PERCEPTION OF JAPANESE CONSONANT LENGTH BY ADVANCED LEARNERS FROM MANDARIN- AND MONGOLIAN-SPEAKING BACKGROUNDS

Kimiko Tsukada[1,2], Yurong[3]

[1]Macquarie University, [2]The University of Melbourne, [3]Inner Mongolia University
kimiko.tsukada@gmail.com, umyurong@yahoo.co.jp

ABSTRACT

Japanese uses durational variation contrastively for both vowels and consonants. It is widely acknowledged that length contrast is difficult for non-native speakers from diverse L1 (first language) backgrounds. We examined the perception of Japanese consonant length (i.e., short/singleton vs long/geminate) by advanced learners from Mandarin- and Mongolian-speaking backgrounds.

Unlike Japanese, consonant length is not contrastive in Mandarin or Mongolian. However, vowel length is contrastive in Mongolian. Furthermore, unlike Japanese and Mandarin which predominantly use open syllables, Mongolian frequently uses closed syllables and permits a wide range of consonants in coda position.

The participants responded to 200 trials via an AXB discrimination task. The overall mean discrimination accuracy was 86%, 96% and 99% for the Mandarin, Mongolian and Japanese groups, respectively. The Mongolian group's advantage over the Mandarin group suggests that the above-mentioned L1-related factors may continue to affect the processing of Japanese consonant length even for highly advanced learners.

Keywords: Japanese consonant length, short/singleton vs long/geminate, advanced learners, Mandarin, Mongolian

1. INTRODUCTION

Contrastive use of consonant length is one of the most prominent characteristics of Japanese phonetics and phonology [1-4] and is known to pose difficulties to non-native learners from diverse first language (L1) backgrounds including Mandarin and Mongolian, i.e., target languages in the present study [5-9].

Consonant length is not contrastive in Mandarin or Mongolian at the level of words [10], but vowel length is contrastive in Mongolian [11]. Furthermore, unlike Japanese and Mandarin which predominantly use open syllables [12, 13], Mongolian frequently uses closed syllables and a wide range of consonants occur in coda position [14].

Mandarin places even greater restrictions than Japanese on the distribution of coda consonants and only two nasal stops (/n/ and /ŋ/) are allowed in that syllable position [10, 15, 16]. Given that geminates are one of the exceptional consonant categories that may close a syllable in Japanese [13], the difference in the preferred L1 syllable structure may also play a significant role in cross-language speech processing.

Furthermore, Mongolian learners of Japanese who participated in the present study were from the Inner Mongolia Autonomous Region in China and Mongolian-Mandarin sequential bilinguals. The above-mentioned cross-linguistic phonetic characteristics, together with Mongolian speakers' exposure to both Mongolian and Mandarin since childhood, were expected to facilitate and further boost their perception of Japanese singleton/geminate in comparison with Mandarin speakers. Indeed, the Mongolian speakers were shown to discriminate Japanese consonant length contrasts more accurately than did the Mandarin speakers regardless of Japanese experience [17, 18].

L1 effects in cross-language speech processing may be most obvious for naïve listeners or inexperienced learners [19]. However, L1 tone language advantage, specifically Vietnamese over English, was observed in Mandarin tone learning even for advanced learners [20]. Research involving highly advanced second/foreign language (L2/FL) learners in languages other than English is still limited. We hope to fill this gap by examining if and how advanced learners of Japanese from Mandarin- and Mongolian-speaking backgrounds differ from each other on the one hand and from native Japanese speakers on the other hand in their perception of Japanese singleton/geminate contrasts. Cross-language comparison involving multiple groups of learners differing in L1 and proficiency levels would advance our current understanding of universal vs language-specific characteristics of how difficult L2/FL sounds are processed and learned. The findings would have implications for extensive L2/FL experience and ultimate attainment in cross-language speech learning as well as pedagogy of Japanese pronunciation.

