Perception of Japanese Consonant Length by Native Speakers of Korean Differing in Japanese Learning Experience

Kimiko Tsukada[1,2], Joo-Yeon Kim[3], Jeong-Im Han[3]

[1]Macquarie University, Australia
[2]The University of Melbourne, Australia
[3]Konkuk University, Korea

kimiko.tsukada@gmail.com, kjooyeon@konkuk.ac.kr, jhan@konkuk.ac.kr

Abstract

The perception of Japanese consonant length contrasts (i.e. short/singleton vs long/geminate) by native and non-native listeners was compared to examine the extent to which difficult foreign language (FL) sounds are processed accurately by native speakers of Korean (NK). Three NK groups differed in their experience with Japanese: non-learners, intermediate and advanced. Via the AXB task, the NK speakers' discrimination accuracy of Japanese consonant length contrasts was assessed and compared to that of a group of 10 native speakers of Japanese (NJ) who served as controls. On average, the NK advanced group did not significantly differ from the NJ group and outperformed the NK non-learner (but not the NK intermediate) group. The NK intermediate and non-learner groups did not differ from each other. However, regardless of experience with Japanese, the NK speakers may benefit from the first language (L1) laryngeal contrasts, associating L1 Korean fortis consonants with Japanese geminates. The NK advanced group appeared less affected than the other two NK groups by Japanese pitch accent patterns in their consonant length perception. The NK advanced learners' results demonstrate that it is possible for non-native speakers to acquire native-like discrimination of consonant length in adulthood.

Index Terms: consonant length, short/singleton, long/geminate, Japanese, Korean

1. Introduction

Durational variation is a phonetic feature of natural languages whether it serves as a lexical cue or not. Japanese is one of the representative languages that use durational variation or length contrastively for both vowels and consonants [1, 2]. For example, *kite* 'wearing' contrasts with *kiite* 'listening' on the one hand and with *kitte* 'stamp' on the other hand. Duration is the primary (though not the only) acoustic cue to differentiate the short and long members of the length contrast [3]. It is widely acknowledged that length contrast is difficult to learn for non-native learners from diverse first language (L1) backgrounds [4-8]. However, incorrect use of contrastive length leads to confusion, embarrassment or misunderstanding. It is therefore important to advance our current knowledge of how individuals with and without exposure to the target languages process and ultimately acquire unfamiliar speech sounds for efficient communication and improved pronunciation pedagogy in foreign languages (FLs). Length contrasts may be difficult to acquire, because they are not as frequent cross-linguistically or as robust as other phonetic contrasts such as the voicing contrast (e.g. 'tip' vs 'dip'), which is supported by multiple co-varying acoustic cues [9].

The perception of Japanese consonant length contrasts (i.e. short/singleton vs long/geminate) by native and non-native listeners was compared to examine the extent to which difficult Japanese sounds are processed accurately by native speakers of Korean (NK). Unlike previous studies, three NK groups differing in their experience with Japanese and a control group of native Japanese (NJ) speakers participated in an AXB discrimination task as described in the Method section below. The question of interest was whether adult language learners are capable of acquiring native-like skills to process difficult non-native sounds.

Japanese has traditionally been a popular language in many Asian countries and Korea is ranked within top 3 countries/regions in terms of the number of institutions (1st with 2,998), teachers (2nd with 15,345) and learners (3rd with 531,511) of Japanese [10]. From a theoretical perspective, examining NK speakers' cross-language perception of consonant length would be insightful. This is because while the contrastive status of consonant length in Korean is still under debate among phoneticians, it has *phonetic* lengthening of fortis consonants, which results in geminate-like sounds [11]. Further, although monomorphemic long consonants are limited, phonological assimilation processes give rise to some long consonants [12, 13]. This may lead to the expectation that the NK speakers show sensitivity to consonant length contrasts in familiar or unfamiliar FLs.

