Perceptual training of vowel length contrast of Japanese by L2 listeners: Effects of an isolated word versus a word embedded in sentences

Mee Sonu\textsuperscript{1,4}, Keiichi Tajima\textsuperscript{2}, Hiroaki Kato\textsuperscript{3} and Yoshinori Sagisaka\textsuperscript{4}

\textsuperscript{1} Faculty of Science and Technology, Sophia University, Japan
\textsuperscript{2} Department of Psychology, Hosei University, Japan
\textsuperscript{3} NICT, Japan
\textsuperscript{4}GITI / LASS Lab, Waseda University, Japan

sonumee@toki.waseda.jp, tajima@hosei.ac.jp, kato.hiroaki@nict.go.jp, sagisaka@giti.waseda.jp

Abstract

In an attempt to improve the perception of vowel length contrasts in Japanese by L2 learners (L1 Korean), we compared two different training methods. The first one involved training L2 learners with sets of isolated words contrasting the vowels (Word training), whereas the other training involved presenting same words within sentences (Sentence training). Word training and sentence training both led to significant improvement in the learners’ ability to identify vowel length contrasts in Japanese, and, both types of training improved the learners’ ability to perceive consonant length contrasts in Japanese, which listeners were not trained to identify. Although the amount of overall improvement was not significantly different between training methods, sentence training showed improvement in a wider range of conditions than word training. These results indicate that word training and sentence training are both effective at improving perception of length contrasts in Japanese, but that sentence training may have some advantage over word training.

Index Terms: perceptual training, vowel length contrast of Japanese, learning effect, presentation contexts, generalization of learning

1. Introduction

Japanese language uses a length (segmental duration) contrast of vowels and consonants for phonemic distinctions [1]. In the learning of the Japanese language, it has been observed that the length contrast of Japanese is one of the most difficult training items for second language (L2) learners [2-4]. This difficulty has been understood as a result of the fact that L2 learners cannot adjust categorical criteria adaptively using vowel and consonant durations that vary with context, such as speaking rate, presentation context, and neighboring segments [3-5]. However, the relationship between the identification difficulty of L2 phonetic contrasts by L2 learners and these contextual factors has not been thoroughly investigated. In addition, effective training methods to assist the L2 learners’ identification of L2 phonetic contrasts have yet to be developed. Thus, the present study aims to propose an effective perceptual training scheme to help L2 learners perceive Japanese phonemic length contrast in the face of contextual factors that affect segment duration.

In terms of perceptual training and learning effect, previous studies have shown that perceptual training helps L2 learners build up new phonetic categories [6,7]. The relationship between the effect of perceptual training and presentation context (presentation of target words in isolation or embedded in a sentence) has also been discussed [3, 4, 8].

Tajima et al. [4] trained native English listeners to identify vowel length contrasts in Japanese using isolated words, and tested their performance before and after training using stimuli varying in various factors. The results indicated that the perceptual training improved accuracies for vowel length contrasts, but did not significantly generalize to consonant length contrasts, when compared with an untrained control group. Tajima et al. [4] pointed out that the limited training effects may be due to the fact that only isolated words spoken at a single speaking rate were used as training materials. To look for a more effective perceptual training scheme for L2 learning, we need to directly investigate how varying the training materials, e.g., varying the presentation context, affects the effect of training.

Several other studies have also investigated the relationship between perceptual training effects and presentation context. Greenspan et al. [8] compared training using isolated words and training using words in sentences. The results showed that isolated word training only improved perception of isolated words. In contrast, sentence training improved perception of not only isolated words but also words in sentences. Similarly, Hirata’s [3] examined the perceptual training of vowel length contrasts using both words in isolation and words in sentences as training materials. Results indicated, firstly, that the effectiveness of perceptual training was similar between word training and sentence training. Second, learning effect in one context generalized to the other. However, training with sentences showed an advantage over training with words.

