ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

NEIGHBORHOOD EFFECTS ON SPOKEN WORD RECOGNITION IN JAPANESE

Shigeaki Amano and Tadahisa Kondo
NTT Communication Science Laboratories
3-1 Morinosato Wakamiya, Atsugi, Kanagawa 2430198, Japan
amano@av-hp.brl.ntt.co.jp kondo@cs.brl.ntt.co.jp

10.21437/Eurospeech.1999-378

ABSTRACT

Reaction time of lexical decision and word recognition score were measured to investigate whether the neighborhood, a word candidate set with single mora substitution with a target word, has inhibitory effects on spoken word recognition in Japanese. Partial correlation analyses were conducted between subjects' performance of spoken word recognition (reaction time and recognition score) and neighborhood properties (density, mean familiarity, maximum familiarity, and sum of familiarity). The results showed significant inhibitory effects of neighborhood on the recognition score but not on the reaction time. It is suggested that some amount of time is necessary for the neighborhood to be activated and to compete with a target word in spoken word recognition in Japanese.

1. INTRODUCTION

1.1. Background

Researches on spoken word recognition suggested that multiple word candidates are activated during the recognition process (e.g., [1], [2], [3]), and that these candidates compete with each other (e.g., [4], [5], [6], [7]). However, there has been no general agreement on the set of word candidates which are activated and compete. Previous studies suggested several different sets of word candidates for spoken word recognition. One of the most well-known sets is a neighborhood. It is defined as a collection of words which differ only one segment (usually a phoneme) from a particular target word (e.g., [8], [9], [10]).

Many studies showed that the neighborhood has significant effects on word recognition performance both in on-line and off-line tasks. For example, the number of words and word frequency in neighborhood negatively correlate with recognition score of a target word [11]. The number of words in neighborhood have inhibitory effects on reaction times in lexical decision and naming for a target word [12]. A low frequency word in a neighborhood have negative priming effects on target word recognition [13]. Recognition of two-syllable target words (spondees) is affected by neighborhood characteristics of each syllable in the word [14]. Older adults have difficulty recognizing a target word if its neighborhood has a large number of words and high frequency words [15]. Taken together, all these studies clearly indicate that the neighborhood has competitive and inhibitory effects on spoken word recognition.

However, these neighborhood studies were carried out in English. The neighborhood effects in Japanese have not been studied. The present study addressed the question whether the neighborhood affects spoken word recognition of Japanese in on-line and/or off-line tasks. For that purpose, performances in a lexical decision task and a word recognition task were measured and their relationships to the neighborhood were analyzed in this study.

1.2. Target Word Properties

The following three variables were used for a target word.

TV1: Reaction time.
TV2: Recognition score.
TV3: Auditory word familiarity.

The 'reaction time' (TV1) of lexical decision was measured in Experiment 1 as an on-line performance, and the 'recognition score' (TV2) with/without noise was measured in Experiment 2 as an off-line performance. 'Auditory word familiarity' (TV3) was used because it is a well-known factor which affects spoken word recognition. For example, Connine, Mullennix, Shernoff, and Yelen [16] showed that high-familiarity spoken words in English produced shorter reaction times than low-familiarity spoken words in a lexical decision task and a delayed naming task. They also showed that the high-familiarity spoken words are recognized better than the low-familiarity spoken words in these tasks. Similarly, Amano and Kondo [17] showed that auditory word familiarity strongly and consistently affects reaction times and recognition scores of Japanese spoken words. These results indicate that it is important to include the auditory word familiarity as the target word variable.

1.3. Definition of Neighborhood

Neighborhood of a Japanese spoken word was defined with the following conditions.

J1: Neighborhood is a set of words with a single mora substitution with a target words.
J2: Words with the same mora sequence with different pitch accent were regarded as different words.
J3: Homophones are regarded as different words.
J4: Population of neighborhood is entire words (about 80,000) in the Japanese word familiarity database in PSYLEX [18].

The neighborhood in English is usually defined using a phoneme as a segment (e.g., [11], [13]). However, as shown in J1 condition, the neighborhood in this study is defined using a mora as a segment. This is because the mora is said to be a segmentation unit of speech perception for Japanese natives ([19], [20]) and therefore it is most probably used in a lexical processing of Japanese spoken words.