In addition to vowel/consonant length contrasts, Japanese uses pitch accents contrastively (e.g. *kaki* (High-Low (HL)) 'oyster' vs *kaki* (LH) 'persimmon'). We were interested in assessing the influence of Japanese pitch accents on NK speakers' consonant length perception. Previous research showed that native and non-native speakers of Japanese from various L1 backgrounds including Korean were differentially affected by the pitch accent type when they identified (non)words containing a singleton or geminate [5, 14, 15]. Phonetic factors such as 1) segmental (vowel) duration (*shorter* in HL, geminate more frequently perceived vs *longer* in LH, geminate less frequently perceived) in the CVC stimuli [14], 2) closure duration (*shorter* in HL, geminate more frequently perceived vs *longer* in LH, singleton more frequently perceived) [15] and 3) durational differences in V2 in CV1C(C)V2 (*shorter* in HL, geminate more frequently perceived vs *longer* in LH, singleton more frequently perceived) [5] have been identified as a possible cue to a length category.

In [5], the NK learners with beginner level proficiency in Japanese misperceived singletons as geminates more frequently when the accent pattern was HL rather than when it was LH. Conversely, misperception of geminates as singletons was more

frequent when the accent type was LH than when it was HL. As most Korean dialects including standard Korean in the Seoul/Kyunggi area have lost their lexical pitch accent system (in particular, the younger generation) [16] and do not use F0 variation for lexical distinctions at the level of words, it is unclear if earlier findings of differential pitch accent effect [5] are replicable or not. Further, the pitch accent effect, if found, may depend on the Japanese proficiency/experience level. For instance, the effect may be reduced by the time the NK learners have gained advanced proficiency in Japanese. Our experimental design including multiple groups of NK speakers enables us to cross-sectionally explore the time course of their Japanese speech learning and may provide valuable information for FL Japanese pronunciation pedagogy.

2. Methods

2.1. Stimuli preparation

2.1.1. Speakers

Six NJ speakers (3 males, 3 females) participated in the recording sessions, which lasted between 45 and 60 minutes. The speakers' age ranged from late twenties to early forties. According to self-report, which was confirmed by the first author who is a NJ speaker originally from Tokyo, all NJ speakers spoke standard Japanese, having been born or having spent most of their life in the Kanto region surrounding the Greater Tokyo Area. The NJ speakers were recorded in the recording studio at the National Institute of Japanese Language and Linguistics, Tokyo.

2.1.2. Speech materials

Table 1 shows 12 Japanese word pairs used in this study. The /(C)VC(C)V/ tokens contained singleton or geminate consonants intervocalically (underlined). As the pitch accent type (HL vs LH) has been reported to affect native [14, 15] and non-native [5] listeners' length perception in Japanese, tokens with both HL and LH were included in the stimuli. Only tokens with stops were considered in this study. As voiced geminates are very limited in Japanese [17, 18], only voiceless stops (/t k/) were used.

Table 1: *Japanese words used with target sounds underlined and bolded.*

	Singleton		Geminate	
stop	HL	LH	HL	LH
/t/	*kato* 'transition'	*heta* 'unskilled'	*katto* 'cut'	*hetta* 'decreased'
	mate 'wait'	*oto* 'sound'	*matte* 'waiting'	*otto* 'husband'
	sate 'well, then'	*wata* 'cotton'	*satte* 'leaving'	*watta* 'broke'
/k/	*ika* 'below'	*ake* 'open'	*ikka* 'lesson one'	*akke* 'appalled'
	kako 'past'	*haka* 'grave'	*kakko* 'parenthesis'	*hakka* 'mint'
		saka 'slope'		*sakka* 'author'
		shike 'rough sea'		*shikke* 'humidity'

To record the stimuli, each word was presented on a computer screen in random order and produced in two separate conditions: one in isolation and the other in a carrier sentence (/sokowa _____ to jomimasu/ 'You read it as _____ there'). The pace of presentation was controlled by the experimenter (the first author). The speech materials were digitally recorded at a sampling rate of 44.1 kHz and the target words were segmented and stored in separate files. To avoid inter-speaker variation in fluency (specifically, the duration of a pause before and after the target word), only tokens produced in isolation were used as experimental stimuli in this study.

