INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Cross-linguistic Perception of the Japanese Singleton/Geminate Contrast: Korean, Mandarin and Mongolian Compared

Kimiko Tsukada[1,4], *Yurong*[2], *Joo-Yeon Kim*[3], *Jeong-Im Han*[3], *John Hajek*[4]

[1]Macquarie University, Australia
[2]Inner Mongolia University, China
[3]Konkuk University, Korea
[4]The University of Melbourne, Australia

kimiko.tsukada@gmail.com, umyurong@yahoo.co.jp, {kjooyeon, jhan}@konkuk.ac.kr, j.hajek@unimelb.edu.au

Abstract

The perception of Japanese consonant length contrasts (i.e. short/singleton vs long/geminate) by native and non-native speakers was compared to examine the extent to which difficult foreign language (FL) sounds are processed accurately. Three groups of participants had Korean, Mandarin or Mongolian as their first language (L1) and had no experience with Japanese. Unlike Japanese, Mandarin and Mongolian do not use consonant length contrastively. The phonemic status of consonant length in Korean is debatable. Further, unlike Japanese and Mandarin which predominantly use open syllables and restrict the occurrence of consonants in coda position, Korean and Mongolian permit a wide range of consonants in that syllable position. Via the AXB task, the participants' discrimination accuracy of Japanese consonant length contrasts was assessed and compared to that of a group of 10 native Japanese speakers who served as controls. The Japanese group was at near ceiling with little individual variation. The Mongolian (but not Korean and Mandarin) group did not significantly differ from the control group when the target token (X) contained a geminate. All non-native groups were significantly less accurate than the control group when X contained a singleton. These results were interpreted as reflecting the participants' L1 quantity system.

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

1. Introduction

Japanese is well-known as a quantity language that uses durational variation (or length) contrastively for both vowels and consonants [1, 2]. For example, *kite* 'wearing' contrasts with *k<u>**ii**</u>te* 'listening' on the one hand and with *ki<u>**tt**</u>e* 'stamp' on the other hand. As such, correct use of contrastive length is important for efficient communication. However, it is widely acknowledged that length contrast is difficult to perceive and produce for non-native speakers from diverse first language (L1) backgrounds [3-8]. To advance our current knowledge of how individuals with and without exposure to the target languages process and ultimately acquire unfamiliar speech sounds for successful communication and to improve pronunciation pedagogy in foreign languages (FLs), solid empirical research is necessary. Length contrasts may be difficult to acquire, because they are not as frequent cross-linguistically or as robust as other phonetic contrasts such as a voicing contrast (e.g. 'tip' vs 'dip'), which is supported by multiple co-varying acoustic cues [9].

In this study, 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 speakers differing in their L1s: Korean, Mandarin and Mongolian. These participants who were naïve to Japanese and a control group of native Japanese speakers participated in an AXB discrimination task as described in the Method section below. The question of interest was if and how adult non-native speakers are able to process difficult non-native sounds in comparison with native speakers.

Unlike Japanese, consonant length is not contrastive at the level of words in Mandarin or Mongolian. However, vowel length is contrastive in Mongolian [10] and this may sensitize native Mongolian speakers to vowel and possibly consonant length in FLs. Alternatively, L1 experience with vowel length may not transfer optimally to cross-language speech processing [11, 12]. Whether consonant length is contrastive or not in Korean is still under debate. While some consider fortis consonants to show *phonetic* lengthening, which results in geminate-like sounds [13], few others consider fortis to be *phonological* geminates [14-16]. In support of this latter view, there are a few hetero-syllabic geminates (e.g. /mul + li/ [mulli] 'physics', /kuk + ki/ [kukk'i] 'national flag') minimally contrasting with words with a singleton (e.g. /mu + li/ [muri] 'group', /ku + ki/ [kugi] 'ball game'). In addition, phonological assimilation processes also give rise to long consonants phonetically as shown in lateralization (e.g. /nan + lo/ → [nallo] 'a stove') [17, 18].

