Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Education in Acoustics
Session 3aED: Learning by Listening: Education in Acoustics Based on Listening (Lecture/Poster Session)

3aED9. Effect of speaking rate variation on the perception of singleton and geminate consonants in Japanese by native and Korean listeners

Mee Sonu*, Takayuki Arai, Keiichi Tajima and Hiroaki Kato

***Corresponding author's address: Faculty of Science and Technology, Sophia University, Kiyoicho, 102-8553, Tokyo, Japan, sonumeephonetic@gmail.com**

Perception of phonemic length contrasts in Japanese is difficult for non-native listeners. To better understand the source of this difficulty, the present study investigated native Korean listeners' perception of consonant length contrasts at different speaking rates. Stimuli were created by modifying the duration of the second consonant of /ereC:e/ along a continuum to /ereCe/, where C was /k/ or /s/. The base words were spoken by a professionally trained native Japanese speaker with a carrier sentence at three rates, fast, normal, slow. Twenty four native Korean and twelve native Japanese listeners participated in a perception test. They listened to one of the created stimuli and identified whether the second consonant was singleton or geminate. Results show that even though Korean listeners' perceptual boundary position between singleton and geminate consonants shifted according to speaking rate in a similar manner as the natives, their boundary width was more variable and larger than native listeners at all speaking rates. These results suggest that Korean listeners have ambiguous criteria for phonemic length contrasts. Results discussed that the perceptual similarity between intervocalic consonant of Korean and geminate consonant of Japanese.

Published by the Acoustical Society of America through the American Institute of Physics

Received 8 Jan 2013; published 2 Jun 2013

1. INTRODUCTION

The ultimate aim of the present study is to understand non-native listeners' perceptual characteristics of the Japanese phonemic length contrast. Japanese phonemic length contrast is difficult for non-native listeners to perceive whose first-language (L1) do not convey lexical information by duration alone. To accomplish this purpose, it is necessary to consider not only the characteristics of Japanese phonemic length contrasts but also second-language (L2) learners' acquisition process of these length contrasts.

In the aspect of the characteristics of Japanese phonemic length contrasts, Japanese language uses length (segmental duration) contrast of vowels and consonants for phonemic distinctions (Fujisaki *et al*. 1975). Moreover, the phonemic length contrasts are not fixed but varies with contextual factors, such as speaking rate, present context, and neighboring segments (Sagisaka and Tohkura, 1984). Thus, in the learning of the Japanese language, the length contrast of Japanese is one of the most difficult training items for L2 learners to acquire (Hirata *et al*. 2007, Tajima *et al*. 2008). To look at the more detail understanding of non-native listeners' criterion to identify the phonemic length contrast of Japanese, Wilson *et al*. (2005) and Kato *et al*. (2004) investigated non-native listeners' perception using synthesized stimulus continuum of a pair of Japanese vowel length. Wilson *et al*. (2005) investigated the influence of speaking rate variations in relation to non-native listeners' perception. The results revealed that native English listeners used fixed-length criterion to identify the Japanese vowel length contrast, because English listeners are unable to identify the Japanese vowel length contrast adapting to the speaking rate variation. Kato *et al*. (2004) reported that there were remarkable and significant difference between English listeners and Japanese native listeners not only in the position of criterion but also in the ambiguity of identification. From the results of previous studies, the speaking rates are one of the indexes to understand the source of difficulty to acquire the phonemic length contrast of Japanese by non-native listeners.

Next, Kinoshita (2011) and Toda (2003) investigated the acquisition orders in relation to the type of consonant (plosive or fricative) in the aspect of the acquisition orders of perception of Japanese length contrast by non-native listeners. Kinoshita (2011) indicated that the plosive consonant contrast was difficult for Korean listeners to acquire than the fricative consonant contrast of Japanese, while Toda (2003) indicated that the fricative consonant contrast was difficult for English listeners to acquire than the plosive consonant contrast. On the basis of the results of those studies, it is presumed that the non-native listeners' perceptual characteristics are influenced by non-native listeners' L1.

In sum, there were two perspectives of the perceptual characteristics of non-native listeners. One is the influence of the speaking rate which is a language-independent factor. The other is the interference of non-native listeners' L1 which is a language-dependent factor. However, Kinoshita (2011) and Toda (2003) have different experimental conditions, and is difficult to compare the two studies directly. The relationship between the influence of the speaking rate is also unclear.

