy environment. They were asked to identify the words. Confusable sounds should induce more mistakes than less confusable sounds. If the sounds are similar, misperceived sounds should be very similar to the actually intended sounds. Therefore, the error patterns in the word identification in noise should provide similarities of sounds in Japanese. Therefore error patterns in the actual experiments should show the same general tendency. This possibility was explored by analyzing the error patterns in the word identification in noise experiment.

There are two main differences between the two previous word identification in noise experiments and this current experiment. First, this study considers how pitch accent patterns contribute to word similarity. In their study, participants typed hiragana characters as responses from the keyboard into a computer so that there was no way to understand whether listeners misperceived pitch-accent patterns or not. In contrast, the responses of the participants in this experiment were recorded on DATs so that word confusability in noise could be analyzed in terms of segments, as well as with respect to word-level prosody. Second, this experiment was able to demonstrate which kinds of phonological information other than a neighborhood density effect were used to recognize words in a noisy environment through multidimensional scaling analyses (MDS).

2. DATA
The data analyzed in this study were reported in Experiment 2 in [5]. The target words were 700 CVCVCV words with a rated auditory familiarity of 5 or higher on a 7-point scale with 7 being highly familiar in the NTT Database Series (Volume 1: [6]). They begin with a voiceless stop ([t, k]), a nasal ([n, m]), or a fricative ([s, z, ʃ, z, ʒ]).

The stimuli were created by adding noise to the audio files such that the signal-to-noise-ratio (SNR) at the point of peak RMS dB was 0 dB SPL. The SNR is the ratio of the amplitude of the stimuli against a constant level of Gaussian noise. The level of the noise was estimated from the peak RMS amplitude of each stimulus file. The noise extended 500 ms before and after each stimulus word. The stimuli were presented to the participants at 75 dB SPL, as measured using a sound level meter.

Participants were 27 native speakers of Tokyo Japanese who were born and raised in the Tokyo area (Tokyo, Kanagawa, Chiba, and Saitama). None of the participants had hearing difficulties.

The word identification data were collected as a secondary task. Participants were asked to write down what they said in hiragana characters after they repeated each word. Some participants used katakana characters and/or kanji characters in their responses. As long as a response was unambiguously interpretable, it was included in the analysis. It was emphasized to them that they should first repeat the words as quickly as possible, and then write down their responses. All the naming responses were recorded onto a digital audio tape (DAT), and the author analyzed the accuracy of spoken responses.

All the responses written by participants in hiragana characters were transcribed by the author in romanization and were saved in a master file for a segment analysis. The written responses were, then, checked against the named responses, as recorded onto the DAT tapes, in order to make sure that the responses were named with the correct accent patterns. If the participant wrote down something different from what he or she said, the transcription was “corrected” to match the oral response. The author first coded correct responses and missed responses as “1” and “0,” respectively.
using a computer script. The responses with a different pitch accent pattern were treated as missed responses and corrected to “0.” Fewer than 1% of the written responses were corrected.

Similarly, for an analysis of pitch accent patterns, the author coded correct pitch accent responses and missed pitch accent responses as “1” and “0,” respectively using a computer script separately in the master file.

3. PITCH ACCENT PATTERNS

All responses in this experiment were analyzed to see how Japanese listeners misperceived pitch accent patterns of the stimuli. In this analysis, the actually perceived accent pattern and the repeated accent pattern were compared. Here, all the responses (N = 18900) were analyzed in terms of pitch patterns only; misidentified segment responses were not considered.

Figure 1 shows frequency counts as a function of correct or missed responses in terms of pitch accent patterns in word identification data. As you can see, more than 86% of the total responses correctly reproduced the pitch accent patterns of the stimulus words.

This clearly suggests that Japanese listeners hardly misheard pitch accent patterns. Information of pitch accent patterns maintained in a noisy environment. Since Japanese listeners employ pitch accent information for lexical access ([1] [2] [3]), this might be related to the fact that pitch accent patterns could change meanings of words in Japanese.

