PERCEPTION OF THE JAPANESE WORD-MEDIAL SINGLETON/GEMINATE CONTRAST BY KELANTAN MALAY SPEAKERS

Mohd Hilmi Hamzah[1], Kimiko Tsukada[2,3], John Hajek[3]

[1]Universiti Utara Malaysia, [2]Macquarie University, [3]The University of Melbourne
hilmihamzah@uum.edu.my, kimiko.tsukada@gmail.com, j.hajek@unimelb.edu.au

ABSTRACT

Both Japanese and Kelantan Malay (KM) use consonant length contrastively, though they differ in terms of word position, i.e., Japanese permits the contrast in word-medial position, while KM restricts such a contrast to word-initial position only. It is thus of theoretical interest to examine the extent to which speakers of these languages process familiar consonant length contrasts (i.e., short/singleton vs long/geminate) in unfamiliar word position.

To this end, we examined the perception of Japanese consonant length by KM participants who had no experience with Japanese. The KM and control Japanese participants responded to 200 trials via an AXB discrimination task. The overall mean discrimination accuracy was 84% and 99% for the KM and Japanese groups, respectively. While there was a clear between-group difference in discrimination accuracy, compared to participants from other first language backgrounds, it appears that the KM group benefitted from their experience with L1 consonant length.

Keywords: Consonant length, short/singleton, long/geminate, Japanese, Kelantan Malay

1. INTRODUCTION

In this study, perception of Japanese consonant length contrasts by native and non-native listeners was compared to examine the extent to which Japanese sounds are processed accurately by speakers of languages with length contrasts differing in word position: word-medial and word-initial positions.

Like Japanese, consonant length is contrastive at the level of words across all consonant groups in Kelantan Malay (KM), as empirically established [1-4]. It was reported that closure duration is the most robust acoustic feature of the singleton/geminate consonant contrast in KM, with non-durational parameters such as root mean square (RMS) amplitude and fundamental frequency (F0) playing important secondary roles in enhancing the word-initial length contrast in this language.

On the one hand, Japanese allows contrastive consonant length in word-medial position, where it is considered to be more perceptually salient [e.g., 5]. On the other hand, consonant length in KM only occurs word-initially, which is more marked and generally avoided across languages [e.g., 6]. It is well known that the word-initial contrast is more perceptually indiscernible, particularly for utterance-initial geminates beginning with voiceless stops as there is ostensibly insufficient acoustic information available for listeners in this particular utterance position [e.g., 7].

However, results from a series of perception experiments [8, 9] show that KM native speakers can reliably differentiate natural stimuli of isolated tokens with word-initial singletons and geminates, including those with voiceless stops in utterance-initial position. The results confirm the primary role of closure duration in geminate perception for KM, at least utterance-medially, while RMS amplitude and F0 show limited perceptual functions in both utterance-initial and medial positions. Given the expectation that word-medial geminates are relatively easier to be distinguished from their corresponding singletons, we hypothesise that KM speakers can potentially perceive the word-medial length contrast in Japanese. That is, their L1 experience with consonant length in KM may transfer optimally to cross-language processing. The roles of other prosodic features, such as syllable structure and intonation with regard to consonant length contrast, appear to be underresearched in KM and cannot be ascertained at this stage.

Empirical work on word-initial geminates has been conducted in a related variety of Pattani Malay [e.g., 10], though research on cross-language speech processing has never been reported between this variety and other foreign languages (FLs), including Japanese. Findings of the current study will have potential contributions for FL pronunciation pedagogy and multilingualism in Malaysia where Japanese is taught as a FL for undergraduate students in most public universities in this country. In this regard, American English learners of Japanese demonstrated a clear advantage over non-learners (93% vs 78%) in their perception of Japanese singleton/geminate contrasts [11]. It would be of theoretical as well as pedagogical importance to determine if KM speakers benefit from a combination of L1 consonant length and FL Japanese learning.

