



# The effect of a word embedded in a sentence and speaking rate variation on the perceptual training of geminate and singleton consonant distinction

Mee Sonu<sup>1</sup>, Keiichi Tajima<sup>2</sup>, Hiroaki Kato<sup>3</sup> and Yoshinori Sagisaka<sup>4</sup>

<sup>1</sup> GITI, Waseda University, Japan

<sup>2</sup> Department of Psychology, Hosei University, Japan

<sup>3</sup> NICT/ ATR, Japan

<sup>4</sup> GITI, Waseda University, Japan

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

## Abstract

Aiming at effective perceptual training of second language learning, we carried out training experiments on Japanese geminate consonants. Native Korean learners were trained to identify geminate and singleton stop of Japanese. Since Korean language has no phonemic contrast between long and short consonants, learners have tried to learn their differences based on their categorical perception through training. To test the training efficiency and find generalization of temporal discrimination, we investigated the perceptual training with a word embedded in sentences and single/multiple speaking rate. Training experiments showed the superiority with a word embedded in sentences and multiple speaking rates. These results suggest that perceptual training which was trained by multiple speaking rates could be effective to perceive temporal discrimination of length contrast of Japanese. However, under the training stimuli was single speaking rate condition, perceptual training have generalized to the limited extent. These results suggest that context factors including speaking rate would affect to identify the length contrast of Japanese to L2 learners.

**Index Terms:** perceptual training, singleton/geminate consonants, speaking rate variation

## 1. Introduction

We have been aiming at an effective perceptual training scheme for second language learning. Phonemic distinction resulting from timing such as geminate and singleton consonants (e.g., /ha-sa-n/ (bankruptcy) versus /ha-s-sa-n/ (emission) in Japanese) are one of the most common and serious problems not only for English speaking learners but also for many Asians learners including Korean speakers. We need to understand the perceptual characteristics of these distinctions to find scientifically relevant training methods. In this paper, we introduce our experimental training results on the identification of consonant length contrasts in Japanese by Korean speaking learners.

In Japanese, vowel and consonant length is phonemically used to distinguish words. Moreover, it has been pointed out that the primary cues are the duration of vowels and consonants (e.g., [2, 3]). It is difficult to identify consonant and vowel length contrast in Japanese for many L2 learners who have no durational contrasts in their L1. Even when learners' L1 have durational contrasts, usually, they do not completely coincide with Japanese timing characteristics controlled by many contextual factors such as mora-timing, speaking rate, phrase length and position in phrase (e.g., [5,6,9]).

There are several studies that examined the effectiveness

of perceptual training of consonant and vowel length contrasts in Japanese in relation to multiple contextual factors (e.g., [4, 8]). Tajima *et al.*, [8] investigated the extent to which native English listeners' perception of Japanese length contrasts would improve through perceptual training of vowel length contrasts. The results suggest that perceptual training improves non-native listeners' perception of Japanese length contrasts only to a limited extent.

In this paper, we report on the experimental training results for Korean learners. To examine the effect of similar training characteristics on timing perception and their extent of generalization, we focused on the perception of geminate and singleton consonants by Korean learners. Sonu *et al.* [7] investigated the extent to which Korean learners' perception of Japanese length contrasts improve through perceptual training of consonant length contrasts using isolated words. The results suggested that perceptual training could improve the identification of consonant contrasts in Japanese by Korean learners. To extend these experimental results, the present study examined the effectiveness of perceptual training using words embedded in carrier sentences. Moreover, we compared between training using a single speaking rate and training using multiple speaking rates. We expected that these training parameter changes will show the key for more efficient training than the one that we have already observed using isolated words.

In the following sections, first we introduce the methodology of the perceptual training experiment which jointly examined the effect of words embedded in sentences single vs. multiple speaking rates. Next, we analyze the results of the perceptual training. In particular, in section 3, we analyze the effects of perceptual training, as it is, and the effects of a word embedded in a sentence and speaking rate variation through appropriate procedures. Finally, in section 4, we sum up our findings and discuss relevant training methods and learners' perceptual characteristics.

## 2. Perceptual training experiment using a word embedded in a sentence and speaking rate variation

In the present study, we carried out perceptual training on Korean L2 learners of Japanese by focusing on the effect of a word embedded in a sentence and of variation in speaking rate.