2.2. Participants

Four groups of listeners participated in an AXB discrimination task. The first group consisted of 12 (NK non-learner; 5 males, 7 females, *mean age* = 21.7 years, *sd* = 3.0) listeners with no experience of Japanese. The second group consisted of 6 (NK intermediate; 2 males, 4 females, *mean age* = 21.3 years, *sd* = 2.1) listeners who passed JLPT (Japanese Language Proficiency Test) at N3 with a mean length of learning of 4.2 years and were considered as intermediate learners. The third group consisted of 6 (NK advanced; 1 male, 5 females, *mean age* = 22.0 years, *sd* = 2.0) listeners who passed JLPT at N1 with a mean length of learning of 8 years and were considered as advanced learners. The fourth and last group consisted of 10 (NJ; 2 males, 8 females, *mean age* = 21.0 years, *sd* = 0.8) native listeners of Japanese who served as controls. None of the NJ listeners participated in the recording sessions. According to self-report, all four groups of listeners had normal hearing.

All listeners were tested individually in a session lasting approximately 30 to 40 minutes in a sound-attenuated laboratory at a university in Korea (for the three NK groups) or USA (for the NJ group). All the NJ listeners were born and spent the majority of their life in Japan. Their mean length of residence in the US was 0.4 years (*sd* = 0.22) at the time of participation. The experimental session was self-paced. The participants heard the stimuli at a self-selected, comfortable amplitude level over the high-quality headphones on a desktop computer in the sound-treated experiment room.

2.3. Procedures

The listeners participated in a forced-choice AXB discrimination test, in which they were asked to listen to trials arranged in a triad (A-X-B). The presentation of the stimuli and the collection of perception data were controlled by the PRAAT program [18]. In the AXB test, the first (A) and third (B) tokens always came from different length categories, and the listeners had to decide whether the second token (X) belonged to the same category as A (e.g. 'yoka$_2$'-'yoka$_1$'-'yokka$_3$') or B (e.g. 'soto$_3$'-'sotto$_1$'-'sotto$_2$'; where the subscripts indicate different speakers).

The listeners listened to 200 trials. The first eight trials were for practice and were not analyzed. The three tokens in all trials were spoken by three different speakers. Thus, X was never acoustically identical to either A or B. This was to ensure that the listeners focused on relevant phonetic characteristics that group two tokens as members of the same length category without being distracted by audible but phonetically irrelevant within-category variation (e.g. in voice quality). This was considered a reasonable measure of listeners' perceptual capabilities in real world situations [20]. All possible AB combinations (i.e. AAB, ABB, BAA, and BBA 48 trials each) were tested.

The listeners were given two ('A', 'B') response choices on the computer screen. They were asked to click on the option 'A' if they thought that the first two tokens in the AXB sequence were the same and to click on the option 'B' if they thought that the last two tokens were the same. No feedback was provided during the experimental sessions. The listeners could take a break after 50 trials if they wished. The listeners were required to respond to each trial, and they were told to guess if uncertain. A trial could be replayed as many times as the listener wished in order to reduce their anxiety, but responses could not be changed once given. The interstimulus interval in all trials was 0.5 s.

3. Results and discussion

We used R version 3.6.0 for statistical analyses and data visualization reported below [21].

3.1. Overall results

The overall mean discrimination accuracy was 98.9%, 94.7%, 88.0% and 80.3% for the NJ, NK advanced, NK intermediate and NK non-learner groups, respectively. Four (66.7%) of the NK advanced learners reached the range set by the NJ group whereas none of the NK intermediate learners or the NK non-learners reached the NJ range.

Table 2 shows discrimination accuracy when the length category of the target token (i.e. X in the AXB sequence) was taken into consideration. Figure 1 shows the distributions of percentages of correct discrimination by the four groups of listeners as a function of the length category of target token (geminate vs. singleton). The number of NK advanced learners whose discrimination accuracy reached the range set by the NJ group was 2 (33.3%) for the target singleton and 5 (83.3%) for the target geminate, respectively. Thus, the NK advanced learners were more native-like when the target token contained a geminate than when it contained a singleton.