1.4. Neighborhood Properties

Neighborhood was described with the following variables.

NV1: Density.
NV2: Mean auditory word familiarity.
NV3: Maximum auditory word familiarity.
NV4: Sum of auditory word familiarity.

The 'density' (NV1) is the number of words in a neighborhood. The 'mean auditory word familiarity' (NV2) is an averaged auditory word familiarity of all the words in a neighborhood. The 'maximum auditory word familiarity' (NV3) is the highest auditory word familiarity of a word in a neighborhood. The 'sum of auditory word familiarity' (NV4) is total amount of auditory word familiarity of words in a neighborhood. The auditory familiarity which ranged from 1 (unfamiliar) to 7 (familiar) was obtained from Japanese word familiarity database in PSYLEX [18]. The auditory word familiarity was used in this study instead of word frequency which is usually used in neighborhood studies in English (e.g., [11], [12], [13], [14], [15]). This is because the auditory word familiarity is measured in the same modality as spoken words but the word frequency is not. The word frequency is usually the counted occurrence of a 'written' word in newspapers and journals. In addition, Amano and Kondo [17] showed that the auditory word familiarity exerts strong effects on spoken word recognition in Japanese. Therefore, the auditory word familiarity is better than the word frequency to describe the neighborhood of spoken words.

1.5. Partial Correlation Analysis

Amano and Kondo's [17] results lead the idea that auditory word familiarity so strongly affects the target word recognition that the neighborhood effects might be very hard to be observed. Therefore, partial correlation was used to analyze the data, so that the effects of the auditory word familiarity can be excluded (Figure 1). By this partial correlation, the pure relationship between the target word and the neighborhood should be observed.

2. EXPERIMENT 1

2.1. Objectives

The objective of Experiment 1 is to investigate the relationships between neighborhood and reaction times of lexical decision for Japanese words.

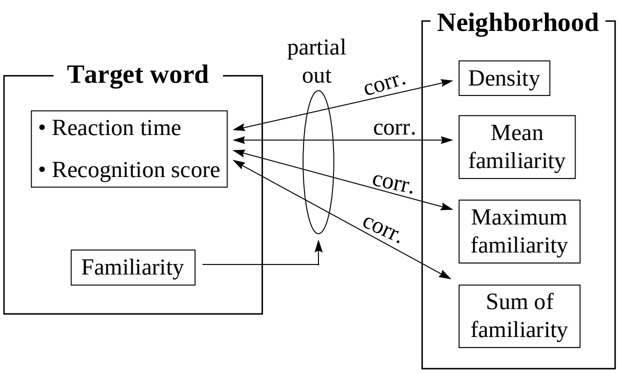

Figure 1. Relationships among variables in a partial correlation analysis between a target word and a neighborhood.

2.2. Method

2.2.1. Subjects

All Japanese participants undertook the kanji-word reading ability test (100 Rakan) [21] as a screening test, and only those who scored between 70% and 80% of accuracy participated in the experiment. Thirty participants (15 males and 15 females) passed the screening test. They were between 18 and 30 years old, and had no hearing impairment.

2.2.2. Stimuli

Japanese words were selected from the word familiarity database in PSYLEX [18], according to the following conditions.

W1: Auditory word familiarity is ranged 4.0-4.5, 4.5-5.0, 5.0-5.5, 5.5-6.0, and 6.0-7.0.
W2: Word length is 4 moras.
W3: Accent pattern is low-high-high-high.
W4: No homophone exists.
W5: Difference between auditory and visual word familiarity is less than 0.15.
W6: No homograph exists.
W7: Orthography consists of two kanji characters.

The W5, W6, and W7 were additional conditions for future visual experiments using the same words as used for the current auditory experiment. Two word sets of 100 words were made by collecting 20 words in each word familiarity range. Nonwords consisted of very low-familiarity words which were selected from the word familiarity database with the following conditions.

N1: Auditory word familiarity is ranged 1.0-2.0.
N2: The conditions from W2 to W7.