In order to study the effect of a word embedded in a sentence used in the perceptual training, each group, including a control group, took an identification test that included not only words embedded in a sentence but also various untrained stimuli. In more specific terms, the learners took the identification test with isolated words as untrained stimuli. Moreover, in order to study the effects of variation in speaking

rate, the learners underwent training using sentences spoken at either a single speaking rate or multiple speaking rates.

## 2.1. Participants

Three listener groups participated. 1) Mixed trained group (Mixed-training): ten native Korean speakers who took five days of training. The training used stimuli spoken at three speaking rates: slow, normal, and fast. 2) Fixed trained group (Fixed-training): ten native Korean speakers who took five days of training. This group was trained with stimuli spoken only at a normal speaking rate. 3) Control group (Control): nine Korean speakers who took only pretest and post-test.

## 2.2. Stimuli and procedure

The experiments consisted of three phases: pretest, five days of training, and post-test. Mixed-training and Fixed-training participated in all three phases. Control participated in pretest and post-test only. Before the experiments, all participants were given a brief description of Japanese length contrasts which was given in Korean. In addition, the length contrasts in Japanese were written as follows. Long vowels were transcribed as “a: i: u: e: o:”. Long consonants were transcribed as “pp tt kk ss zz jj” or as “ssh tch”.

## 2.3. Testing 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 real word pairs, 15 pairs contrasted in vowel length, and the other 15 pairs contrasted in consonant length. Similarly, of the 30 real word pairs which were embedded in sentences, 15 pairs contrasted in vowel length, and the other 15 pairs contrasted in consonant length. All word pairs were selected based on a Japanese lexical database [1]. 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 target consonant was one of the following voiceless consonants: “t”, “p”, “k”, “s”, “sh”, “ch”.

All the stimuli were taken from the same speech database that was used in [8] (see [8] for details). The talkers were four professionally trained native Japanese talkers who had been trained as voice actors or actresses and spoke standard Tokyo Japanese comfortably. They produced each test word 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 two Japanese words comprising a minimal pair appeared as clickable buttons in the computer program window. 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  $\times$  3 rates, 60 words in sentence  $\times$  3 rates). 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, Mixed-training and Fixed-training received identification training for geminate and singleton consonants in Japanese. In contrast, Control group took the test only. The post-test for control group was carried out seven days after the pretest.

## 2.4. Training procedure

The training stimuli consisted of 60 real word minimal pairs contrasting in geminate and singleton consonants, which were embedded in a sentence. The words were different from those used in the test. We also tried to balance the number of words containing various geminate and singleton consonants. Mixed-training and Fixed-training were presented with different stimuli.

Mixed-training was trained with stimuli spoken at three speaking rates (120 real words in sentence  $\times$  3 speaking rates (slow-normal-fast)  $\times$  10 sessions).

Meanwhile, Fixed-training was trained with stimuli spoken only at a normal speaking rate (120 real words  $\times$  3 repetitions  $\times$  10 sessions). Each session consisted of 360 trials, broken into six blocks of 60 trials each. Both training groups (Mixed-training and Fixed-training) took 2 sessions (720 trials) a day.

In each session, the training stimuli were presented once each in a random order. 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. Mixed-training and Fixed-training underwent five days of perceptual training between pretest and post-test. 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 of the perceptual training

In subsequent sections, we report the results of the perceptual training by comparing identification accuracies in the pretest and post-test. Additionally, all statistical tests were conducted on arcsine transformed values of the identification accuracies.

### 3.1 Overall performance

Figure 1 indicates box-plots of the identification accuracies in pretest and posttest for Mixed-training, Fixed-training and Control.

First, mean accuracy for the Mixed-training was 81.4% (standard deviation (SD)=7.3) in pretest, but rose to 87.1% (SD=4.9) in post-test. Next, mean accuracy for the Fixed-training was 82.1% (SD=6.9) in pretest, but rose to 85.9% (SD=5.6) in post-test as well. Last, control group mean accuracy was 82.1% (SD=5.7) in pretest and 83.6% (SD=7.4) in post-test.

The results of listeners’ accuracies were submitted to repeated measures analyses of variance (ANOVA), with test (pretest, post-test) as a within-subjects variable and group (Mixed-training, Fixed-training and Control) as a between-subjects variable. Results revealed a significant main effect of test [ $F(1,25)=33.3, p<.001$ ]. The main effect of group was not significant [ $F(2,25)=0.149, n.s.$ ]. There was no significant interaction between group and test [ $F(2,25)=3.018, n.s.$ ].