2. METHOD

2.1. Stimuli preparation

2.1.1. Speakers

The experimental stimuli and procedure were identical to those used in our previous research ([21, 22]). 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 (native speaker of Japanese 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

	Singleton		Geminate	
	word	gloss	word	gloss
Alvolar	he**t**a	unskilled	he**tt**a	decreased
	ka**t**o	transient	ka**tt**o	cut
	ma**t**e	wait	ma**tt**e	waiting
	o**t**o	sound	o**tt**o	husband
	sa**t**e	well, then	sa**tt**e	leaving
	wa**t**a	cotton	wa**tt**a	broke
Velar	a**k**e	open	a**kk**e	appalled
	ha**k**a	grave	ha**kk**a	mint
	i**k**a	below	i**kk**a	lesson one
	ka**k**o	past	ka**kk**o	parenthesis
	sa**k**a	slope	sa**kk**a	author
	shi**k**e	rough sea	shi**kk**e	humidity

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

Table 1 shows 12 Japanese word pairs used in this study. The /(C)V**C(C)**V/ tokens contained singleton (n = 96) or geminate (n = 96) consonants intervocalically (underlined and bolded). Only tokens with stops were considered in this study. As voiced geminates are limited in Japanese [2, 23, 24], only voiceless stops (/t, k/) were used. On average, the closure durations were 96 ms and 262 ms for singletons and geminates, respectively. The geminate-to-singleton ratios were 2.7 for alveolars (/t/-/tː/) and 2.8 for velars (/k/-/kː/), respectively. These durational values are in good agreement with what has been reported in previous research [e.g., 25] (see, however, [24] for alveolars).

2.2. Participants

Three groups of young adults participated in an AXB discrimination task. The first group consisted of 10 (5 males, 5 females) native speakers of Mandarin (*mean age* = 24.9 years, *sd* = 1.9) who were on short-term exchange and living in Tokyo, Japan for less than 1 year. They were born and raised in various provinces of China. The second group consisted of 10 (5 males, 5 females) native speakers of Mongolian (*mean age* = 23.3 years, *sd* = 2.0). Two of the male Mongolians lived in Japan for more than 3 years while the other three lived in Japan for less than 1 year. One of the female Mongolians was staying in Tokyo for less than 2 months while the rest participated in the study in China. All non-native learners had passed JLPT (Japanese Language Proficiency Test, according to which, the easiest level is N5 and the most difficult level is N1) at N1 and were considered highly advanced learners. The last and a control group consisted of 10 (2 males, 8 females) native speakers of Japanese (*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 stay 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 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 computer.

2.3. Procedure

The participants completed a two-alternative 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 [26]. 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., 'kato$_2$'-'kato$_1$'-'katto$_3$') or B (e.g., 'kako$_3$'-'kakko$_1$'-'kakko$_2$'; 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. 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 [27]. 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 select the option 'A' if they thought that the first two tokens in the AXB sequence were the same and to select 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 every 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 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 [28]. The packages used include ez [29] and tidyverse [30].

3.1. Overall results

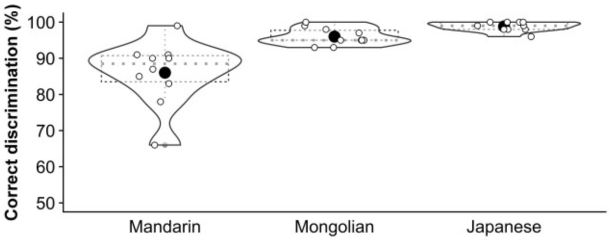

Figure 1: Accuracy (%) of length discrimination by three groups of participants. The horizontal line and the black circle in each box indicate the median and mean, respectively.

Figure 1 shows the distributions of percentages of correct discrimination by the three groups of participants. The overall mean discrimination accuracy was 86%, 96% and 99% for the Mandarin, Mongolian and Japanese groups, respectively. The Japanese group was at near ceiling with little individual variation. Of the two learner groups, the Mandarin group was much more variable than the Mongolian group as clearly seen in Figure 1.