The N1 condition assures that the stimuli are not recognized as words and that they do not violate Japanese phonotactic rules (e.g., transitional probability of mora). One nonword set was made by collecting 100 nonwords. Two stimulus sets of 200 stimuli were made by combining the word set 1 and the nonword set (hereafter called stimulus set 1), and the word set 2 and the nonword set (hereafter called stimulus set 2). Another 10 words and 10 nonwords were selected for practice trials from the word familiarity database according to the same conditions described above. Speech files of these stimuli were extracted from PSYLEX [18]. They were digitally recorded files (16 bit - 16 kHz sampling rate) pronounced by one Japanese female speaker. All stimuli were digitally adjusted to have the same root-mean-square power.

2.2.3. Procedure

The stimuli were diotically presented to the subjects through headphones at 74 dBSPL. In each trial, a warning signal (1 kHz pure tone, 150 ms) was presented first. One second after it, a word or a nonword was presented. The subjects were instructed to push one key with their first finger and another key with their second finger as soon as possible when they thought that they heard a word or a nonword respectively. The subjects were also instructed to try to avoid errors as much as possible. The interval between successive trials was about two seconds. Reaction times were measured from the word/nonword stimulus onset. A half of the subjects heard the stimulus set 1, the other half heard stimulus set 2. Each subject heard 200 experimental trials after three sets of 20 practice trials. There was a short break after every 50 trials. The order of the stimuli was randomized for each subject.

2.3. Results

Responses were determined as errors when one of the following conditions was satisfied.

E1: wrong key is pressed
E2: reaction times which is more than 1600 ms
E3: reaction times which is less than 160 ms

The data from the two subjects were excluded from the analyses because their error rates were greater than 25%. Otherwise, the erroneous responses were excluded from the analyses. Partial correlation coefficients between lexical decision time and neighborhood variables were calculated by excluding the factor of familiarity of the target word. The coefficients were .103 between reaction time and density, .083 between reaction time and mean auditory word familiarity, .030 between reaction time and maximum auditory word familiarity, and .109 between reaction time and sum of auditory word familiarity. None of the partial correlation coefficients was significant.

2.4. Discussion

Although the partial correlation coefficients were positive values suggesting tendency of weak inhibitory effects, none of them was significant in Experiment 1. Therefore, the results indicate that neighborhood does not affect reaction times of lexical decision for spoken words. This result seems to be inconsistent with Goldinger's study [12] which showed that the density of a neighborhood has inhibitory effects on reaction time of lexical decision in English. However, direct comparison is difficult because he used only high-frequency target words. It might be argued that the neighborhood is not activated in a time span in which lexical decision is made, or even if it is activated, its activation level is not high enough to compete with a target word. If so, the effects can be observed in an off-line task which allow the neighborhood to be activated at a response point of a subject. This possibility was investigated in the next experiment.

3. EXPERIMENT 2

3.1. Objectives

Experiment 2 was conducted to investigate the relationship between neighborhood and recognition score of a word.

3.2. Method

3.2.1. Subjects

Forty Japanese adults (20 males and 20 females) aged between 18 and 30 years participated in the experiment. They scored from 70 to 80% of accuracy in the kanji-word reading ability test (100 Rakan) [21]. They had no hearing impairment.

3.2.2. Stimuli

The 200 words in Experiment 1 were used as stimuli in Experiment 2. To prevent the ceiling effect, the words were prepared by adding the Hoth noise. The Hoth noise had a spectral shape with about 5dB/oct decreasing in higher frequency. Frequency range of the noise was between 80 Hz and 8 kHz. The noise started 100 ms before the onset of a word stimulus, and lasted about 100 ms to 200 ms after the end of the word stimulus. The onset of the noise was tapered for 20 ms to prevent an audible click. There were five signal-to-noise ratios (-5.0, -2.5, 0.0, and 2.5 dB, or no noise which is hereafter called 'org'). The 10 word stimuli used for the practice trial in Experiment 1 were also prepared in the same noise-adding operation for a practice trial in Experiment 2. All stimuli were digitally adjusted to have the same root-mean-square power after the noise-adding operation. All the 'org' stimuli were also digitally adjusted to have the same root-mean-square power.