Both Greenspan et al.’s [8] and Hirata’s [3] studies suggest that sentence training may be more advantageous than word training. However, their tasks differ from each other and from Tajima et al.’s [4] study. For example, Hirata [3] trained listeners to identify the number of moras in Japanese words, and found an advantage for sentence training over word training. However, it has not been confirmed whether a sentence training advantage would also occur when listeners are trained to identify words in a minimal pair identification task. Moreover, both studies examined whether word training generalized to sentence training, and whether sentence training generalized to word training, but neither studies examined whether training generalizes to other broader contexts, e.g., whether training vowel length contrasts generalizes to perception of consonant length contrasts.

Accordingly, the first purpose of the present study is to investigate the effectiveness of using words in isolation and words in sentences in training learners to identify length contrasts in Japanese, using the same task as Tajima et al. [4]. Specifically, one group of learners was trained with isolated words (Word training), and another group was trained with words embedded in sentences (Sentence training).

The second purpose of the present study is to investigate whether the effect of perceptual learning differs between word training and sentence training. In particular, the present study examined the extent to which training perception of vowel
length contrasts generalizes to the perception of consonant length contrasts, and whether the extent of generalization differs between the two types of training. There are, arguably, no phonemic vowel length contrasts in modern Seoul Korean. Accordingly, it is expected that Korean listeners may be less sensitive to phonetic distinctions based on vowel length.

The paper is organized as follows. Section 2 describes the perceptual training method and the tests. Section 3 shows the results of the perceptual training study. Section 4 interprets the results and addresses the relationship between presentation context and perceptual learning effects. Finally, section 5 sums up the results and suggests future directions.

2. Methods

The present study conducted vowel length contrast training to Korean L2 learners of Japanese. The experiments consisted of three phases: pretest, five days of training, and post-test.

2.1. Participants

37 Korean native listeners participated in the study after 170 hours of Japanese learning on average. No participant had a history of hearing or speech disorders. The participants were assigned to the following two training groups at random: 1) Word training: 19 native Korean listeners. The training used stimuli spoken as isolated words. 2) Sentence training: 18 native Korean listeners.

2.2. Test stimuli and procedure

The test stimuli consisted of 30 real word pairs in isolation and the same 30 real word pairs embedded in sentences. Among the 30 pairs, 15 pairs contrasted in vowel length, and the other 15 pairs contrasted in consonant length. All stimuli were taken from the same speech database as [9]. The word pairs were also reasonably phonetically balanced.

For test stimuli contrasting in vowel length, the vowel length contrast appeared in either the word-initial syllable, as in /koi/ (love) versus /ko:i/ (kindness), or the word-final syllable, as in /kaze/ (wind) versus /kaze:/ (taxation). For test stimuli contrasting in consonant length, we tried to balance the number of words containing various geminate and singleton consonants.

The talker was one professionally trained native Japanese speaker who had been trained as voice actresses and spoke standard Tokyo Japanese comfortably. She produced each test word or sentence at three speaking rates: slow, normal, and fast.

Listeners took the test in a quiet room. The task was a single-stimulus, two-alternative forced-choice identification task. On each trial, alphabetical transcriptions of the two Japanese words comprising a minimal pair appeared as clickable buttons in the computer program window. For sentences, an English transcription of the carrier sentence was also presented, with the target word replaced by an underline. Above the presentation style considered the participants who were not able to read the Japanese character. At the same time, listeners heard one of the words, or a sentence containing one of the words, presented through headphones. Listeners’ task was to select the word they heard by clicking the appropriate button. The trials were self-paced. The test consisted of 360 trials (60 words in isolation × 3 speaking rates, 60 words in sentence × 3 speaking rates). Within each block, the stimuli were presented once each in a random order. We divided the trials into six blocks of 60 trials each. Listeners could take short breaks between blocks of trials. All the groups took the same test twice, once in the pretest and once in the post-test.

Between pretest and post-test, the training groups received identification training for long and short vowel contrast in Japanese. In contrast, the control group took the pretest and post-test only. The post-test for the control group was carried out seven days after the pretest.

2.3. Training stimuli and procedure

Both training groups took the same amount of identification training. However, each group was trained by different kinds of stimuli. The stimuli were as follows.