These results indicated that the perception of the distinction between singleton and geminate consonants in Japanese improved from pretest to post-test, but the effectiveness of the perceptual training was not significantly different compared with the control group.

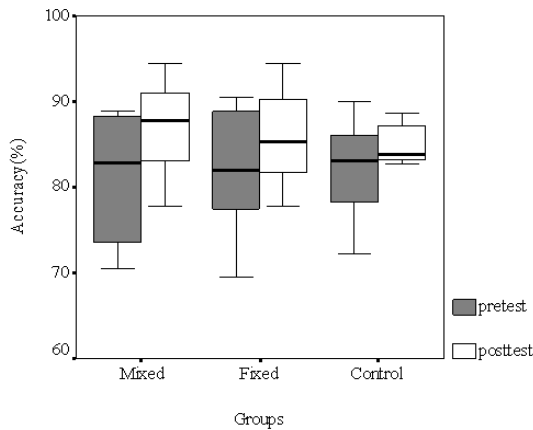


Fig1. Identification accuracies as a function of listener group (Mixed-training, Fixed-training, and Control) and test phase (pretest and post-test)

### 3.2 The effect of a word embedded in a sentence

Figure 2 shows the accuracies separately for isolated words and words embedded in sentences, for the two training groups. Figure 2(a) is the group which was trained with three different speaking rates and Figure 2(b) is the group which was trained with one speaking rate. The results were as follows.

First, mean accuracy for isolated words for the Mixed-training was 80.8% (SD=8.5) in pretest, but rose to 85.6% (SD=5.0) in post-test. Similarly, mean accuracy for words embedded in sentences for the Mixed-training was 82.0% (SD=6.7) in pretest, but rose to 88.5% (SD=5.3) in post-test.

However, for Fixed-training, there was a different tendency compared with Mixed-training. Mean accuracy for isolated words was 81.6% (SD=7.1) in pretest, and rose to only 82.9% (SD=7.2) in post-test. In comparison, mean accuracy for words embedded in sentences was 82.7% (SD=7.5) in pretest, but rose to 88.9% (SD=4.8) in post-test.

The results of listeners' accuracies were submitted to separate repeated measures ANOVAs for words and sentences, with test (pre and post-test) as a within-subjects variable and group (Mixed-training, Fixed-training) as a between-subjects variable.

Results revealed significant main effects of test for isolated words [ $F(1,18)=6.44, p<.05$ ] and test for sentences [ $F(1,18)=86.9, p<.001$ ]. However, main effect of group for isolated words was not significant [ $F(1,18)=0.95, n.s.$ ], and the main effect of group for sentences was not significant either [ $F(1,18)=0.49, n.s.$ ]. None of the interactions was significant.

These results indicate that perceptual training using words embedded in a sentence is effective for improving perception of consonant length contrasts in Japanese by Korean speaking learners. Mixed-training groups' posttest mean accuracies significantly rose from the pretest for both words and sentences. In contrast, Fixed-training group's post-test mean accuracies rose to a limited extent. There were difference tendencies in the identification accuracies depending on contextual factor, although there was no statistical difference.

These results show that perceptual training with a word embedded in a sentence could affect performance to a limited extent. However, these results do not provide strong evidence of effective perceptual training methods using a word embedded in a sentence.

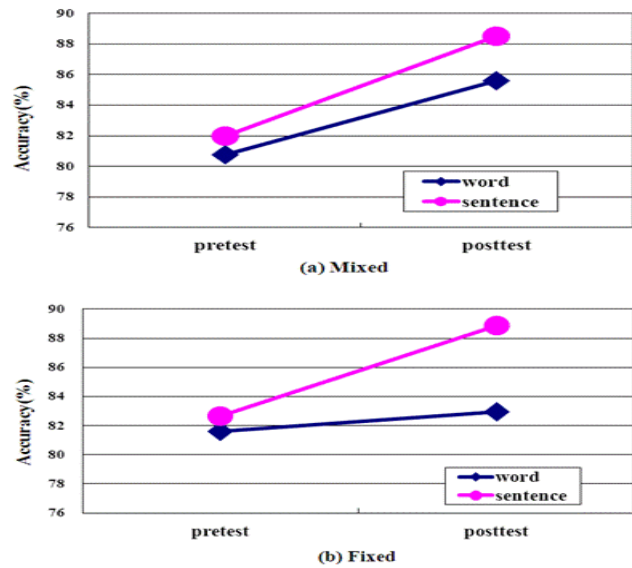


Fig2. Identification improvements after training ((a) Multiple speaking-rates training (b) Single speaking-rate training, ●: an isolated word. ◆: a word embedded in sentence)