One of the approaches in the present study is to investigate the perception of consonant length contrast with speaking rate variations by Korean native listeners using synthesized stimulus continuum particularly on the boundary position and width as similar methodology with Wilson *et al*.(2005) and Kato *et al*.(2004). In addition, the experimental task of consonant length contrast includes two types of consonants, plosive or fricative. Our predictions are as follows. If the speaking rate is more reliable to identify the consonant length contrast by Korean listeners, it would have similar tendency to the results of Wilson *et al*.(2005) and Kato *et al*.(2004) in which non-native listeners use the fixed-length criterion to identify the length contrast of Japanese. On the other hand, if the influence of listeners' L1 is more reliable to identify the consonant length contrast, it would have similar results as Kinoshita (2011) in which Korean listeners have more difficulty to identify the plosive consonant contrast rather than to identify the fricative consonant contrast.

2. EXPERIMENT: CATEGORICAL PERCEPTION PERFORMANCE

2.1 Methods

The present study examined the categorical perception performance of Japanese native and Korean native listeners in categorizing the consonant length-based phonemic contrast at different speaking rates. Stimuli were created by modifying the duration of the second consonant of /ereC:e/ along a continuum to /ereCe/, where C was /k/ or /s/ using PRAAT (2012). These words were chosen as considering of an ideal pair that differs in phonemic length alone. First, these words have three or for moras, which constitute the minimum structure required to avoid

TABLE 1. Stimuli of categorical perception test

Stimuli	Carrier sentence	Extent of manipulation (ms)
/erek:e/-/ereke/	/korekara ___ to kakimasu/ (I will write _____ now)	Fast 20 - 100 (10 ms /step)
		Normal 40 - 140 (10 ms /step)
		Slow 60- 200 (20 ms /step)
/eres:e/-/erese/		Fast 60-110 (10 ms /step)
		Normal 60-150 (10 ms/step)
		Slow 70-210 (20 ms/step)

placing the length contrast in the initial or final syllable. Next, the vowel /e/ was chosen for avoiding the influencing of frequency words in given context because it does not result in real words; the consonants /k/ and /s/ do not change. allophonically when they precede in vowel, and are rarely devoiced. Finally, the stimuli were chosen /erek:e/-/ereke/ and /eres:e/-/erese/ by considering above the reasons.

The base material was chosen from a database (Tajima et al. 2008). The stimuli were spoken at three speaking rates (fast, normal, slow) by a professionally trained native Japanese voice actor who spoke standard Tokyo Japanese comfortably and recorded in a sound-proof booth. All stimuli were produced embedded in a carrier sentence, /kore kara ____ to kakimasu/ (I will write _______ now). The carrier sentence contained four moras preceding the target word, and five moras following the target word. A continuum gradually varying between /ereC:e/ and /ereCe/ was modified to provide stimuli in length by removing part of the second consonant (TABLE 1). Accordingly, in terms of /erek:e/ - /ereke/, 9 steps of stimuli at the fast speaking rate, 11 steps at the normal speaking rate, and 8 steps at the slow speaking rate were modified by removing part of the closure duration of the second consonant of /erek:e/. All stimuli had 10 repetitions. In total, one session consisted of 280 trials ([9 steps at fast + 11 steps at normal + 8 steps at slow] ×10 repetitions). In terms of /eres:e/ - /erese/, 6 steps of stimuli at the fast speaking rate, 10 steps at the normal speaking rate, and 8 steps at the slow speaking rate were modified by removing the second consonant of /eres:e/. All stimuli had 10 repetitions. In total, one session consisted of 240 trials ([6 steps at fast + 10 steps at normal + 8 steps at slow] ×10 repetitions). In each session, the stimuli were presented in a random order. Listeners sat in front of a laptop computer and heard the stimuli through headphones at a comfortable listening level. In each trial, they listened to a stimulus and were asked to identify it by choosing one of two words that contrasted in the length of the second consonant (singleton or geminate).