Word-training group was trained with stimuli spoken as isolated words (160 real isolated word × 10 sessions × 3 repetitions). The 160 words consisted of 80 real word pairs that minimally contrasted in vowel length. Each session consisted of 480 trials, broken into eight blocks of 60 trials each. Word training group took 2 sessions (960 trials) a day.

Meanwhile, sentence-training group was trained with the same 160 words as above which were embedded in carrier sentences (160 real words in sentence × 10 sessions × 3 repetitions). In each session, the training stimuli were presented once each in a random order, broken into eight blocks of 60 trials each. Sentence training group took 2 sessions (960 trials) a day.

Training trials were the same as test trials, except that listeners received feedback immediately after they chose a response. If a listener responded incorrectly, the same trial was repeated, until the listener responded correctly. Moreover, unlike the test, listeners were able to click the replay button for an unlimited number of times when they wanted to listen to the stimulus again. The training for each day took approximately 40-60 min, with a mild tendency for sessions to become shorter as listeners accumulated training. The training was conducted in the same room environment as the pretest and post-test.

3. Results

3.1. Overall effect of perceptual training

Figure 1 shows the accuracy in the pretest and post-test for word training and sentence training groups. Overall accuracy shows a similar tendency between word training and sentence training.

Firstly, for the word training group, accuracy for both vowel length contrast and consonant length contrast improved after the training. Moreover, vowel length contrast showed a higher rate of increase in accuracy compared with consonant length contrast. Accuracy for geminate and singleton consonants was 77.8% (standard deviation (SD) = 9.2) in pre-test, and 81.4% (SD = 10.8) in post-test. Accuracy for long and short vowel contrast was 78.7% (SD = 8.4) in pre-test, and 87.6% (SD = 6.3) in post-test.

Next, the sentence training group showed a similar tendency as the word training group. Accuracy for geminate and singleton consonants was 80.1% (SD = 7.3) in pre-test, and 84.5% (SD = 3.8) in post-test. Accuracy for long and short vowel contrasts was 83.4% (SD = 4.2) in pre-test, and 89.9% (SD = 3.0) in post-test.

A three-way repeated-measures ANOVA with group (word training, sentence training) as a between-subjects variable and test (pre, post-test) and contrast type (vowel, consonant) as within-subjects variables was conducted. The main effect of test [F(1,70)=63.4, p<0.01] and the test-by-contrast type interaction [F(1,70)=7.0, p<0.01] were significant. Other main effects and interactions were not significant. Further analysis of the simple main effect and
multiple comparisons using Bonferroni revealed that both word training and sentence training groups significantly improved accuracy after the training. Moreover, vowel length contrasts, which listeners were trained to identify, showed a significantly higher rate of increase in accuracy than consonant length contrasts, which listeners were not trained to identify. However, there were no systematic differences between the two training groups.

These overall results indicate that perceptual training of vowel length contrasts improves listeners’ ability to perceive vowel length contrasts. Moreover, accuracy of the consonant length contrasts, which listeners were not trained to identify, also improved after the training. These results suggest that perceptual training of vowel length contrasts generalized to consonant length contrasts. However, the rate of increase of accuracy for consonant length contrasts is significantly smaller than for vowel length contrasts. Finally, between word training and sentence training groups, the pattern of increase in accuracy from pretest to post-test did not significantly differ from each other.

### 3.2. Further analysis of the effect of presentation context types

Overall results in section 3.1 indicated that there were no significant differences between the two training groups on the amount of increase in accuracy from pretest to post-test. To address the influence of perceptual training, this section analyzes the results in more detail.

Figure 2 shows the accuracy in the pretest and post-test for both training groups, separately for word stimuli and sentence stimuli, and separately for long and short vowels (for vowel length contrasts) and singleton and geminate consonants (for consonant length contrasts). Listeners’ accuracies were submitted to a three-way repeated-measures analysis of variance (ANOVA) with group (word training, sentence training) as a between-subjects variable and test (pre, post-test) and contrast type (8 levels: singleton consonant, geminate consonant, short vowel, and long vowel for isolated words, and singleton consonant, geminate consonant, short vowel, and long vowel (for words embedded in sentences) as within-subjects variables. The main effect of test $[F(1,280)=72.2, p<0.001]$ and the test-by-contrast type interaction $[F(7,280)=6.3, p<0.001]$ were significant. Other factors were not significant. Further analysis of the simple main effect and multiple comparisons using Bonferroni revealed several differences between the word training and sentence training groups.