One-way analysis of variance (ANOVA) with group (Mandarin, Mongolian, Japanese) reached significance [$F(2, 27) = 15.4$, $p < .001$, $\eta^2_G = .53$]. According to the post-hoc *t*-tests, all between-group differences were statistically significant [$t(9.4 – 13.9) = -3.2 – -4.5$, $p < .05$]. The Japanese group was significantly more accurate than the Mongolian group which, in turn, was more accurate than the Mandarin group.

3.2. Comparison of the direction of category change (Geminate (G) > Singleton (S) or Singleton > Geminate)

Figure 2 shows the distributions of percentages of correct discrimination for trials differing in the direction (from G to S, from S to G) of length category change.

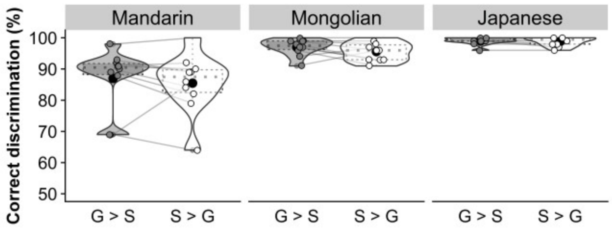

Figure 2: Accuracy (%) of length discrimination for trials differing in the direction of category change (G > S, S > G). The light lines connect individual participants' scores.

Two-way ANOVA with group (Mandarin, Mongolian, Japanese) and direction of category change (G > S, S > G) reached significance only for the main effect of group [$F(2, 27) = 15.3$, $p < .001$, $\eta^2_G = .49$]. Lack of significant two-way interaction suggests that the overall group effect was mirrored whether the length category changed from geminate to singleton or from singleton to geminate within a trial.

3.3. Comparison of the length category (Geminate vs Singleton) of the target token (X in AXB)

Figure 3 shows the distributions of percentages of correct discrimination for trials differing in the length category (Geminate, Singleton) of the target token.

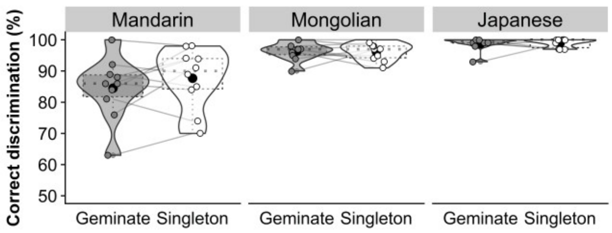

Figure 3: Accuracy (%) of length discrimination for trials differing in the length category (Geminate, Singleton) of the target token.

Two-way ANOVA with group (Mandarin, Mongolian, Japanese) and length (Geminate, Singleton) reached significance only for the main effect of group [$F(2, 27) = 15.9$, $p < .001$, $\eta_G^2 = .49$]. Again, lack of significant two-way interaction suggests that the overall group effect was mirrored whether X in the AXB sequence was singleton or geminate.

3.4. Comparison of the place of articulation (Alveolar vs Velar) of the target token (X in AXB)

Figure 4 shows the distributions of percentages of correct discrimination for trials differing in the place of articulation (Alveolar, Velar) of the target token.

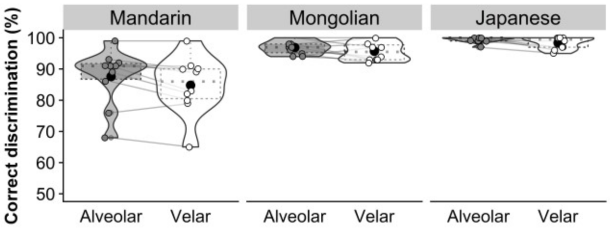

Figure 4: Accuracy (%) of length discrimination for trials differing in the place of articulation (Alveolar, Velar) of the target token.