Furthermore, unlike Japanese which predominantly uses open syllables [19, 20], Korean and Mongolian frequently use closed syllables and a wide range of consonants occur in coda position [21, 22]. Mandarin, on the other hand, places greater restrictions on the distribution of coda consonants and only two nasal stops (/n/ and /ŋ/) are allowed in that syllable position [21, 23]. Given that geminates are one of the exceptional consonant categories that may close a syllable in Japanese [20], the difference in the preferred L1 syllable structure may also play a significant role in cross-language speech processing. Taken together, these cross-linguistic phonetic characteristics may lead to the expectation that the Korean and Mongolian speakers perceive Japanese consonant length contrasts more accurately than Mandarin speakers.

http://dx.doi.org/10.21437/Interspeech.2021-21

In addition to vowel/consonant length contrasts, Japanese uses pitch accents contrastively (e.g. *kaki* (High-Low (HL)) ‘oyster’ vs *kaki* (LH) ‘persimmon’). Previous research has shown that non-native learners of Japanese from various L1 backgrounds including Korean and Mandarin were differentially affected by the pitch accent type when they identified (non)words containing a singleton or geminate [4]. Specifically, Korean and Mandarin speakers who were introductory learners of Japanese were strongly biased towards geminates and misperceived singletons as geminates more frequently than misperceiving geminates as singletons when the accent pattern was HL. Conversely, misperception of geminates as singletons was more frequent when the accent type was LH than when it was HL. However, for both groups of learners, the direction of misperception was more balanced when the accent type was LH [4].

As most Korean dialects including standard Korean in the Seoul/Kyunggi area have lost their lexical pitch accent system (in particular, the younger generation) [26] and do not use F0 variation for lexical distinctions at the level of words, it is unclear if native Korean speakers without Japanese experience might show a differential pitch accent effect or not. As Mandarin uses F0 variation for lexical distinctions [23], native Mandarin speakers may be skilled in detecting pitch accent differences in Japanese. However, what role their experience with lexical tones may play in the perception of Japanese consonant length with respect to speakers from other non-tonal L1s appears to be under researched.

Similarly, detailed information on how Japanese pitch accents might influence native Mongolian speakers’ perception is limited. While there is agreement among phoneticians that “Mongolian stress (interpreted here as accent) has no distinctive function (no lexical or morphological function) [10, p. 63],” its exact nature (e.g. location, variable vs fixed, pitch vs stress accent) seems to be unresolved, as there is not yet sufficient research on prosodic characteristics of Mongolian [27, 28]. Thus, this study provides rare insights into how ethnic minorities, such as Mongolians, living in China process Japanese sounds. Cross-language comparison involving multiple groups of participants would improve our current understanding of universal vs language-specific characteristics of how difficult FL sounds are processed and have implications for FL pronunciation learning.

2. Methods

2.1. Stimuli preparation

2.1.1. Speakers

Six (3 males, 3 females) native speakers of Japanese 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 native Japanese speaker originally from Tokyo, all speakers spoke standard Japanese, having been born or having spent most of their life in the Kanto region surrounding the Greater Tokyo Area. The 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 (n = 96) or geminate (n = 96) consonants intervocalically (underlined). Only tokens with stops were considered in this study. As voiced geminates are very limited in Japanese [29, 30], only voiceless stops (/t, k/) were used.

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.

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

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

2.2. Participants

Four groups of young adults participated in an AXB discrimination task. The first group consisted of 12 (5 males, 7 females) native speakers of Korean (KRN: *mean age* = 21.7 years, *sd* = 3.0) who were students at Konkuk University in Seoul, Korea. The second group consisted of 13 (6 males, 7 females) native speakers of Mandarin (MND: *mean age* = 20.7 years, *sd* = 0.9). The third group consisted of 12 (6 male, 6 females) native speakers of Mongolian (MGL: *mean age* = 20.1 years, *sd* = 0.7). The Mandarin and Mongolian speakers were students at Inner Mongolia University in Hohhot, China. The Mongolian speakers were born and raised in Inner Mongolia Autonomous Region and Mongolian-Mandarin sequential bilinguals. None of these participants had experience learning Japanese. The fourth and a control group consisted of 10 (2 males, 8 females) native speakers of Japanese (JPN: *mean age* = 21.0 years, *sd* = 0.8) who were students at University of Oregon in Eugene, OR, USA. All Japanese speakers 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. None of the Japanese speakers participated in the recording sessions. According to self-report, all four groups of participants had normal hearing.

All participants were tested individually in a session lasting approximately 30 to 40 minutes in a sound-attenuated laboratory or a quiet room at their own university. 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 or notebook computer.