2. METHOD

2.1. Stimuli preparation

2.1.1. Speakers

The experimental stimuli and procedures were identical to those used in previous research ([11, 12]). 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 second author, 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 [13-15], 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., 16] (see, however, [15] for alveolars).

2.2. Participants

Two groups of young adults participated in an AXB discrimination task. The first group consisted of 12 (6 males, 6 females) native speakers of KM (*mean age* = 22.5 years, *sd* = 0.8) who were undergraduate students at Universiti Utara Malaysia in Kedah, Malaysia. They were born and raised in Kelantan, Malaysia and are fluent in Kelantan Malay. They can also speak the standard variety of Malay. None of these participants had experience learning Japanese.

The second 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 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 participants had normal hearing.

All participants were tested individually in a session lasting approximately 20 to 25 minutes in a quiet room at their university. The experimental session was self-paced. The participants heard the stimuli at a self-selected, comfortable amplitude level over the high-quality speakers on a notebook 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 [17]. 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., ‘saka$_2$’-‘saka$_1$’-‘sakka$_3$’) or B (e.g., ‘oto$_3$’-‘otto$_1$’-‘otto$_2$’; where the subscripts indicate different speakers).

The participants listened to a total of 200 unique 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 [18]. 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 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 [19]. The packages used include ez [20] and tidyverse [21].

3.1. Overall results

Figure 1 shows the distributions of percentages of correct discrimination by the two groups of participants. The overall mean discrimination accuracy was 84% and 99% for the KM and Japanese groups, respectively. The Japanese group was at near ceiling with little individual variation. A comparison via the Welch two-sample *t*-test showed that the difference between the KM and Japanese groups was significant [$t(11.5) = -5.3$, $p < .001$].

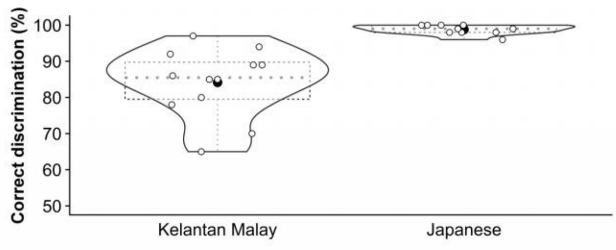

Figure 1: Accuracy (%) of length discrimination by two groups of participants. The horizontal line and the black circle in each box indicate the median and mean, respectively.

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

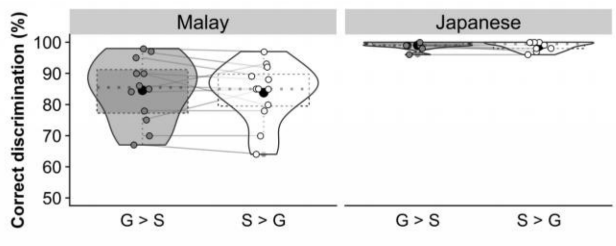

Figure 2: Accuracy (%) of length discrimination for trials differing in the direction of category change. The light lines connect individual participants' scores.

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. The question of interest was if the participants' discrimination accuracy differed between trials that started with a geminate (and ended with a singleton (i.e., G-S-S, G-G-S)) and trials that started with a singleton (and ended with a geminate (i.e., S-G-G, S-S-G)).

Two-way analysis of variance (ANOVA) with group (KM, Japanese) and direction of category change (G > S, S > G) reached significance only for the main effect of group [$F(1, 20) = 23.2$, $p < .001$, $\eta_G^2 = .52$]. As clearly seen in Figure 2, the Japanese group was significantly more accurate than the KM group whether the length category changed from geminate to singleton or from singleton to geminate within a trial. Neither group was biased with respect to the direction of category change.

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. The question of interest was if the participants' discrimination accuracy differed between trials in which X in AXB was a geminate and trials in which X in AXB was a singleton.