2.3 Modeling

Categorical perception characteristics were represented by two measures (McMurray and Spivey 1999, Kato et al. 2004). These two indexes defined as a boundary position and a boundary width which is based on a mathematical modeling of listeners' responses. The present study attempted to model the listeners' response by using the logistic function formulated as

$$y = 1/(1+\exp(-b_0 - b_1 x)) \quad (1)$$

where x is the change in consonant duration and y is the proportion of the geminate consonants responses; b_0 and b_1 are coefficient of x. This modeling provides the following indexes that represent different aspects of listeners'perceptual characteristics (see also Fig. 1).

(a) Boundary Position: x-intercept where y is 0.5, i.e., $f^{-1}(0.5)$. This index represents the listener's categorical boundary between short and long phonemes.

(b) Boundary Width: This index represents the listener's sharpness of identification. The present study decided to measure the interval from x-intercept where y is 0.25 to x-intercept where y is 0.75 , i.e., $f^{-1}(0.75) - f^{-1}(0.25)$

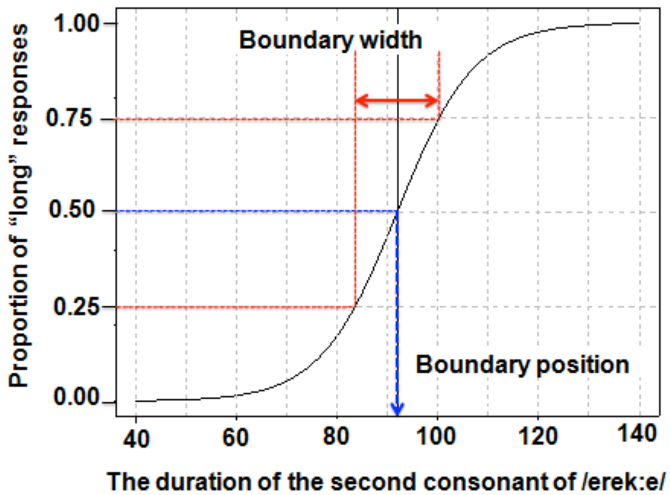

FIGURE 1. An example of the modeling of categorical identification between short and long phonemes by the logistic function. The modeling provides two indexes that represent different aspects of perception: Boundary position and boundary width.

3. RESULTS

The obtained perceptual indexes from the logistic modeling are calculated. The present study investigated the following points. (a) **Boundary Position**: the difference between the boundary position of Korean listeners and Japanese listeners investigate. In addition, how the boundary position is moved by the speaking rates was investigated. (b) **Boundary Width**: the difference between the boundary widths of Korean listeners and Japanese listeners was investigated. Also, the boundary width was investigated to determine the influence of speaking rate variation.

Fig. 2 (a) and (b) indicate the boundary position of (a) /erek:e/-/ereke/ and (b) /eres:e/-/erese/ by Korean and Japanese listeners as a function of the speaking rate. For Korean listeners' /erek:e/-/ereke/, mean boundary position were 54.8 ms (s.d.=9.5) at the fast speaking rate, 92.7 ms (s.d.=10.0) at the normal speaking rate, and 113.1 ms (s.d.=14.8) at the slow speaking rate. For native listener' /erek:e/-/ereke/, mean boundary position were 55.0 ms (s.d.=5.1) at the fast speaking rate, 91.9 ms (s.d.=7.6) at the normal speaking rate, and 126.5 ms (s.d.=17.8) at the slow speaking rate. Moreover, For Korean listeners' /eres:e/-/erese/, mean boundary position were 80.0 ms (s.d.=11.0) at the fast speaking rate, 103.3 ms (s.d.=10.0) at the normal speaking rate, and 132.2 ms (s.d.=12.8) at the slow speaking rate. For native listener' /eres:e/-/erese/, mean boundary position were 87.2 ms (s.d.=9.1) at the fast speaking rate, 97.5 ms (s.d.=10.8) at the normal speaking rate, and 141.9 ms (s.d.=16.0) at the slow speaking rate.

For further understanding, a two-way repeated-measures ANOVA, with group (Korean, Japanese) as a between-subjects variable, and speaking rate (fast, normal, slow) as within-subjects variables were conducted with correction for sphericity based on Green-house and Geisser' s (1959) method because the variance was not equivalent.