2.3. Procedures

The participants completed a forced-choice AXB discrimination task, 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 [31]. In the AXB task, the first (A) and third (B) tokens always came from different length categories, and the participants 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 participants listened to a total of 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. Male and female voices were presented together in half of the trials while tokens from three male or three female speakers were presented in the other half of the trials. Thus, X was never acoustically identical to either A or B. This was to ensure that the participants 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 participants’ perceptual capabilities in real world situations [32]. All possible AB combinations (i.e. AAB, ABB, BAA, and BBA, 48 trials each) were tested.

The participants 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 participants could take a break after 50 trials if they wished. The participants 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 participants 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

We used R version 3.6.0 for statistical analyses and data visualization reported below [33].

3.1. Overall results

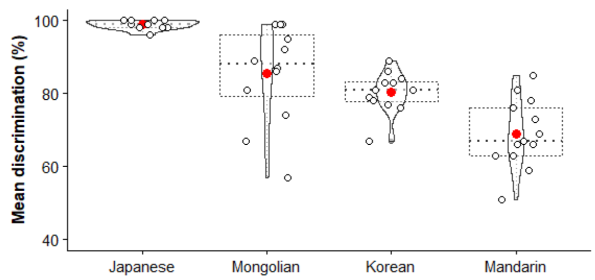

Figure 1: *The distributions of discrimination accuracy (%) by four groups of participants. The horizontal line and the red circle in each box indicate the median and mean, respectively.*

Figure 1 shows the distributions of percentages of correct discrimination by the four groups of participants. The overall mean discrimination accuracy was 99%, 86%, 80% and 69% for the JPN, MGL, KRN and MND groups respectively. Three (25%) of the MGL speakers reached the range set by the JPN group whereas none of the KRN or the MND speakers reached the JPN range. A three-way repeated-measures ANOVA with group (G: JPN, MGL, KRN, MND) as a between-subjects factor and length category (L: geminate, singleton) and pitch accent (P: HL, LH) as within-subjects factors yielded a significant main effect of G [$F(3, 43) = 20.1$, $p < .001$, $\eta_G^2 = .53$] and Group x Length [$F(3, 43) = 3.8$, $p < .05$, $\eta_G^2 = .02$] and Length x Pitch [$F(1, 43) = 8.9$, $p < .01$, $\eta_G^2 = .01$] two-way interactions. The significant interactions were explored in Sections 3.3. and 3.4.

3.2. Results by word pair

Table 2 shows discrimination accuracy by four group of participants according to word pair. Unlike the JPN control group, the non-native groups showed varying degrees of length discrimination accuracy according to the pairs of words presented to them. As pointed out by one reviewer, the first vowel of the target word *shike* was devoiced expectedly by five of the six speakers. However, there is no evidence that this devoicing had a systematic influence on the participants’ discrimination of Japanese consonant length.

Table 2: *Mean discrimination accuracy (%) for each word pair by four groups of participants. Standard deviations are in parentheses.*

Pair	JPN	MGL	KRN	MND
heta-hetta	100 (0)	93 (14.4)	88 (7.2)	77 (15.9)
kato-katto	99 (1.9)	84 (19.1)	77 (15.7)	68 (9.5)
mate-matte	100 (0)	88 (16.6)	87 (13.8)	79 (14.6)
oto-otto	100 (0)	87 (14.4)	84 (17.3)	71 (14.8)
sate-satte	100 (0)	89 (13.4)	76 (6.8)	67 (20.5)
wata-watta	97 (4.2)	86 (12.5)	81 (7.2)	65 (12.0)
ake-akke	98 (2.9)	88 (13.2)	81 (9.5)	63 (9.7)
haka-hakka	98 (3.1)	78 (20.8)	83 (9.3)	62 (15.0)
ika-ikka	99 (2.5)	89 (17.0)	87 (8.5)	79 (17.1)
kako-kakko	98 (2.9)	84 (17.5)	73 (6.9)	65 (15.3)
saka-sakka	98 (2.9)	83 (15.8)	67 (10.0)	62 (14.3)
shike-shikke	99 (1.9)	81 (18.0)	83 (8.5)	72 (10.9)

3.3. The effect of contrastive consonant length

Figure 2 shows the distributions of percentages of correct discrimination by the four groups of participants as a function of the length category of target token (geminate vs. singleton). Table 3 shows discrimination accuracy when the length category of the target token (i.e. X in the AXB sequence) was taken into consideration.

