INTERSPEECH 2009 BRIGHTON

Perceptual training of singleton and geminate stops in Japanese language by Korean learners

Mee Sonu[1], Keiichi Tajima[2], Hiroaki Kato[3] and Yoshinori Sagisaka[4]

[1] GITI, University of Waseda, Japan
[2] Department of Psychology, University of Hosei, Japan
[3] NICT/ATR, Japan
[4] GITI, University of Waseda, Japan

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

10.21437/Interspeech.2009-525

Abstract

We aim to build up an effective perceptual training paradigm toward a computer-assisted language learning (CALL) system for second language. This study investigated the effectiveness of the perceptual training on Korean-speaking learners of Japanese in the distinction between geminate and singleton stops of Japanese. The training consisted of identification of geminate and singleton stops with feedback. We investigated whether training improves the learners' identification of the geminate and singleton stops in Japanese. Moreover, we examined how perceptual training is affected by factors that influence speaking rate. Results were as follows. Participants who underwent perceptual training improved overall performance to a greater extent than untrained control participants. However, there was no significant difference between the group that was trained with three speaking rates and the group that was trained with normal rate only.

Index Terms: singleton/geminate stops, perceptual training, learners of Japanese

1. Introduction

We have purpose to make clear how to acquire the Japanese length contrast by second language learners. In addition, we aim to build up an effective perceptual training paradigm toward a CALL system for second language.

As a fist step toward the way of perceptual training of Japanese length contrast, this study tested the effectiveness of the perceptual training on Korean learners of Japanese in their identification of singleton and geminate stops in Japanese.

In Japanese, vowel and consonants length can be used phonemically to signal different words. For example, the vowel length distinction is phonemic in Japanese; e.g., /hodo:/ (pedestrian) versus /ho:do:/ (broadcasting). Similarly, the consonant length distinction is phonemic in Japanese; e.g., /iso/ (beach) versus /isso/ (rather) and /haken/ (dispatch) versus /hakken/ (discovery).

The primary acoustic cue of the length contrast in both vowels and consonants are said to be the duration of the vowel or consonant (e.g., [1, 2]).

However, the duration of the vowel and consonant are determined by many factors, including neighboring segments, speaking rate, and other acoustical factors; phonemic length is not determined in absolute terms but rather in relative terms (e.g., [3, 4]). Such a complex characteristic of the length contrast makes it difficult to acquire by second-language (L2) learners (e.g., [5–7]).

Particularly, the geminate and singleton stops of Japanese are difficult to identify as a length contrast by Korean-speaking learners [7]. In Korean, consonants are not divided into voiceless and voiced as in Japanese. Korean is not a voicing language but an aspirating language.

Specifically, Korean has three types of consonants. For example, /ita/ (be) versus /it^{h}a/ (later) versus /itt^{h}a/ (there -is). Johnson and Oh [8] contend that intervocalic tense consonants in Korean that are phonologically geminates are more compressible than singletons. They are called geminate reduction.

Moreover, the closure of the lenis consonant is shorter than the closure in the bare tense consonants, which in turn is shorter than the closure duration in the geminate tense consonants. Accordingly, it is possible to use perceptual cues in a different way between Korean-speaking learners and native Japanese speakers when they perceive the consonant duration. That is, it might be more difficult for Korean-speaking learners to identify the geminate and singleton stops than speakers of other languages like English that have no phonemic distinction using consonant durations.

The effects of perceptual training in terms of Japanese length contrasts have been investigated for English-speaking learners (e.g., [9–12]). However, it has not been thoroughly investigated, and also we do not know whether or not English native speakers and Korean native speakers show the same progress during the perceptual training.

Consequently, the present study addresses the following questions.

1) Is the perceptual training of the identification of singleton and geminate stops in Japanese effective to distinguish the length contrast by Korean-speaking learners? If it is true, do the results of test coincide with those observed for English-speaking learners?
2) Is the effectiveness of the perceptual training different among presentation stimuli? Hirata et al. [12] showed that training with two rates was more effective than training with one rate. Namely, Hirata et al. [12] said that the more variability in the rates of training stimuli, the more effective for non-native speakers' perceptual learning. Hirata et al. [12] and Pisoni and Lively [13] showed that phonetic variability provided a useful guide to identify Japanese vowel length. If it is true, the group trained by multiple speaking rates might achieve a higher perceptual accuracy than the group trained by a static speaking rate.