The main effect of speaking rate [$F(1.7,64)=449.7$, $p<0.01$] were significant. In contrast, the main effect of group [$F(1,32)=1.65$, n.s.] were not significant. The rate-by-group interaction [$F(2,64)=6.7$, $p<0.05$] were significant. For more detail understanding, the rate-by-group interaction was analyzed by examining the simple main effect of speaking rate for group (Korean and Japanese). Results revealed that the simple main effect of speaking rate was significantly different all speaking rates (Fast < Normal < Slow, $p<0.01$). The examining the simple main effect of group for speaking rate was not significant. Further analysis conducted the multiple comparisons with Bonferroni correction showed that the all speaking rate have significantly different for both Japanese and Korean listeners: fast < normal < slow ($p<0.01$).

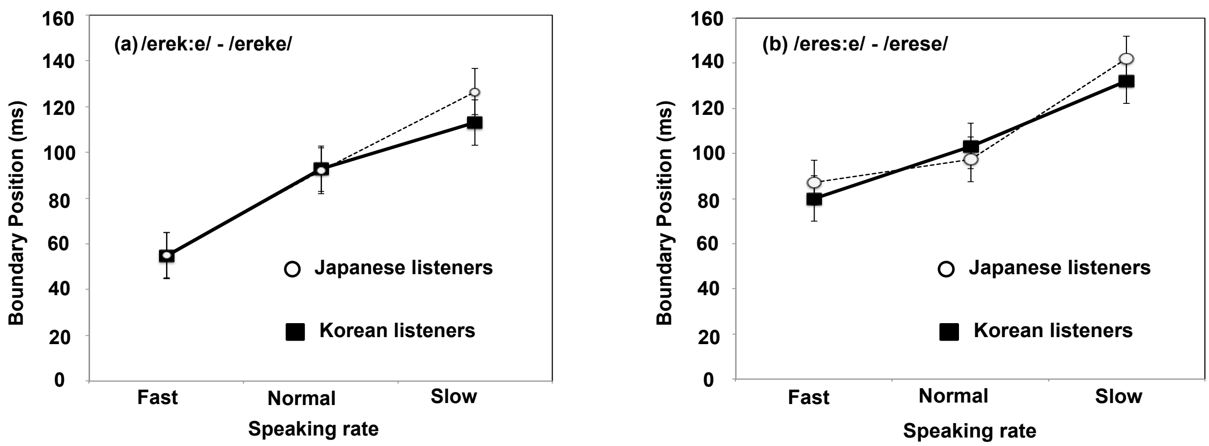

FIGURE 2. Boundary position of the participants' perceptual characteristics as a function of group (Korean listeners, Japanese listeners) and speaking rate (fast, normal, slow). (a) indicates the mean boundary position of identification between /erek:e/ and /ereke/. (b) indicates the mean boundary position of identification between /eres:e/ and /erese/. Each whisker indicates the standard error of the mean.