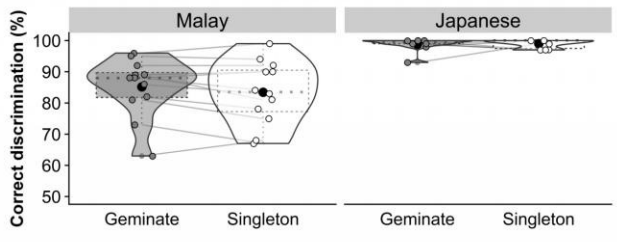

Figure 3: Accuracy (%) of length discrimination for trials differing in the length category of the target token.

Two-way ANOVA with group (KM, Japanese) and length (geminate, singleton) reached significance only for the main effect of group [$F(1, 20) = 22.8$, $p < .001$, $\eta_G^2 = .52$]. As clearly seen in Figure 3, the Japanese group was significantly more accurate than the KM group whether X in the AXB sequence was singleton or geminate. Neither group was biased with respect to the length category of the target token.

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. The question of interest was if the participants' discrimination accuracy differed between trials in which X in AXB was alveolar and trials in which X in AXB was velar.

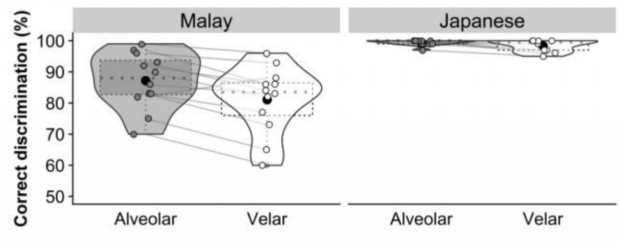

Figure 4: Accuracy (%) of length discrimination for trials differing in the place of articulation of the target token.

Two-way ANOVA with group (KM, Japanese) and place (alveolar, velar) reached significance for the main effects of group [$F(1, 20) = 23.3$, $p < .001$, $\eta_G^2 = .52$] and place [$F(1, 20) = 22.3$, $p < .001$, $\eta_G^2 = .07$]. The two-way interaction was also significant [$F(1, 20) = 10.0$, $p < .01$, $\eta_G^2 = .03$]. As seen in Figure 4, while the Japanese group was not affected by the place factor (99% for alveolar vs 98% for velar), the KM group was more accurate when the target token was alveolar (87%) than when it was velar (81%).

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

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. The question of interest was if the participants' discrimination accuracy differed for trials in which X in AXB varied in both place and length at the same time (i.e., alveolar geminate, velar geminate, alveolar singleton, velar singleton).

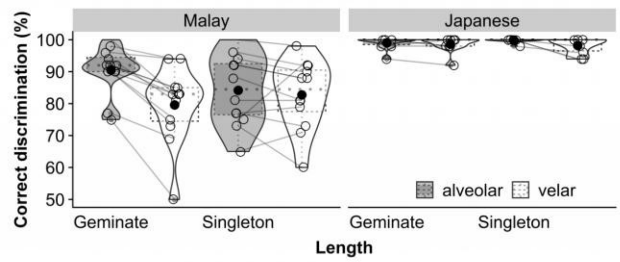

Figure 5: Accuracy (%) of length discrimination for trials differing in the place of articulation and the length category of the target token.

Three-way ANOVA with group (KM, Japanese), length (geminate, singleton) and place (alveolar, velar) reached significance for the main effects of group [$F(1, 20) = 22.9$, $p < .001$, $\eta_G^2 = .49$] and place [$F(1, 20) = 25.4$, $p < .001$, $\eta_G^2 = .06$]. The interactions involving the place factor were all significant [group x place: $F(1, 20) = 11.5$, $p < .01$, $\eta_G^2 = .03$, length x place: $F(1, 20) = 7.3$, $p < .05$, $\eta_G^2 = .02$, group x length x place: $F(1, 20) = 9.6$, $p < .01$, $\eta_G^2 = .03$]. As seen in Figure 5, the influence of place, which was virtually limited to the KM group, was clearer when the target token was geminate (91% for alveolar vs 80% for velar) than when it was singleton (84% for alveolar vs 83% for velar).