### 3.3 The effect of single/multiple speaking rates

Figure 3(a) shows identification accuracies as a function of speaking rate for isolated words. For Mixed-training, mean accuracy in isolated words which were spoken at a slow rate was 80.5% (SD=10.5) in pretest, and 86.8% (SD=4.4) in post-test. Mean accuracy of words in normal speaking rate for the Mixed-training was 85.5% (SD=7.6) in pretest, and rose to 89.2% (SD=7.0) in post-test. Mean accuracy of words in fast speaking rate for the Mixed-training was 76.3% (SD=12.1) in pretest, but rose to 80.8% (SD=6.7) in post-test.

By contrast, for Fixed-training mean accuracy for words in isolation in slow speaking rate was 80.5% (SD=7.5) in pretest, and 83.2% (SD=7.5) in post-test. And mean accuracy for words in normal speaking rate was 86.3% (SD=7.9) in pretest, and 88.2% (SD=6.3) in post-test. Finally, words in fast speaking rate was approximately equal between pretest (77.5% (SD=8.3)) and post-test (77.5% (SD=9.8)).

Figure 3(b) shows identification accuracies for words embedded in sentences as a function of speaking rate. First, for Mixed-training mean accuracy in slow speaking rate was 84.0% (SD=9.3) in pretest, and rose to 90.1% (SD=5.5) in post-test. Similarly, in normal speaking rate, mean accuracy was 86.0% (SD=5.0) in pretest, and rose to 89.7% (SD=6.1) in post-test. Finally, accuracy in fast speaking rate was 76.0% (SD=8.0) in pretest, and markedly rose to 85.0% (SD=7.9) in post-test.

Similarly, Fixed-training also shows increases in accuracy from pretest to post-test. Mean accuracies for slow speaking rate was 85.7% (SD=8.7) in pretest, and rose to 91.7% (SD=4.6) in post-test. Mean accuracies for normal speaking rate was 84.5% (SD=10.4) in pretest, and rose to 91.2% (SD=5.5) in post-test. Last, mean accuracies for fast speaking rate was 77.8% (SD=6.1) in pretest, and rose to 83.8% (SD=7.0) in post-test.

Together, Mixed-training showed an increase in identification accuracies from pretest to post-test for all speaking rates. Meanwhile, Fixed-training showed only a relatively small increase in identification accuracies in the post-test. These results mean that multiple speaking rates leads to greater generalization of training than a single speaking rate.

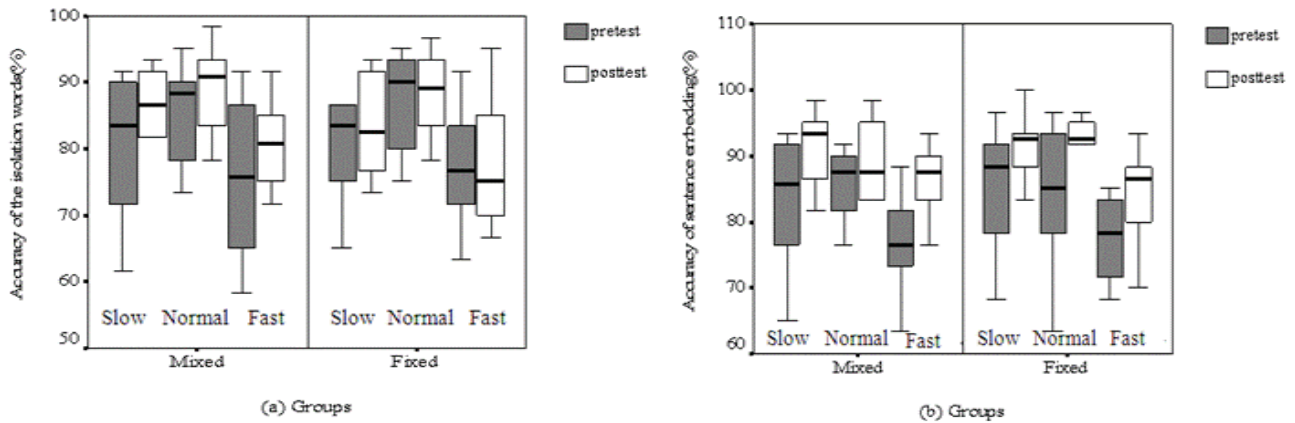


Fig3. Identification accuracies of (a) a isolated words contrast and (b) a word embedded in sentences as a function of listener groups (Mixed-training and Fixed-training) and test phrase speaking rates (Slow-Normal-Fast).