3.2.3. Procedure

The word stimuli were diotically presented to the subjects through headphones at 70 dBSPL. The subjects were instructed to type the presented word in 'hiragana' using a computer keyboard with a 'romaji-to-hiragana' conversion software. They were also instructed to make a guess and type their answer even when they could not identify the word. Forty words in each word familiarity range were assigned to different signal-to-noise ratio, so that one subject did not hear a particular word more than once in different noise conditions for preventing unexpected memory effects. The assignment of these was counterbalanced across subjects. Each subject undertook 200 trials for the word stimuli after the 10 practice trials using practice stimuli. There was a short break after 100 trials. The order of the stimuli was randomized for each subject.

3.3. Results

Word recognition scores were obtained by dividing the number of correct answers by the total number of answers for each spoken word at each signal-to-noise ratio. Partial correlation coefficients between word recognition score and neighborhood variables (density, mean familiarity, maximum familiarity, and sum of familiarity) were obtained by excluding the factor of familiarity of the target word, and they were plotted in Figure 2 as a function of signal-to-noise ratio. Partial correlation coefficients between averaged recognition score and neighborhood variables were also plotted in Figure 2 designated as 'All'.

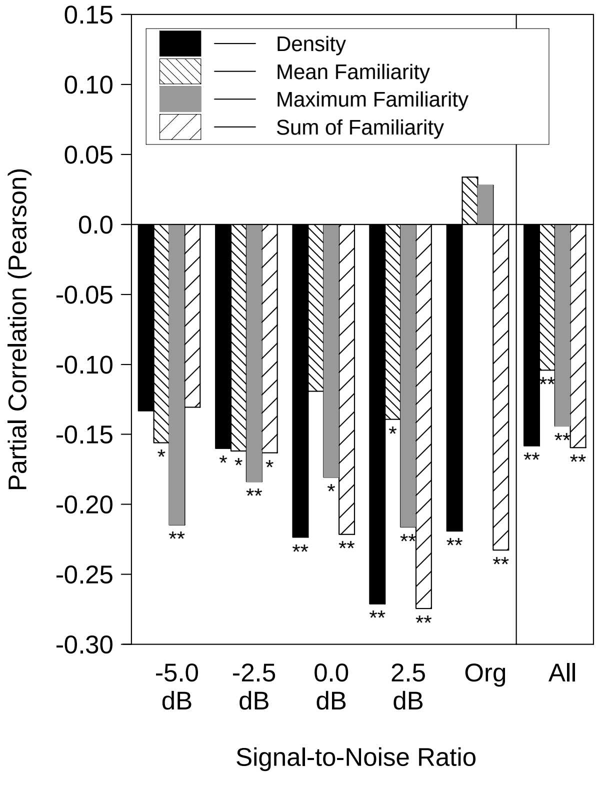

Figure 2. Partial correlation coefficient between word recognition score and neighborhood variables (density, mean familiarity, maximum familiarity, and sum of familiarity) as a function of signal-to-noise ratio. Single and double asterisks respectively represent 5% and 1% significance.

3.4. Discussion

Although a few partial correlation coefficients were not significant, almost all of them were negative and significant. Especially, the partial correlation coefficients between averaged recognition score and neighborhood variables were all negative and significant. Therefore, the result of Experiment 2 indicates that the neighborhood has inhibitory effects on the spoken word recognition in an off-line task in Japanese. That is, the higher the neighborhood variables (density, mean word familiarity, maximum word familiarity, and sum of word familiarity) are, the worse the word recognition score is. Taken together with the results of Experiment 1 that the neighborhood is not effective on spoken word recognition in an on-line task, it is suggested that the neighborhood requires some amount of time to be activated enough to compete with a target word.

This idea on the time course of lexical competition is supported by the study of Vroomen and de Gelder [7]. They found that facilitative cross-modal priming effects are larger for auditory primes with few or no competitors than for auditory primes with many competitors, which suggest that competitors have an inhibitory effect on word recognition according to their numbers (i.e., density). More importantly, they also found that the difference of priming effects according to the number of competitors disappeared when there was no interstimulus interval between auditory prime and visual target word, which can be interpreted that lexical competitors require some amount of time to be activated. Although their task is not identical to this study, their results are consistent with the present results in term of time at which lexical competitors interact with a target word.