4. SEGMENTS

This section discusses similarities of sounds in noise. The segments in the word identification data were analyzed. In this analysis, mispronunciations of the words about accent patterns are ignored because I am concentrating on consonant – vowel confusion. Furthermore, word positions were also ignored. The target words had the same CVCVCV structure. The mean number of errors of consonants and vowels in each of the two analyses were analyzed.

Table 1 shows the mean number of errors of consonants and vowels in each of the two analyses.

<table>
<thead>
<tr>
<th>Subjects analysis</th>
<th>Consonants</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>263 (SD: 46.57)</td>
<td>96 (SD: 20.99)</td>
</tr>
<tr>
<td>Items Analysis</td>
<td>10 (SD: 10)</td>
<td>4 (SD: 5)</td>
</tr>
</tbody>
</table>

Table 1: The mean number of errors of consonants and vowels in each of the two analyses.

Both subjects and items analyses showed that errors occurred significantly more often for the consonants than for the vowels (F1 = 1, 26) = 870.003, p < 0.001, F2 [1, 699] = 409.488, p < 0.001). Therefore, consonants and vowels were analyzed separately in MDS.

4.1. Multidimensional Scaling (MDS)

Similarities of sounds were computed from the error patterns of the responses from participants. In order to explore the underlying structure, MDS is applied to the similarities. The confusion matrices from the error patterns in a word identification in noise experiment generally provide information about which pairs of sounds are more confusable than others. The idea here is that if two sounds are similar, they are also confusable in noise. If the similarities of sounds are submitted to multidimensional scaling, the output should show the auditory similarities among sounds. If Japanese listeners classified the sounds based on auditory features, there should be meaningful dimensions in Multi-Dimensional scaling (MDS).

MDS is a statistical technique that is useful for uncovering meaningful organization in complex sets of data. Multidimensional scaling was proposed to help understand people’s judgments on the similarity of members of a set of objects. The matrix of similarity judgments was subjected to the multidimensional scaling to create the multidimensional map space. The map that results from a MDS treatment of a set of data is essentially a representation of the psychological relationships among sounds. If two sounds are more similar, the psychological distance in the MDS map should be less. That is, if the “auditory space” is “warped” by linguistic experience (as much cross-linguistic perception work suggests), then this is “auditory similarity.” I do not just mean general, universal “auditory space” here. This MDS analysis does not take position into account — so this is closer to “auditory similarity” than raw counts. Finally, the resulting dimensions must be interpreted to determine the most accurate number of solutions necessary to give the best final MDS solution for the data.

In this study, first, similarities of sounds were analyzed based on the error patterns. Correlation computes measures of similarity. Here, correlations between intended sounds that participants actually hear and perceived sounds that they thought they heard were investigated. The more similar the sounds, the higher the correlation should be. Correlations used in this analysis are a matrix of Pearson product-moment correlation coefficients. Pearson correlations vary between -1 and +1. A value of 0 indicates that neither of two variables can be predicted from the other by using a linear equation. A Pearson correlation of 1 or -1 indicates that one variable can be predicted perfectly by a linear function of the other. The calculated correlations were submitted to multidimensional scaling. The output was a multidimensional “map” to provide auditory similarity space for sounds in Japanese.

4.1.1. Vowels

First, vowels and consonants of the responses were separately tabulated. Some of the responses including a moraic nasal or a palatalized /n/ were excluded from the analyses, since no corresponding input sounds existed. That is, the matrices were designed to make similarity symmetrical. The vowel data and the consonant data were converted into proportions since the frequencies of different segments in the stimulus words were not balanced. Then the data were submitted to Pearson’s correlation analyses to compute similarities among vowels. This similarity matrix was then submitted to MDS in order to explore the underlying structure.