Table 2: *Mean discrimination accuracy (%) by four groups of listeners. Standard deviations are in parentheses.*

Group	Target geminate	Target singleton
NJ	98.8 (2.4)	99.0 (1.6)
NK advanced	96.2 (5.4)	93.2 (6.0)
NK intermediate	88.3 (6.5)	87.8 (7.0)
NK non-learner	80.6 (7.7)	80.0 (6.2)

Table 3: *Results of one-way ANOVA assessing the effects of Group and multiple comparison tests (significance level at .05).*

df	*F*-value	*p*-value	Between-group comparisons
3, 10.9	40.8	< .001	NJ > NK intermediate, NK non-learner; NK advanced > NK non-learner

A two-way repeated-measures analysis of variance (ANOVA) with group (G: NJ, NK advanced, NK intermediate, NK non-learner) as a between-subjects factor and length category (L: singleton, geminate) as a within-subjects factor yielded a significant main effect of G only [$F(3, 30) = 30.2$, $p < .001$, $\eta_G^2 = .73$]. Table 3 shows the results of one-way ANOVA which assessed the effect of Group (adjusted for not assuming

equal variances) and Dunnett's Modified Tukey-Kramer pairwise multiple comparison *post hoc* tests. The NJ group was significantly more accurate than the NK intermediate and NK non-learner groups, who did not differ from each other. The NK advanced group did not differ from the NJ group. Further, the NK advanced learners outperformed the NK non-learners, but not the NK intermediate learners.

Individual data points in Figure 1 show that the NK intermediate group consists of 3 stronger and 3 weaker learners. In other words, this showed a bimodal distribution with half of the learners patterning like advanced learners and the other half like non-learners. This finding suggests that the NK intermediate group may be at a critical point in their Japanese language development. As we only tested 6 each of the NK intermediate and NK advanced learners, it would be valuable to examine a larger number of learners in future work.
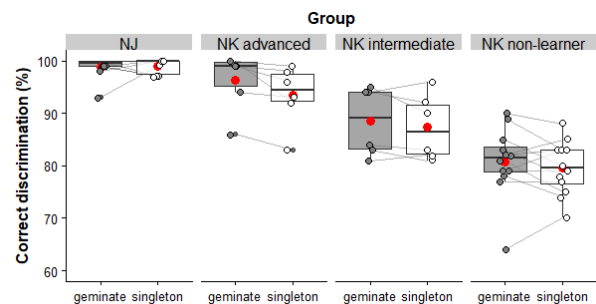

Figure 1: *The distributions of discrimination accuracy (%) by four groups of listeners as a function of length category of the target token (geminate, singleton). The light lines connect individual listeners' scores. The bold horizontal line and the red circle in each box indicate the median and mean, respectively. The bottom and top lines of the box indicate the first and third quartiles. The small black points outside the box are outliers.*

3.2. The effect of pitch accent

Figure 2 shows the distributions of percentages of correct discrimination for the target geminate and singleton by the four groups of listeners as a function of pitch accent type (HL vs LH).
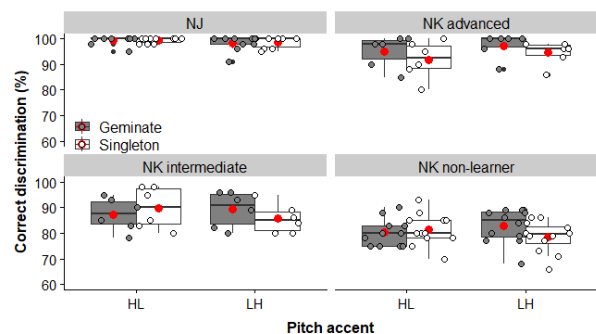

Figure 2: *The distributions of discrimination accuracy (%) for trials with the target singleton or geminate by four groups of listeners as a function of pitch accent (HL, LH).*

It can be observed that the pitch accent type had varying effects on different listener groups. Specifically, the groups

with less or no exposure to Japanese (i.e. NK intermediate and NK non-learner) were slightly more accurate in perceiving the target singleton than the target geminate when the words carried a HL accent (NK intermediate: 87.3% for geminate vs 89.8% for singleton, NK non-learner: 78.3% for geminate vs 81.4% for singleton). In the words with a LH accent, on the other hand, the same groups of listeners were more accurate in perceiving the target geminate than the target singleton (NK intermediate: 89.3% for geminate vs 85.7% for singleton, NK non-learner: 82.8% for geminate vs 78.6% for singleton). The opposite direction of bias according to the pitch accent type was absent in the NJ (HL: 99.3% for geminate vs 99.4% for singleton, LH: 98.3% for geminate vs 98.7% for singleton) and NK advanced (HL: 95.2% for geminate vs 91.8% for singleton, LH: 97.3% for geminate vs 94.5% for singleton) groups.