To answer these questions, we carried out perceptual training to distinguish the geminate and singleton stops in Japanese by Korean learners.

2. Methods

2.1. Participants

Three listener groups participated. 1) Group KT1 (Korean Training 1): ten native Korean learners of Japanese who took five days of training. The training used stimuli spoken at three speaking rates: slow, normal, and fast. 2) Group KT2 (Korean Training 2): eleven native Korean learners of Japanese who took five days of training. This group was trained with stimuli spoken only at a normal speaking rate. 3) Group KC (Korean Control): nine Korean Japanese learners 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. Group KT1 and Group KT2 participated in all tree phases. Group KC 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 mm nn jj" or as "ssh tch".

2.3. Testing procedure

The test stimuli consisted of 30 real word pairs. Among the 30 real word pairs, 15 pairs contrasted in vowel length, and the other 15 pairs contrasted in consonant length. The word pairs were selected based on a Japanese lexical database [14]. 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 stimuli were taken from the same speech database that was used in [9] (see [9] for details). The talkers were four professionally trained native speakers of Japanese 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, English transcriptions of two Japanese words comprising a minimal pair appeared as clickable buttons in the computer program window (see Fig.1). Listeners' task was to select the word they heard by clicking the appropriate button. In the test, listeners were allowed to listen to the stimulus only once. The trials were self-paced. The test consisted of 180 trials (60 words x 3 rates). The stimuli were presented once each in a random order. We divided the trials into six blocks of 30 trials each. Listeners could take short breaks between blocks of trials.

All groups (KT1, KT2 and KC) took the same test twice in the pretest and post-test. Between pretest and post-test, groups KT1 and KT2 received identification training for geminate and singleton stops of Japanese. In contrast, group KC took the test only. The post-test for group KC was carried out seven days after the pretest.

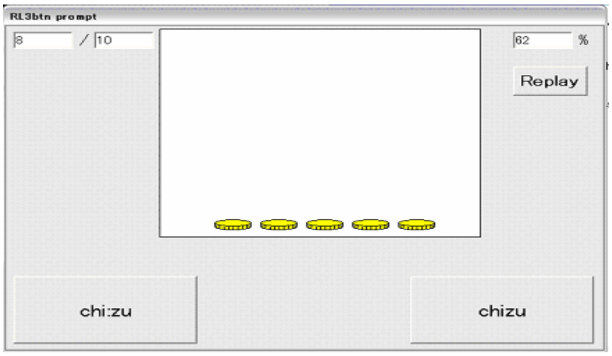

Figure 1: *The example of the perceptual training. When the listener takes the test, there is no replay button.*

2.4. Training procedure

The training stimuli consisted of 60 real word minimal pairs contrasting in geminate and singleton stops. We also tried to balance the number of words containing various geminate and singleton consonants. Groups KT1 and KT2 were presented with different stimuli. Group KT1 was trained with stimuli spoken at three speaking rates (120 real words x 3 speaking rates (fast-normal-slow) x 10 sessions). Group KT2 was trained with stimuli spoken only at a normal speaking rate (120 real words x 3 repetitions x 10 sessions).

Each session was broken into six blocks of 30 trials each. Both training groups (KT1 and KT2) 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.

Groups KT1 and KT2 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.

2.5. Analysis

Listeners' performance is reported as percent-correct identification accuracies. All statistical tests were conducted on arsine transformed values of the identification accuracies. Repeated-measures analyses of variance (ANOVA) were carried out to verify the effectiveness of the training.

First of all, we analyzed whether the two groups that underwent perceptual training (KT1 and KT2) improved identification of geminate and singleton consonants or not compared with the control group (KC).

Next, if the perceptual training was effective, we analyzed whether the effectiveness of training was different depending on presentation context. In the present study, we focused on speaking rate.