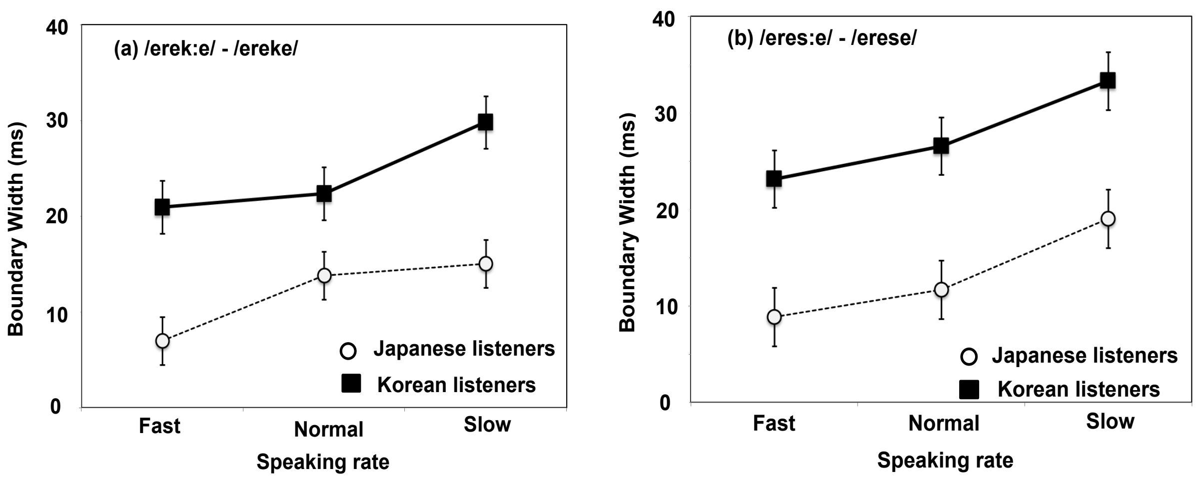

FIGURE 3. Boundary width of the participants' perceptual characteristics as a function of group (Korean listeners, Japanese listeners) and speaking rate (fast, normal, slow). (a) indicates the mean boundary width of identification between /erek:e/ and /ereke/. (b) indicates the mean boundary width of identification between /eres:e/ and /erese/. Each whisker indicates the standard error of the mean.

Next, /eres:e/ - /erese/ also were submitted a two-way repeated-measures ANOVA, with group (Korean, Japanese) as between-subjects variable, and speaking rate (fast, normal, slow) as within-subjects variables. The results indicated that the main effect of speaking rate [$F(1.6,68)=389.3$, $p<0.01$] were significant, the main effect of group [$F(1,34)=1.1$, n.s.] were not significant. The rate-by-group interaction [$F(1.6,68)=554.9$, $p<0.01$] were significant. The rate-by-group interaction was analyzed by examining the simple effect of speaking rate for group (Korean and Japanese). Results revealed that the effect of speaking rate was significantly different all speaking rates (Fast < Normal < Slow, $p<0.01$). The examining the simple main effect of group for speaking rate was not significant. Further analysis conducted the multiple comparisons with Bonferroni correction showed that the all speaking rate have significantly different for both Japanese Korean listeners: fast < normal< slow ($p<0.01$).

These results indicated that both Korean and Japanese listeners have shifted the boundary position according to the speaking rate variation. In addition, the shifting of boundary position revealed not only the identification of /erek:e/ - /ereke/ but also the identification of /eres:e/ - /erese/ by Korean listeners. Last, the boundary position did not significantly different between Korean listeners and Japanese listeners. The obtained results show that the effect of the speaking rate on the identification of singleton and geminate consonants does not differ depending on the type of consonant for Korean listeners. Although the difference in the type of consonant could not be compared straightforwardly where the consonant is /k/ and /s/ because the present examination did not control the other contextual factors, e.g., inherent duration and neighboring segments, overall performance suggests that it is not different for Korean listener to perceive the length contrast of consonants, /k/ or /s/, in Japanese.

Fig. 3 (a) and (b) indicated the boundary width of (a) /erek:e/-/ereke/ and (b) /eres:e/-/erese/ by Korean and Japanese listeners as a function of speaking rate.For Korean listeners' /erek:e/-/ereke/, mean boundary width were 21.0 ms (s.d.=11.1) at the fast speaking rate, 22.4 ms (s.d.=9.0) at the normal speaking rate, and 29.8 ms (s.d.=21.2) at the slow speaking rate. For native listener' /erek:e/-/ereke/, mean boundary position were 7.0 ms (s.d.=5.4) at the fast speaking rate, 13.8 ms (s.d.=3.8) at the normal speaking rate, and 15.1 ms (s.d.=6.1) at the slow speaking rate. Moreover, For Korean listeners' /eres:e/-/erese/, mean boundary position were 23.2 ms (s.d.=11.7) at the fast speaking rate, 26.6 ms (s.d.=15.8) at the normal speaking rate, and 33.3 ms (s.d.=27.0) at the slow speaking rate. For native listener' /eres:e/-/erese/, mean boundary position were 8.9 ms (s.d.=3.5) at the fast speaking rate, 11.7 ms (s.d.=6.8) at the normal speaking rate, and 19.2 ms (s.d.=14.2) at the slow speaking rate.