Two-way ANOVA with group (Mandarin, Mongolian, Japanese) and place (Alveolar, Velar) reached significance for the main effects of group [$F(2, 27) = 15.4$, $p < .001$, $\eta_G^2 = .52$] and place [$F(1, 27) = 11.5$, $p < .01$, $\eta_G^2 = .02$], but not the two-way interaction. This suggests that the overall group effect was mirrored for both places of articulation and the participants were slightly more accurate when the target token was alveolar (95%) than when it was velar (93%).

3.5. Comparison of length discrimination at alveolar (/t/-/tː/) and velar (/k/-/kː/) places of articulation

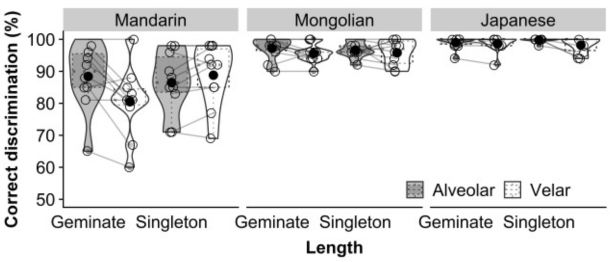

Figure 5: Accuracy (%) of length discrimination for trials differing in the place of articulation (Alveolar, Velar) and the length category (Geminate, Singleton) of the target token.

Figure 5 shows the distributions of percentages of correct discrimination for trials differing in the place of articulation (Alveolar, Velar) and the length category (Geminate, Singleton) of the target token.

Three-way ANOVA with group (Mandarin, Mongolian, Japanese), length (geminate, singleton) and place (Alveolar, Velar) reached significance for the main effects of group [$F(2, 27) = 15.9$, $p < .001$, $\eta_G^2 = .46$] and place [$F(1, 27) = 12.1$, $p < .01$, $\eta_G^2 = .02$]. Some of the interactions involving the place factor were significant [length x place: $F(1, 27) = 8.3$, $p < .01$, $\eta_G^2 = .02$, group x length x place: $F(2, 27) = 9.6$, $p < .001$, $\eta_G^2 = .04$]. As seen in Figure 5, the Mandarin group was differentially affected by the place of articulation when the target token was geminate (88% for Alveolar vs 81% for Velar) as opposed to when it was singleton (87% for Alveolar vs 89% for Velar). This pattern was absent in the Japanese and Mongolian groups.

4. DISCUSSION

This study examined how advanced learners from Mandarin- and Mongolian-speaking backgrounds perceive Japanese consonant length contrasts known to be difficult for non-native speakers. Vowel length is contrastive in Mongolian, but not in Mandarin. Further, unlike Japanese and Mandarin, Mongolian frequently uses closed syllables. We were interested in determining if the Mongolian learners of Japanese continue to outperform the Mandarin counterparts at an advanced level of proficiency.

The Mongolian group's advantage over the Mandarin group observed in this study may be due to the former group's early bilingualism (in Mongolian and Mandarin) and further examination with monolingual Mongolian learners of Japanese is needed. Our results are consistent with the view on Mandarin tone learning by L1 Vietnamese learners that "L1 experience exerts specific and lasting influences on L2 tone perception, well into advanced levels of proficiency" [20].

5. CONCLUSIONS

We investigated the perception of Japanese singleton/geminate by highly advanced learners from Mandarin- and Mongolian-speaking backgrounds to assess the lasting influence of L1. We observed that the Mongolian group resembled the control Japanese group to a greater extent than did the Mandarin group. This suggests that L1-related factors may continue to affect the processing of Japanese consonant length even for highly advanced learners.

6. ACKNOWLEDGEMENTS

This research is supported by the Sumitomo Foundation Fiscal 2020 Grant for Japan-related Research Projects (Grant Number: 208012). We thank reviewers for their time and input for this work.