3. Results

3.1. Overall performance

In the statistical analyses of this subsection, the two trained groups KT1 and KT2 were merged into a single trained group, which was referred to as KT. Figure 2 shows boxplots of the identification accuracies in pretest and post-test for groups KT1, KT2, and KC. Mean accuracy for the trained groups (KT1 and KT2) was 80.2% (standard deviation (SD)=6.3) in pretest, but rose to 84.0% (SD=5.5) in post-test. Meanwhile, control group mean accuracy was 81.4% (SD=8.2) in pretest and 81.9% (SD=8.4) in post-test.

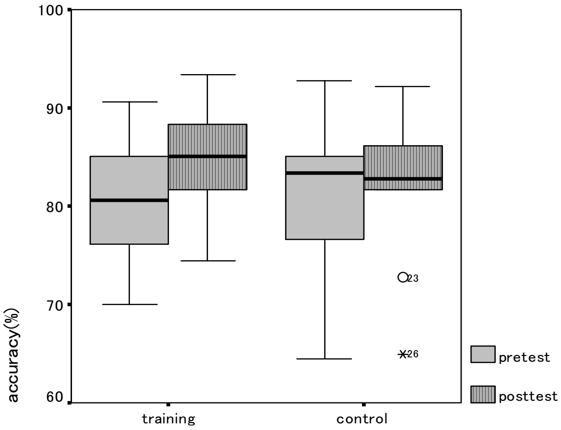

Figure 2: *Identification accuracies as a function of listener group (KT and KC) and test phase (pretest and post-test).*

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 (trained, control) as a between-subjects variable. Results revealed a significant main effect of test [F(1, 28)=10.92, p<.01]. The main effect of group was not significant [F(1, 28) =0.05]. A significant interaction was found between test and group [F(1, 28)=6.82, p<.05].

These results mean that trained listeners significantly improved the identification of the length contrast of Japanese from the pretest to the post-test compared with control group.

These results indicate that the perceptual training is effective in improving identification of length contrasts in Japanese by Korean learners.

3.2. Effect of the speaking rate

In this subsection, we took a closer look at the effect of training by analyzing the two trained groups separately. Figure 3 shows boxplots of the identification accuracies in pretest and post-test for group KT1 which was trained with stimuli produced at three speaking rates and group KT2 which was trained with stimuli spoken only at a normal rate.

For group KT1, mean identification accuracy was 78.5% (SD=7.2) in pretest, but rose to 82.3% (SD=6.5) in post-test. For group KT2, mean identification accuracy was 81.7% (SD=5.2) in pretest, but rose to 86.8% (SD=3.2) in post-test.

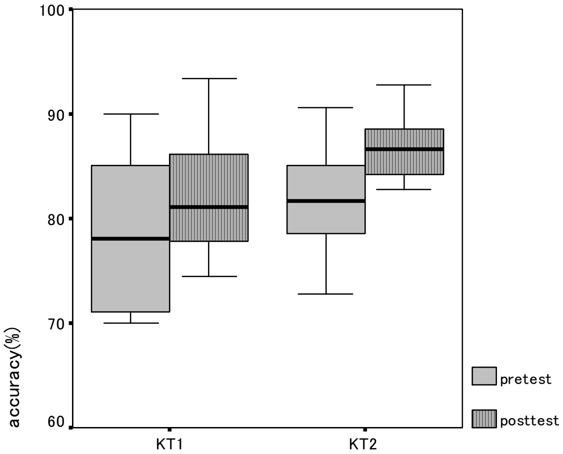

Figure 3: *Identification accuracies for group KT1's pretest and post-test and group KT2's pretest and post-test.*

The results of listeners' accuracies were submitted to repeated measures analyses of variance (ANOVA) with group (KT1, KT2) as a between subjects variable and test (pretest and post-test) as a within-subjects variable. If the perceptual training that incorporated speech rate variation was more effectively than training that used only normal-rate stimuli, results might show a significant interaction between group and test. Results were revealed a significant main effect of test [F(1, 19)=22.12, p<.001]. The main effect of group was not significant [F(1, 19)=2.35, n.s.]. No significant interaction was found between test and group [F(1, 19)=0.64, n.s.].