A two-way repeated-measures ANOVA with group (Korean, Japanese) as a between-subjects variable, and speaking rate (fast, normal, slow) as within-subjects variables were conducted with correction for sphercity based on Greenhouse and Geisser' s (1959) method. The main effect of speaking rate [$F(1.55,64)=8.0$, $p<0.01$] were significant. The main effect of group [$F(1,32)=11.3$, $p<0.01$] were significant. The rate-by-group interaction [$F(1.55,64)=1.3$, n.s.] were not significant. As a further analysis conducted simple main effect test of rate. The results revealed that fast speaking rate were significantly low boundary width compare with both normal speaking rate ($p<0.05$) and slow speaking rate ($p<0.01$). Between the normal and slow speaking rate were not significant. Fig. 3 (b) indicated the results of the boundary width of /eres:e/-/erese/. A two-way repeated-measures ANOVA with group (Korean, Japanese) as a between-subjects variable, and speaking rate (fast, normal, slow) as within-subjects variables were conducted. The main effect of speaking rate [$F(1.55,68)=6.4$, $p<0.01$] were significant. The main effect of group [$F(1,34)=8.9$, $p<0.01$] were significant. The rate-by-group interaction [$F(1.55,68)=0.01$, n.s.] were not significant. As a further analysis conducted simple main effect test of rate. The results revealed that slow speaking rate were significantly high boundary width compared with both fast ($p<0.05$) and normal speaking rate ($p<0.05$). Between the fast and normal speaking rate were not significant.

These results indicated that the Korean listeners' boundary widths of singleton and geminate consonants were significantly larger than those of Japanese listeners for both types of consonant at all speaking rates. These results revealed that the Korean listeners have larger identical extent which led to them misidentifying singleton and geminate consonants.

4. GENERAL DISCUSSION

The present study investigated how non-native listeners perceive the singleton and geminate consonants of Japanese at speaking rate variation through a categorical perception test. A mathematical modeling of the listeners' response was analyzed from two points of view: a) boundary position and b) boundary width. Overall performance indicated that although the boundary position shifted to cope with different speaking rates for Korean and Japanese listeners, the Korean listeners have a more variable boundary width than native listeners at all speaking rates. Obtained results discussed as the following points.

First is the influencing of the speaking rate which is the language-independent factor. Our prediction was that if the influence of the speaking rates were more reliable, the boundary position did not shift as similar as Wilson *et al.* (2005) ' results. However, the results revealed that the boundary position of Korean listeners shifted along with the speaking rate. Although the condition of stimuli did not exactly correspond with that in the present study, these observation differences are presumed to have been caused by listeners' L1. Specially, Korean has three kinds of voiceless stops: aspirated /p^h t^h k^h/, tense /p* t* k*/, and lax /p t k/. Investigating the similarity between the consonant contrast of Korean and geminate consonant of Japanese, Sonu *et al*. (2012) addressed Korean intervocalic tense consonants, which is an epenthetic /s/, in an phonological system to make the compound noun has similar to geminate consonant aspect of not only segmental duration but also psycho-acoustical factor. If perceptual similarity between geminate consonants in Japanese and intervocalic tense consonant in Korean exists, Korean listeners can use the criteria of L1 to identify the singleton and geminate consonants in Japanese.

Next, from the view point of the L2 learning process, the present study of the boundary width has larger extent than Japanese listeners at all speaking rates. These results suggest that even though the Korean listeners identified singleton and geminate consonants at the different speaking rates, they did not differentiate between singleton and geminate consonants accurately. These results also suggest that the Korean listeners' misidentifications were caused not by an inability to cope with the speaking rate variation but an ambiguous criterion. Moreover, Korean listeners have different perception of consonant length contrasts compared with Japanese listeners' results. These results correspond with Kato *et al*. (2004) which was English listeners have remarkably different compared with Japanese native listeners. These ambiguous criteria easily affect the contextual factor for non-native listeners to identify phonemic length contrast by non-native listeners. Moreover, perception to use the ambiguous criteria means that the other contextual factors could also be primary criterion to identify the phonemic length contrasts that were not substantial factors for native listeners. Sonu *et al*. (2011) found that the misidentification of singleton and geminate consonants by Korean listeners correlated with the loudness of stimuli in some part. It is possible that there are specific word-components that are difficult for Korean listeners to identify. According to our prediction, it is more difficult for Korean listeners to identify the consonant /k/ than /s/. The results of the within the contrast type of consonant, /k/ and /s/ was not remarkably different for either boundary position or boundary width. These results could not support both Kinoshita (2011) and Toda (2003).