4. DISCUSSION

This study examined how KM speakers may perceive Japanese consonant length contrasts known to be difficult for non-native speakers. Consonant length is contrastive in their L1 Malay, but only word-initially. Thus, we were interested in determining if the KM speakers are able to successfully transfer their L1 experience and perceive familiar singleton/geminate contrasts in unfamiliar word-medial position.

While the KM group was less accurate than the Japanese group in discriminating Japanese consonant length, it is notable that their overall score (84%) was intermediate between the previously noted two groups of American English speakers with (93%) and without (78%) Japanese language experience [11]. It thus appears that the KM group benefitted, to some extent, from their experience with L1 consonant length despite lack of Japanese experience.

5. CONCLUSIONS

In the current study, the speakers of KM did not match the Japanese speakers in discriminating Japanese singleton/geminate contrasts word-medially. This may be because they lack specific knowledge of the phonetic characteristics of Japanese singletons and geminates. Only the non-native group was affected by the place factor and discriminated Japanese length contrasts more accurately when alveolar rather than velar occurred in the target position (Figure 4). Our results suggest that experience with Malay singletons and geminates may be helpful but may not automatically transfer to native-like processing of Japanese singletons and geminates.

In the future, we intend to include speakers of other varieties of Malay (e.g., Kedah Malay) who do not use consonant length contrastively in any word positions. Further, given that there are many Malaysian learners of Japanese, it would be pedagogically valuable to include Malay-speaking *learners* of Japanese from different backgrounds and examine if there is additional benefit of Japanese learning within different settings.

6. ACKNOWLEDGEMENTS

This research was supported by Universiti Utara Malaysia (UUM) through University Grant (SO Code: 21403) and 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] Hamzah, M. H., Fletcher, J., Hajek, J. 2011. Durational correlates of word-initial voiceless geminate stops: The case of Kelantan Malay. *Proc. 17th ICPhS* Hong Kong, 815–818.

[2] Hamzah, M. H., Fletcher, J., Hajek, J. 2015. Word-initial voiceless stop geminates in Kelantan Malay: Acoustic evidence from amplitude/F0 ratios. *Proc. 18th ICPhS* Glasgow.

[3] Hamzah, M. H., Fletcher, J., Hajek, J. 2016. Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay. *J. Phon.* 58, 135–151.

[4] Hamzah, M. H., Fletcher, J., Hajek, J. 2020. Non-durational acoustic correlates of word-initial consonant gemination in Kelantan Malay: The potential roles of amplitude and f0. *JIPA* 50(1), 23–60.

[5] 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.

[6] Frej, M. S., Carignan, C., Best, C. T. 2017. Acoustics and articulation of medial versus final coronal stop gemination contrasts in Moroccan Arabic. *Proc. 18th Interspeech* Stockholm, 210–214.

[7] Ridouane, R., Hallé, P. A. 2017. Word-initial geminates: From production to perception. In: Kubozono, H. (ed.), *The Phonetics and Phonology of Geminate Consonants*. Oxford University Press, 34–65.

[8] Hamzah, M. H., Hajek, J., Fletcher, J. 2016. The role of closure duration in the perception of word-initial geminates in Kelantan Malay. *Proc.* 16th *SST* Sydney, 85–88.

[9] Hamzah, M. H., Fletcher, J., Hajek, J. 2019. The secondary roles of amplitude and F0 in the perception of word-initial geminates in Kelantan Malay. *Proc. 19th ICPhS* Melbourne, 2753–2757.

[10] Abramson, A. S. 2003. Acoustic cues to word-initial stop length in Pattani Malay. *Proc. 15th ICPhS* Barcelona, 387–390.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

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

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

[21] 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.