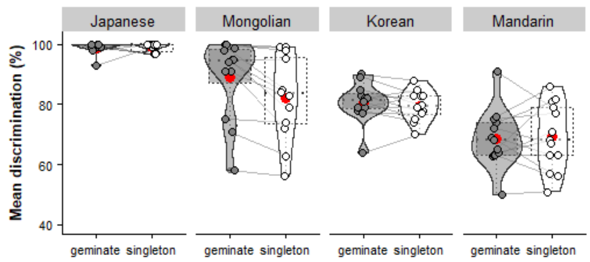

Figure 2: *The distributions of discrimination accuracy (%) by four groups of participants as a function of length category of the target token (geminate, singleton). The light lines connect individual participants’ scores.*

A two-way repeated-measures ANOVA with group (G: JPN, MGL, KRN, MND) as a between-subjects factor and length category (L: geminate, singleton) as a within-subjects factor yielded a significant main effect of G [$F(3, 43) = 21.9$, $p < .001$, $\eta_G^2 = .58$] and Group x Length interaction [$F(3, 43) = 3.6$, $p < .05$, $\eta_G^2 = .03$].

Table 3: *Mean discrimination accuracy (%) by four groups of participants. Standard deviations are in parentheses.*

Group	Target geminate	Target singleton
JPN	99 (2.4)	99 (1.6)
MGL	89 (14.0)	83 (14.7)
KRN	81 (7.7)	80 (6.2)
MND	69 (11.1)	70 (12.8)

Table 4 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 pattern of between-group difference depended on the length category of the target token. For the target geminate, the JPN group outperformed the KRN and MND groups, but not the MGL group. The MND group was significantly less accurate than all other groups. It is notable that the MGL group did not significantly differ from the JPN group and differed only from the MND group in this context. For the target singleton, all three non-native groups, who did not differ from one another, were significantly less accurate than the native JPN group. The lack of significant difference across non-native groups may be due to large individual variability for the MGL and MND groups as shown in Figure 2. It is interesting that the native vs non-native difference was clearer for the target singleton than for the target geminate. In Japanese phonetics, the geminate is called special mora and considered as more marked than singleton or non-special mora. The simple effect of Length did not reach significance for any of the non-native groups.

Table 4: *Results of one-way ANOVA assessing the effects of Group and multiple comparison tests (significance level at .05).*

Length	*df*	*F*	*p*	Between-group comparisons
Geminate	3, 21.0	58.2	< .001	1. JPN > KRN > MND; 2. MGL > MND
Singleton	3, 20.2	78.7	< .001	JPN > MGL, KRN, MND

3.4. The effect of pitch accent

A Length (geminate, singleton) x Pitch (HL, LH) ANOVA did not show significant effects for either of the main factors, but a two-way interaction reached significance [$F(1, 46) = 8.1$, $p < .01$, $\eta_G^2 = .006$]. The difference in mean discrimination accuracy according to pitch accent type was modest for both length categories [geminate: 83% for HL vs 84% LH, singleton: 84% for HL vs 80% for LH]. Neither the simple effect of Length nor the simple effect of Pitch was significant.

4. Discussion

This study examined the perception of Japanese consonant length contrasts by native and non-native speakers whose L1 was Korean, Mandarin or Mongolian. The non-native groups had no experience with Japanese. In summary, the Mongolian group did not differ from the Japanese group and the Mandarin group was less accurate than all the groups for the target geminate. For the target singleton, all non-native groups were less accurate than the Japanese group.

Of the three non-native groups, the Mongolian group resembled the Japanese group the most and the Mandarin group the least. As mentioned in the Introduction, Mongolian has vowel (but not consonant) length contrasts and this may be helpful for consonant length perception. However, our results differ from those reported in the previous study [11], in which native German speakers did not benefit from L1 experience with vowel length in perceiving Italian consonant length. This calls for further cross-linguistic investigation. It is possible that the Mongolian group benefitted from having extensive experience in two languages. We intend to include Mongolian speakers from Mongolia (not China) to determine the role of bi/multilingualism in our future work.

Mandarin is clearly more limited in the range of permissible syllable structures than Korean and Mongolian which allow a wide range of consonants in coda position. Mandarin may be even more restricted than Japanese which permits both nasals and obstruents as geminates. Due to these constraints, Mandarin speakers may face greater challenges than Korean and Mongolian speakers when they need to process Japanese consonant length.