#### 4. General discussion and conclusion

In this section, we sum up our findings, and discuss relevant training methods and the learners' characteristics of perception.

First, perceptual training, as it is, could help the learners identify singleton and geminate consonants in Japanese, irrespective of the context in which they are presented as well as the speaking rate. In this experiment, Mixed-training and Fixed-training groups had significantly higher scores in the post-test than in the pretest. We assumed that the training groups were repeatedly exposed to stimuli under various conditions in the pre and posttests, thus leading to the overall higher rate of accuracy in the posttest. However, the results of these two groups were not significantly different from those of the control group.

Subsequently, we found that training using words embedded in sentences and multiple speaking rates led to robust generalization in Korean speaking learners' identification of Japanese length contrasts. All training groups showed generalization in the sentence context. However, training generalized to a different extent, depending on the training stimuli.

Specifically, the Mixed-training group, which was trained with three speaking rates (slow-normal-fast), could generalize to a wider extent than the Fixed-training group, which was trained with only a single speaking rate. These results suggest that it is crucial for L2 learners to be exposed to high-variability stimuli in order to perceive the contrast between single and geminate consonants in Japanese.

Overall, perceptual training might be more effective when the training word embedded in the sentence is presented in multiple speaking rates rather than only in a single speaking rate.

To obtain the above results, we suggest that the stimuli be used repeatedly with multiple speaking rates so that the embedded sentences are effective in facilitating temporal distinction of Japanese.

#### 5. Summary and future works

In this paper, in order to provide effective perceptual training to L2 learners of Japanese, we have introduced our training experimental results in Japanese.

We observed the following results. First, perceptual training could improve the learners' ability to identify consonant length contrasts in Japanese, irrespective of contextual factors and despite the limited extent of learning.

Next, depending on the training conditions, the extent of generalization of learning differed. Training experiments showed that carrier sentences and multiple speaking rates had an advantage.

In the future, it is necessary to investigate the perceptual training of vowel length contrasts for Korean speaking learners. As contextual factors, type of length contrast might be a crucial factor for effective perceptual training. Such future work might be able to provide hints for developing relevant perceptual training methods for second language learning.

#### 6. Acknowledgements

This work was supported in part by Grant-in-Aid for Scientific Research B, No.20300069 of JSPS and No.2105072 of JSPS.

#### 7. References

- [1] Amano, S. and Kondo, T. "Nihongo-no Goi-Tokusei (Lexical Properties of Japanese)", Vol.1, Sanseido, Tokyo, 1999.
- [2] Fukui, S., "Perception for the Japanese stop consonants with reduced and extended durations", The Bulletin of the Phonetic Society of Japan, 159:9-12, 1978 (in Japanese).
- [3] Fujisaki, H., Nakamura, K., and Imoto, T., "Auditory perception of duration of speech and non-speech stimuli", in G. Fant and M. Tatham (eds.), Auditory Analysis and Perception of Speech, 197-219, Academic Press, 1975.
- [4] Hirata, Y., Whitehurst, E., and Cullings, E., "Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts", J. Acoust. Soc. Am., 116:2384-2394, 2007.
- [5] Min, K., "Perception of the geminate consonants of Japanese by Korean learners", Jpn. Lang. Edu., 62:179-193, 1987 (in Japanese).
- [6] Sagisaka, Y. and Tohkura, Y., "Phoneme duration control for speech synthesis by rule", Trans. Inst. Electr. Inf. Commun. Eng., J67-A: 629-636, 1984.
- [7] Sonu, M., Tajima, K., Kato, H., and Sagisaka Y., "Perceptual training of singleton and geminate stops in Japanese language by Korean learners", in Proceedings of Interspeech 2009, 1747-1750
- [8] Tajima, K., Kato, H., Rothwell, A., A-Yamada, R., and Munhall, K., "Training English listeners to perceive phonemic length contrasts in Japanese", J. Acoust. Soc. Am., 123:397-413, 2008.
- [9] Toda, T., Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese, University Press of America, Lanham, MD, 2003.