7. REFERENCES

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

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

[3] Kubozono, H. 2013. Introduction to the special issue on Japanese geminate consonants. *Journal of East Asian Linguistics* 22, 303–06.

[4] Vance T. J. 2008. *The Sounds of Japanese.* Cambridge University Press.

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

[6] Hung, H. Y. 2012. Perception of Japanese geminate stops among Taiwanese learners of Japanese. *Journal of the Phonetic Society of Japan* 16, 15–27.

[7] Liu, Y. L. 2014. The process of acquiring geminates by native speakers of Mongolian (in Japanese). *Nihongo Kenkyu (Japanese Language Studies)* 34, 1–13. (http://nihongo.hum.tmu.ac.jp/tmu_j/archive31-40.html)

[8] Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., Sagisaka, Y. 2013. 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* 22, 373–398.

[9] Tsukada, K, Cox, F., Hajek, J., Hirata, Y. 2018. Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Second Language Research* 34(2), 179–200.

[10] Feng, Q., Busà, M. G. 2022. Mandarin Chinese-speaking learners' acquisition of Italian consonant length contrast. *System* 111, Article 102938.

[11] Iivonen, A., Harnud, H. 2005. Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *JIPA* 35(1), 59–71.

[12] Kubozono, H. 1988. *Phonetics/Phonology.* Kuroshio.

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

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

[15] Lee, W.-S., Zee, E. 2003. Standard Chinese (Beijing). *JIPA* 33(1), 109–112.

[16] Meng, Y., Kotzor, S., Xu, C., Wynne, H. S. Z., Lahiri, A. 2021. Mismatch negativity (MMN) as an index of asymmetric processing of consonant duration in fake Mandarin geminates. *Neuropsychologia* 163, Article 108063.

[17] Tsukada, K., Yurong, Kim, J.-Y., Han, J.-I., Hajek, J. 2021. Cross-linguistic perception of the Japanese singleton/geminate contrast: Korean, Mandarin and Mongolian compared. *Proc.* 22nd *Interspeech* Brno, 3910–3914.

[18] Tsukada, K., Yurong. 2022. Non-native perception of the Japanese singleton/geminate contrast: Comparison of Mandarin and Mongolian speakers differing in Japanese experience. 23rd *Interspeech* Incheon, 3068–3072.

[19] Best, C. T., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn, O.-S., Munro, M. J. (eds), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege.* John Banjamins Walter de Gruyter, 13–34.

[20] Pelzl, E., Liu, J., Qi, C. 2022. Native language experience with tones influences both phonetic and lexical processes when acquiring a second tonal language. *Journal of Phonetics* 95, Article 101197.

[21] Tsukada, K., Idemaru, K., Hajek, J. 2018. The effects of foreign language learning on the perception of Japanese consonant length contrasts. *Proc.* 17th *SST* Sydney, 37–40.

[22] Tsukada, K., Hajek, J. 2022. Adaptation to L3 phonology? Perception of the Japanese consonant length contrast by learners of Italian. *Proc.* 18th *SST* Canberra, 171–175.

[23] Hussein, Q., Shinohara, S. 2019. Partial devoicing of voiced geminate stops in Tokyo Japanese. *J. Acoust. Soc. Am.* 145, 149–163.

[24] Sano, S. 2019. The distribution of singleton/geminate consonants in spoken Japanese and its relation to preceding/following vowels. *Proc. 19th ICPhS* Melbourne, 1833–1837.

[25] Hayes-Harb, R. 2005. Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language and Linguistics* 4, 1–29.

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

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

[28] R Core Team. 2019. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[29] Lawrence, M. A. 2016. *ez: Easy Analysis and Visualization of Factorial Experiments.* R package version 4.4-0. https://CRAN.Rproject.org/package=ez

[30] Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R. et al. 2019. Welcome to the tidyverse. *Journal of Open Source Software* 4(43), 1686.