Future work also includes obtaining detailed information about distributional frequency of syllable structures in Mongolian and Korean and examine if and how they are related to Japanese consonant length perception accuracy. It would be interesting to include participants from L1s which predominantly use open syllables other than Mandarin (e.g. French or Spanish [34, 35]) to further examine the role of syllable structure in cross-language speech processing. Furthermore, we intend to continue our investigations with learners of Japanese who share the same L1s with the participants in this study.

5. Conclusions

By directly comparing three non-native groups and a native control group, we found that the Mongolian group perceived Japanese consonant length contrasts most accurately and the Mandarin group least accurately with the Korean group showing an intermediate level of accuracy. This may be related to difference in syllable structure and phonological status of durational variation in their L1. Our results demonstrate that it is possible for some non-native speakers to perceive difficult FL sounds as accurately as native speakers in certain contexts, highlighting the importance of including multiple groups of participants from diverse linguistic experience.

6. Acknowledgements

Part of the data collection for this study was conducted while the first author was affiliated with the University of Oregon on the 2018 Endeavour Research Fellowship. We thank Kaori Idemaru for the use of Spoken Language Research Laboratories and participants for making the study possible. We also thank three anonymous reviewers for their time and input. The first and second authors are supported by the Sumitomo Foundation Fiscal 2020 Grant for Japan-related Research Projects (entitled “Comparison of Japanese sound processing by Mongolian speakers and Mandarin speakers”).

7. 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] 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.

[4] 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.

[5] 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.

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

[7] 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.

[8] T. Uchida, "Characteristics of auditory cognition of long vowels and double consonants for Chinese students in learning Japanese language," *Japanese Journal of Educational Psychology,* vol. 41, pp. 414–423, 2003

[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] A. Iivonen and H. Harnud, H. "Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt," *Journal of the International Phonetic Association,* vol. 35, pp. 59–71, 2005.

[11] H. Altmann, I. Berger, and B. Braun, "Asymmetries in the perception of non-native consonantal and vocalic length contrasts," *Second Language Research,* vol. 28, pp. 387–413, 2012.

[12] K. Tsukada, Y. Hirata, and R. Roengpitya, "Cross-language perception of Japanese vowel length contrasts: Comparison of listeners from different first language (L1) backgrounds," *Journal of Speech, Language, and Hearing Research,* vol. 57, pp. 805–814, 2014.

[13] A. Kochetov 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.

[14] J.-I. Han, "On the Korean tensed consonants and tensification," *Proceedings of the 28 meeting of the Chicago Linguistics Society, University of Chicago,* pp. 206–233, 1992.

[15] D.-I. Choi, "Korean tense consonants as geminates," *Kansas Working Papers in Linguistics,* vol. 20, pp. 25–38, 1995.

[16] S.-C. Ahn and G. Iverson, "Dimensions in Korean laryngeal phonology," *Journal of East Asian Linguistics,* vol. 13, pp. 345–379, 2004.

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

[18] 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.

[19] H. Kubozono, *Phonetics/Phonology,* Tokyo: Kuroshio, 1998.

[20] M. Shibatani, *The Languages of Japan,* Cambridge: Cambridge University Press, 1990.

[21] J.-I. Han, J.-Y. Kim, and T.-H. Choi, "The role of orthography in lexical processing of the phonological variants in second language," *Journal of Psycholinguistic Research,* vol. 50, pp. 437–445, 2021.

[22] H. Kuribayashi, "The development of reduced vowels and the change of syllabic structure in Mongolian," *The Study of Sounds,* vol. 22, pp. 209–223, 1988.

[23] W.-S. Lee and E. Zee, "Standard Chinese (Beijing)," *Journal of the International Phonetic Association,* vol. 33, no. 1, pp. 109–112, 2003.

[24] 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.

[25] 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.

[26] 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.

[27] A. Karlsson, "The intonational phonology of Mongolian," In S.-A. Jun (Ed.), *Prosodic Typology II: The Phonology of Intonation and Phrasing,* pp. 187-215. Oxford University Press, 2014.

[28] Yurong, "Corpus-based comparative study of prosodic features in Japanese and Mongolian," *Hakuho Foundation Japanese Research Fellowship Final Report,* 2017.

[29] 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.

[30] 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.

[31] 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).

[32] 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.

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

[34] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics,* vol. 11, pp. 51–62, 1983.

[35] P. Delattre and C. Olsen, "Syllabic features and phonic impression in English, German, French and Spanish," *Lingua,* vol. 22, pp. 160–175, 1969.