5. CONCLUSION

The first purpose of the present study was to investigate the perception of consonant length contrast with speaking rate variations by Korean native listeners particularly focusing on the boundary position and width. The results through the categorical perception performance of Korean native listeners in categorizing the consonant length-based phonemic contrast at different speaking rates did not show similar manner with English listeners who did not follow the variation in speaking rate. Korean native listeners shifted the boundary positions along with the speaking rate variation. These results suggest that non-native listeners' perception of length contrast of Japanese is influenced by the listeners' Ll. It might be related with the perceptual similarity of the listeners' Ll.

The second purpose of the present study was to investigate Korean listeners' boundary position and width. Results revealed that even though Korean listeners' perceptual boundary position between singleton and geminate consonants shifted according to speaking rate in a similar manner as the natives, the boundary width was remarkably and significantly different compared with Japanese native listeners at all speaking rates. These results suggest that Korean listeners have an ambiguous criteria for phonemic length contrasts.

The third and last purpose of the present study was to investigate whether the Korean listeners have a different tendency of perception within the contrast type of consonants. The results of the present study did not show remarkable differences within the contrast type of consonants.

ACKNOWLEDGMENTS

This work was supported in part by Grant-in-Aid for Scientific Research No. 20300069, No.2402004 of JSPS.

REFERENCES

Fujisaki, H., Nakamura, K., and Imoto, T. **(1975)**. "Auditory perception of duration of speech and non-speech stimuli," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. Tatham, (Academic Press, London), pp. 197–219.

Greenhouse, S. W., and Geisser, S. **(1959)**. "On methods in the analysis of profile data," Psychometrika **24**, 94–112.

Hirata, Y., Whitehurst, E., and Cullings, E. **(2007)**. "Training native English speakers to identify Japanese vowel length contrasts with sentences at varied speaking rates," J. Acoust. Soc. Am. **121**, 3837–3845.

Kato, T., Tajima, K., Rothwell, A., Akahane-Yamada, R., and Munhall, K. **(2004)**. "Perception of phonemic length contrasts in Japanese with or without a carrier sentence by native and non-native listeners," Proc. 18th Int. Congr. Acoust., 609–612.

Kinoshita, N. **(2011)**. *The acquisition and teaching of Japanese rhythm* (in Japanese) (Ph.D. dissertation, Waseda University, Tokyo).

McMurray, B. and Spivey, M. **(1999)**. "The categorical perception in consonants: The interaction of learning and processing," Proc. Chicago Linguist. Soc. **35**, 205–220.

Paul, B. and Weenink, D. **(2012)**. Praat: doing phonetics by computer [Computer program]. Version 5.3.39, http://www.praat.org/.

Sagisaka, Y. and Tohkura, Y. **(1984)**. "Phoneme duration control for speech synthesis by rule" (in Japanese), Trans. Inst. Electron. Comm. Eng. Jpn. **J67-A**, 629–636.

Sonu, M., Tajima, K., Kato, H., and Sagisaka, Y. **(2011)**. "Perceptual studies of Japanese geminate insertion phenomena based on timing control characteristics," Proc. 17th Int. Congr. Phon. Sci., 1886–1889.

Sonu, M., Tajima, K., Kato, H., and Sagisaka, Y. **(2012)**. "Perceptual characteristics of Japanese geminate consonants by Korean native listeners: Focusing on the relationship between geminate insertion and Korean intervocalic tense consonants" (in Japanese), Inst. Electron. Inf. Commun. Eng. Tech. Rep. SP2011-156. **111(471)**, 7–12.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. **(2008)**. "Training English listeners to perceive phonemic length contrasts in Japanese," J. Acoust. Soc. Am. **123**, 397–413.

Toda, T. (**2003**). *Second Language Perception and Production: Acquisition of Phonological Contrasts in Japanese* (University Press of America, Lanham, MD).

Wilson, A., Kato, H., and Tajima, K. **(2005)**. "Native and non-native perception of phonemic length contrasts in Japanese: effects of speaking rate and presentation context," J. Acoust. Soc. Am. **117**, 2425(A).