Firstly, for the word training group, isolated word stimuli showed a significant improvement in accuracy from pretest to post-test only for words containing long vowels ($p<0.001$). Words containing singleton and geminate consonants and short vowels did not show a significant improvement from pretest to post-test. In contrast, for sentence stimuli, accuracy significantly improved for sentences in which the target word contained singleton consonants ($p<0.05$) and long vowels ($p<0.001$).

Next, for the sentence training group, isolated word stimuli showed a significant improvement from pretest to post-test for words containing singleton consonants ($p<0.001$) and long vowels ($p<0.001$). Words containing geminate consonants and short vowels did not show a significant improvement. Sentence stimuli showed a similar tendency as isolated words. That is, sentences in which the target word contained singleton consonants ($p<0.05$) and long vowels ($p<0.001$) showed significant improvement in accuracy from pretest to post-test.

The above results suggest that it is possible that the effect of training is different depending on the training group, even though the difference is small. Specifically, the sentence training group showed slightly more generalization to untrained contrasts (consonant length contrasts) than did the word training group.

### 4. Discussion and conclusion

This section sums up our findings, and discusses the relationship between the perceptual training methods and the learning effects.

The first goal of the present study was to investigate whether word training and sentence training both improve learners’ perception of length contrasts, using the same task as Tajima et al. [4]. The results showed that overall performance significantly improved as a result of training for both word training and sentence training groups. In addition, between the training groups, there were no significant different in the amount of increase in accuracy. This result suggests that perceptual training using words in isolation and words embedded in sentences are both effective in improving L2 learners’ overall ability to perceive length contrasts in Japanese. This result agrees with Hirata [3].

The second purpose of the present study was to investigate whether the effect of perceptual learning differs between word training and sentence training. In particular, the present study examined the extent to which training perception of vowel length contrasts generalizes to the perception of consonant length contrasts, and whether the extent of generalization differs between the two types of training. Overall results (Figure 1) showed that both training groups improved performance not only for vowel length contrasts, which listeners were trained to identify, but also for consonant length contrasts, which listeners were not trained to identify. This suggests that both types of training in the present study generalized to untrained contrasts. Further analysis (Figure 2) indicated, however, that the sentence training group showed improvement from pretest to post-test in an untrained contrast type (singleton consonants) for both word and sentence stimuli, while the word training group showed improvement in an untrained contrast type for only sentence stimuli. This potentially suggests that sentence training is advantageous to word training, showing greater generalization of training to untrained conditions. These results agree with Hirata [3], and extends her findings by suggesting the possibility that sentence training generalizes to untrained contrast types. Moreover, Greenspan’s experiments [8] showed that the training with
isolated words generalized only to isolated word contexts. On the other hand, in the present study, the training with isolated words generalized not only to word contexts but also to sentence contexts. These differences should be more precisely studied in the future using the same experimental frameworks.

5. Summary and future plan
To look for more effective perceptual training schemes for L2 learners, the training on vowel length contrast of Japanese was compared between word training and sentence training. Moreover, we analyzed the generalization characteristics. Results showed that the perceptual training improved the learners’ identification accuracies both by word training and by sentence training. In addition, both trainings improved the accuracies of consonant contrasts which listeners were not trained to identify. It was also found that sentence training generalized to a wider range of conditions than word training.

As a future plan, to understand the effectiveness of perceptual training in more detail, we investigate the sentence training characteristics more in detail related to the speech rate variations and contrast types, and compare the obtained results with those of an untrained control group.

6. Acknowledgements
This work was supported in part by Grant-in-Aid for Scientific Research B, No.20300069 of JSPS and Waseda University Grant for Special Research Project 2010B-367.

7. References