A three-way repeated-measures ANOVA with group (G: NJ, NK advanced, NK intermediate, NK non-learner) as a between-subjects factor and length category (L: singleton, geminate) and pitch accent type (P: HL, LH) as within-subjects factors yielded a significant main effect of G [$F(3, 30) = 29.9$, $p < .001$, $\eta_G^2 = .68$] and Length x Pitch [$F(1, 30) = 11.1$, $p < .01$, $\eta_G^2 = .03$] and Group x Length x Pitch [$F(3, 30) = 4.1$, $p < .05$, $\eta_G^2 = .03$] interactions. The significant Length x Pitch interaction arose, because while the geminate in the target position was slightly more accurately perceived when it occurred in the words with a LH accent (91.1%) than in the words with a HL accent (89.1%), the pattern of discrimination accuracy was different for the singleton in the target position (HL: 90.0% vs LH: 88.6%). However, this influence of pitch accent depended on the listener group, which resulted in a significant three-way interaction as shown in Figure 2.

It is interesting that the NK intermediate and non-learner groups resembled each other and differed from the NK advanced group in response to the stimuli differing in the pitch accent. Specifically, the less experienced NK listeners' length discrimination was more accurate for the target geminate than for the target singleton when the pitch accent was LH, but this pattern was reversed when the pitch accent was HL. These results differ from those reported in previous research [5], possibly due to methodological differences (identification in [5] vs discrimination in this study).

Alternatively, the diverging results may be related to the Korean phrasal prosody (intonation). In Korean, when the accentual phrase (AP) begins with a fortis or aspirated consonant, the following vowel is associated with a H tone. When the AP begins with a lenis consonant, on the other hand, the following vowel is associated with a L tone. Thus, in the former case, the AP is realized as HH-LH, but the latter AP shows LH-LH phrasal tones. Relatively, Korean listeners with limited exposure to Japanese might show such L1-prosody transfer.

The pitch accent effect was negligible for the NK advanced group who discriminated consonant length contrasts more accurately when the geminate rather than the singleton was in the target position regardless of the pitch accent type. It may be the case that the NK speakers learned to resist the influence of pitch accent in the process of attaining higher proficiency in Japanese. However, the NK advanced learners were not completely native-like in the sense that they showed a slight perceptual bias towards geminates. The NJ group was not susceptible to pitch accent nor length and showed excellent discrimination of length contrasts in all contexts. With respect to the pitch accent effect, we agree with an anonymous reviewer

that it would be interesting to involve older NK learners of Japanese (who still have pitch accent), or possibly dialectal groups (e.g. Seoul/Kyunggi vs Kyungsang) differing in retention of pitch accent.

In summary, the NK advanced group was indistinguishable from the NJ group in many respects and significantly outperformed the NK non-learner, but not the NK intermediate, group. However, all NK groups including the non-learner group were generally accurate (> 80%). In addition, the NK non-learners did not significantly differ from the NK intermediate learners. These findings suggest that, regardless of the Japanese experience, the NK speakers may benefit from exposure to 1) L1 laryngeal contrasts, associating Korean fortis consonants with Japanese geminates [11] and/or 2) morphologically derived long Korean consonants, arising from phonological assimilation processes [12, 13]. To verify this possibility, it would be necessary to examine listeners from other L1 backgrounds. From a pedagogical perspective, the NK intermediate learners may be at a critical point in their Japanese language acquisition, as they showed a bimodal distribution with half of them patterning like advanced learners and the other half like non-learners.

4. Conclusions

By comparing the discrimination accuracy of Japanese consonant length across three NK groups and a NJ control group, we confirmed that the NK advanced group differentiated themselves from the less experienced NK groups and approximated to the NJ group. This suggests that non-native learners are capable of attaining native-like accuracy in cross-linguistic perception in adulthood and also highlights the importance of including multiple groups of participants covering a wide range of proficiency/experience with the target language. While the NK advanced learners did not significantly differ from the NJ listeners in their overall accuracy of Japanese consonant length perception, we do not know how long it took the NK learners to reach this level of proficiency. We believe that the NK intermediate learners are an interesting group that deserves closer attention. To gain a better understanding of the time course of NK learners' second language/FL speech learning, it would be valuable to conduct a longitudinal study.