These results indicate that Korean-speaking learners in general improved their identification of Japanese length contrasts while they were not significantly affected by the difference in variation of trained speaking rate.

4. Discussion and Conclusion

First of all, we investigated the effectiveness of the perceptual training. The results revealed that the groups which took training significantly improved the accuracy of the identification of the geminate consonants. These results coincide with Tajima et al. [9]. Tajima et al. [9] showed that English-speaking learners also improved their accuracy by taking a similar perceptual training. However, Tajima et al. [9] trained the learners by vowel length contrast, whereas the present study did by consonant length contrast. Therefore, it has to be verified whether or not the training by vowel length is also effective for Korean-speaking learners.

Next, we investigated the effectiveness of variations in the condition of perceptual training. We introduced the variability of speaking rate in the training stimuli. We predicted that the training with multiple rates would improve the perceptual accuracy more than that with a single (normal) rate. However, our results were not the case, which do not coincide with the case in Hirata et al. [12]. Namely, both KT1 and KT2 achieved significantly higher scores in the post-test than in the pretest.

Moreover, the group KT2, which was trained only normal rate, tended to have a higher accuracy than the group KT1, which was trained by three different rates. The reason might relate to the participants' proficiency level of Japanese. In this study, we did not control the participants' proficiency level.

However, the beginner-level participants tended to have difficulty in adjusting themselves with the multiple-rate training in group KT1. Some of the beginner learners in group KT1 got lower post-test scores than their pretest scores. Meanwhile, all beginner learners in group KT2 improved their accuracy after the training. Thus, it needs to further examine the effect of stimulus variability in terms of learners' proficiency levels.

In conclusion, Korean-speaking learners who were trained to identify singleton and geminate stops in Japanese show overall improvement in performance in comparison with the control group. In addition, the way of training has to be considered with not only conditions of the stimuli but also individual factors of learners. Further research is needed to refine a condition of the stimuli considering individual factors.

5. Acknowledgements

This work was supported in part by Waseda University RISE research project of "Analysis and modeling of human mechanism in speech and language processing" and the Grant-in-Aid for Scientific Research (B) No.20300069, JSPS.

6. References

[1] 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).

[2] 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.

[3] Amano, S. and Hirata, Y., "Correspondence of perception and production boundaries between single and geminate stops in Japanese", in Proc. Interspeech, 2869–2872, 2008.

[4] Sagisaka, Y. and Tohkura, Y., "Phoneme duration control for speech synthesis by rule", Trans. Inst. Elecrton. Inf. Commun. Eng., J67-A:629–636, 1984.

[5] Toda, T., Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese, University Press of America, Lanham, MD, 2003.

[6] Uchida, T., "Categorical perception of relatively stead-state speech sound duration in Japanese moraic phoneme", J. Phonet. Soc. Jpn., 2:71–86, 1998 (in Japanese).

[7] Min, K., "Perception of the geminate consonants of Japanese by Korean learners", Jpn. Lang. Edu., 62:179–193, 1987 (in Japanese).

[8] Johnson, K. and Oh, M., "Intervocalic consonant sequences in Korean", Ohio State Univ. Working Papers in Linguistics, 45:85–97, 1995.

[9] 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.

[10] Hirata, Y., "Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts", Comp. Assist. Lang. Learn., 17:357–376, 2004.

[11] Hirata, Y. and Whiton, J., "Effects of speaking rate on the single/geminate stop distinction in Japanese", J. Acoust. Soc. Am., 118:1647–1660, 2005.

[12] 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.

[13] Pisoni, D. B. and Lively, S. E., "Variability and invariance in speech perception: A new look at some old problems in perceptual learning", in W. Strange (eds.), Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research, 433–459 , York, Timonium, MD, 1995.

[14] Amano, S. and Kondo, T. "*Nihongo-no Goi-Tokusei* (Lexical Properties of Japanese)", Vol.1, Sanseido, Tokyo, 1999.

[15] Min, K., "The cause of the occurrence of geminates insertion: Evidence from Korean learners' production of the Japanese voiceless stops as geminates", J. Phonet. Soc. Jpn., 11:58–70, 2007 (in Japanese).