Figure 2 shows the two-dimensional scaling solutions for similarity of vowels as the best fit for the vowel data. The R² value for this solution is 0.99957 and stress is 0.00654. Dimension 1 represents the vowel height, and is graphed along the horizontal axis. Dimension 2 is graphed along the vertical.
and /r/ are not considered sonorants in an auditory-based view. However, the data suggest that not all the sounds are perceived as [±sonorant]. Among consonants, nasals (/m/, /n/, /g/), glides (/j/, /w/), and liquids (/r/) are considered as sonorants articulate in general from an articulatory point of view. However, the data suggest that not all the sounds are auditorily considered as sonorants. Phonetic realization of /w/ and /r/ are not considered sonorants in an auditory-based psychological space. [w] is located near [k] which is apart from nasals and [j]. [r] is very close to [d], suggesting that they are auditorily very similar. Recall that Dutch listeners were not able to respond to Japanese [r] with the visual target ‘r’ whereas when the visual target was changed from ‘r’ to ‘d’, they were able to detect [r] in a phoneme monitoring experiment [9]. This is a piece of evidence that a phonetic realization of /r/ is a flap in Japanese. It is natural to treat /g/ as a sonorant, once we remember that /g/ = [ŋ] in Tokyo Japanese. /ŋ/ in the onset position is realized as [ŋ] after a vowel or a moraic nasal in Tokyo Japanese, and there is a shift from [ŋ] to [ɡ] apparently now in progress [10]. Since all the counts for [ɡ] in this analysis actually occur in this phonological environment, ‘g’ naturally appears to be [ŋ]. However, in MDS, [ŋ] is apart from a cluster of nasals and it is about midway between a cluster of nasals (/m/ and /n/) and a cluster of voiceless stops (/p/ and /t/). In the future, /ɡ/ is expected to be moving towards voiced stops so that voiced consonants would be classified in terms of [±sonorant] more clearly.

5. DISCUSSION AND CONCLUSIONS

The response patterns in the word identification in noise experiment provided interesting findings. First, more than 86% of the total responses was identified correctly in terms of pitch accent patterns. Japanese listeners hardly misperceived pitch accent patterns of words in a noisy condition. This may imply that pitch accent patterns constrain the neighbors in Japanese – perhaps the words that have the same pitch accent pattern might be only considered as neighbors.

The analyses also demonstrated that Japanese listeners classified sounds in terms of phonological features in a noisy condition. Confusion matrices of vowels and consonants submitted to MDS enabled us to see how Japanese listeners perceived sounds and arranged them within the auditory similarity map. Both MDSs for consonants and vowels provided interpretable dimensions, both of which turned out to be phonological features.

The outputs of the MDS represent auditory similarity spaces for consonants and vowels. The five Japanese vowels were classified in terms of Height and Backness. Height is realized to measure an acoustic parameter, the first formant (F1). It is also related to sonority, and the bigger distance between /i/, /u/ vs. /e/, /o/ than between /e/, /o/ vs. /a/ supports this. Backness is an acoustic property of the second formant (F2).

Similarly, the phonological features that emerged from MDS for consonants were auditory features: Voice and Sonorant. Voicing and Sonority are not independent: they are based on activity of the vocal folds. Dimensions used to describe consonants and vowels are really basic features of sounds, all of which are related to the sonority scale of sounds in the language.

In conclusion, the results of error patterns in the word identification in noise data revealed that similarities of sounds that are highly related to the sonority scale of sounds in Japanese and pitch accent patterns contribute to word recognition in a noisy condition in Japanese.

6. REFERENCE

[1] Luce, P.A. Neighborhoods of words in the mental lexicon, (Research on speech perception technical report 6), Bloomington, IN, Indiana University, 1986.

* I thank Keith Johnson, Mary Beckman, Mark Pitt, JJ Nakayama, and Takashi Otake for their help. I am also grateful for support from the National Institute for Health for a grant entitled “Cross-linguistic studies of spoken language processing” (R01DC04421, PI: Keith Johnson), which supported this study.
Figure 1: Frequency counts as a function of correct or missed responses in terms of pitch accent patterns.

Figure 2: MDS for vowels. Dimensions 1 and 2 represent vowel height (F1) and backness (F2), respectively.

Figure 3: MDS for consonants. F = [f], SS = [ʃ], C = [ts], CC = [c], KY = [kʰ], Z = [ʒ], Y = [j]. Dimensions 1 and 2 represent [±voice] and [±sonorant], respectively.