5. Acknowledgements

6. References

[1] T. J. Vance, *The Sounds of Japanese.* Cambridge: Cambridge University Press, 2008.

[2] P. Ladefoged, *A Course in Phonetics.* Fort Worth: Harcourt Brace & Company, 1993.

[3] H. Fujisaki, K. Nakamura, and T. Imoto, "Auditory perception of duration of speech and nonspeech stimuli," *Annual Bulletin, Research Institute of Logopedics and Phoniatrics,* vol. 7, pp. 45–64, 1973.

[4] M. S. Han, "The timing control of geminate and single stop consonants in Japanese: A challenge for non-native speakers," *Phonetica,* vol. 49, pp. 102–127, 1992.

[5] Y. Minagawa and S. Kiritani, "Discrimination of the single and geminate stop contrast in Japanese by five different language groups," *Annual Bulletin, Research Institute of Logopedics and Phoniatrics,* vol. 30, pp. 23–28, 1996.

[6] M. Sonu, H. Kato, K. Tajima, R. Akahane-Yamada, and Y. Sagisaka, "Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes," *Journal of East Asian Linguistics,* vol. 22, pp. 373–398, 2013.

[7] T. Toda, "Issues regarding geminate consonants in Japanese language education," *Journal of the Phonetic Society of Japan,* vol. 11, pp. 35–46, 2007.

[8] K. Tsukada, F. Cox, J. Hajek, and Y. Hirata, "Non-native Japanese learners' perception of consonant length in Japanese and Italian," *Second Language Research,* vol. 34, pp. 179–200, 2018.

[9] L. Lisker, "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," *Language and Speech,* vol. 29, pp. 3–11, 1986.

[10] The Japan Foundation, *Survey on Japanese-Language Education Abroad.* https://www.jpf.go.jp/j/about/press/2019/029.html 2019

[11] A. Kochtov and Y. Kang, "Supralaryngeal implementation of length and laryngeal contrasts in Japanese and Korean," *Canadian Journal of Linguistics/Revue canadienne de linguistique,* vol. 62, pp. 18–55, 2017.

[12] B. Pajak and R. Levy, "The role of abstraction in non-native perception," *Journal of Phonetics,* vol. 46, pp. 147–160, 2014.

[13] B. Pajak, S. Creel, and R. Levy, "Difficulty in learning similar-sounding words: A developmental stage or a general property of learning?" *Journal of Experimental Psychology: Learning, Memory, and Cognition,* vol. 42, pp. 1377–1399, 2016.

[14] H. Kubozono, H. Takeyasu, M. Giriko and M. Hirayama, "Pitch cues to the perception of consonant length in Japanese," *Proceedings of the 17th International Congress of Phonetic Sciences,* pp. 1150–1153, 2011.

[15] E. Ofuka, "Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceding/following vowels," *Journal of the Phonetic Society of Japan,* vol. 7, pp. 70–76, 2003.

[16] H. Lee and A. Jongman, "Acoustic evidence for diachronic sound change in Korean prosody: A comparative study of the Seoul and South Kyungsang dialects," *Journal of Phonetics,* vol. 50, pp. 15–33, 2015.

[17] Q. Hussein and S. Shinohara, "Partial devoicing of voiced geminate stops in Tokyo Japanese," *The Journal of the Acoustical Society of America,* vol. 145, pp. 149–163, 2019.

[18] S. Kawahara, "The phonetics of *sokuon,* or geminate obstruents," In: H. Kubozono (ed), *Handbook of Japanese Phonetics and Phonology.* Berlin: Walter de Gruyter, pp. 43–78, 2015.

[19] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer* [version 6.0.19], retrieved from http://www.praat.org (Last viewed June 13, 2016).

[20] W. Strange and V. L. Shafer, "Speech perception in second language learners: The re-education of selective perception," In: J. G. Hansen Edwards and M. L. Zampini (eds), *Phonology and Second Language Acquisition.* John Benjamins, pp. 153-191, 2008.

[21] R Core Team, *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ 2019.