The present study showed that the neighborhood affects on spoken word recognition of Japanese in an off-line task. However, it does not mean that the neighborhood defined in this study is the best word candidate set for the recognition. Other kind of word candidate set might be more suitable for representing lexical competition. One of the possibilities is the different kind of neighborhood which includes words with single phoneme deletion or addition to the target words in addition to the substitution (e.g., [11], [12], [15]). The other possibility is the cohort (e.g., [22]) which shares the initial part of a word. Further studies should be conducted on these issues.

4. CONCLUSION

The present study investigated neighborhood effects on spoken word recognition in Japanese. The inhibitory neighborhood effects were found on recognition scores, but not on lexical decision times for Japanese spoken words. It is suggested that the neighborhood requires some times to be activated for lexical competition with a target word.

5. REFERENCES

[1] Shillcock, R. (1990), Lexical hypotheses in continuous speech. In: G. T. M. Altmann (ed.) *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MIT Press, pp. 24-49.

[2] Connine, C. M., Blasko, D. G., and Wang, J. (1994), Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, 56, pp. 624-636.

[3] Zwitserlood, P. (1989), The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32, pp. 25-64.

[4] McQueen, J. M., Norris, D., and Cutler, A. (1994), Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, pp. 621-638.

[5] Hamburger, M. B., and Slowiaczek, L. M. (1996), Phonological priming reflects lexical competition. *Psychonomic Bulletin and Review*, 3, pp. 520-525.

[6] Slowiaczek, L. M., and Hamburger, M. B. (1992), Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, pp. 1239-1250.

[7] Vroomen, J., and de Gelder, B. (1995), Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21, pp. 98-108.

[8] Frauenfelder, U. H. (1990), Structure and computation in the mental lexicon. In: H. Haken & M. Stadler (eds.) *Synagetics of cognition*. Berlin, Springer-Verlag. pp. 406-414.

[9] Frauenfelder, U. H., Baayen, R. H., Hellwig, F. M., and Schreuder, R. (1993), Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language*, 32, pp. 781-804.

[10] Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. (1985), Speech perception, word recognition and the structure of the lexicon. *Speech Communication*, 4, pp. 75-95.

[11] Luce, P. A. (1986), Neighborhoods of words in the mental lexicon. *Research on Speech Perception, Technical Report*, 6, Bloomington IN, Indiana University, pp. 1-91.

[12] Goldinger, S. D. (1989), Neighborhood density effects for high frequency words: Evidence for activation-based models of word recognition. *Research on Speech Perception, Progress report*, 15, Bloomington IN, Indiana University, pp. 163-186.

[13] Goldinger, S. D., Luce, P. A., and Pisoni, D. B. (1989), Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, pp. 501-518.

[14] Cluff, M. S., and Luce, P. A. (1990), Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance*, 16, pp. 551-563.

[15] Sommers, M. S. (1996), The structural organization of the mental lexicon and its contribution to age-related declines in spoken-word recognition. *Psychology and Aging*, 11, pp. 333-341.

[16] Connine, C. M., Mullennix, J., Shernoff, E., and Yelen, J. (1990), Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, pp. 1084-1096.

[17] Amano, S. and Kondo, T. (in press), Familiarity effect on spoken word recognition in Japanese. *Proceeding of the 14th International Congress of Phonetic Sciences*.

[18] Amano, S. and Kondo, T. (1999), *PSYLEX*. Sanseido, Tokyo.

[19] Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993), Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, pp. 258-278.

[20] Cutler, A., and Otake, T. (1994), Mora or Phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, pp. 824-844.

[21] Kondo, T. and Amano, S. (1998), Reading ability test for Kanji word (100 Rakan): New estimation method for language skill using word familiarity. *Proceeding of the 62nd annual meeting of the Japanese Psychological Association*, p.711. (in Japanese).

[22] Marslen-Wilson, W. D., and Welsh, A. (1978), Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, pp. 29-63.