Cross-language Perception of the Japanese Singleton/Geminate Contrast: Case of Vietnamese Speakers With and Without Japanese Language Experience

Kimiko Tsukada¹², Dịch Mục Đào³, Trang Le Thi Huyen⁴

¹Macquarie University, Australia  
²The University of Melbourne, Australia  
³University of Social Sciences and Humanities, Vietnam National University - Ho Chi Minh City, Vietnam  
⁴Waseda University, Japan

kimiko.tsukada@gmail.com, dich.daovns@gmail.com, huyentrangle295@gmail.com

Abstract

To examine the effect of Japanese learning experience on consonant length, we compared the perception of Japanese singleton/geminate contrasts by three groups of Vietnamese speakers with varying Japanese experience. Two of the Vietnamese groups consisted of learners of Japanese with one group participating in Vietnam and the other in Japan. The third Vietnamese group consisted of participants inexperienced in Japanese. A group of 10 native Vietnamese speakers participated as controls. Unlike Japanese, Vietnamese does not use consonant length contrastively.

The participants responded to 200 trials via the AXB discrimination task. The overall mean discrimination accuracy was 67%, 80%, 91% and 99% for the non-learner group, the learner group in Vietnam, the learner group in Japan and the native Japanese group, respectively. A clear difference between the two learner groups in Japan and Vietnam demonstrates learnability of Japanese consonant length beyond early childhood in an immersion setting. At the same time, a non-negligible difference between the native Japanese and the advanced learner groups suggests genuine difficulty of Japanese consonant length.

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

1. Introduction

Japanese, which is a popular foreign language in both East and West, uses durational variation contrastively for both vowels and consonants [1, 2]. For example, yokka ‘leisure’ contrasts with yokka ‘eight days’ on the one hand and with yokka ‘four days’ on the other hand. It is widely acknowledged that length contrast is difficult for non-native speakers from diverse L1 (first language) backgrounds including Vietnamese, which is the target language of this study [3, 4]. Unlike Japanese, consonant length (i.e., short/singleton vs long/geminate) is not contrastive in Vietnamese [4, 5]. Thus, we were interested in how different experience with Japanese may affect Vietnamese speakers’ perception of difficult Japanese contrasts.

Recently, the number of Vietnamese learners of Japanese has been rapidly increasing [6]. Vietnam is currently ranked within the top 6 countries/regions of the world in terms of the number of learners (169,582) of Japanese ([7]). Within Japan, Vietnam (31,643 or 14.4%) is the second largest country of origin of non-native learners of Japanese after China (67,027 or 30.5%) as of 2022 ([8]). As such, improving our understanding of how to facilitate pronunciation pedagogy for Vietnamese learners of Japanese is significant.

This research compared three groups of Vietnamese speakers and a control group of native Japanese speakers to determine how they may differ in their perception of Japanese consonant length contrasts on account of their experience with Japanese. Two of the three Vietnamese groups consisted of learners of Japanese with one group participating in Japan (Tokyo) and the other group in Vietnam (Ho Chi Minh City). The third group who also participated in Ho Chi Minh City consisted of participants inexperienced in Japanese. In addition to the countries of residence, the two learner groups substantially differed in their Japanese proficiency based on the Japanese-Language Proficiency Test (JLPT), according to which, the easiest level is N5 and the most difficult level is N1 [9]. As described in the Participants section, the learners in Japan had all passed the JLPT at N1 and were considered highly advanced learners. The learners in Vietnam were at the JLPT N3 level which requires “the ability to understand Japanese used in everyday situations to a certain degree” (https://www.jlpt.jp/e/about/levelsummary.html) and were considered (pre-)intermediate learners.

The main objective of the present study was to evaluate the effect of Japanese language experience on native Vietnamese speakers’ perception of Japanese short/singleton vs long/geminate contrasts. As pointed out by some researchers [10, 11], previous second/foreign language (L2/FL) research on consonant length contrasts mostly focused on non-native learners of Japanese from English [12, 13], Korean [14] or, more recently, Cantonese [15, 16] backgrounds. We aim to increase our current understanding on contrastive length perception by focusing on the learner population from a different L1 background that has been substantially growing. Furthermore, due to recruitment difficulties, research involving highly advanced L2/FL learners in languages other than English is still limited. We hope to fill this gap by examining if and how advanced Vietnamese learners perceive Japanese singleton/geminate contrasts.

The questions addressed in this study are as follows. Is there a difference across four (three Vietnamese and one Japanese) groups of participants in their perception of...
Japanese consonant length? Specifically, do the two groups of learners, one in Vietnam (less experienced/proficient) and the other in Japan (more experienced/proficient), differ from each other on the one hand and from the native Japanese speakers on the other hand?

2. Method

2.1. Stimuli preparation

2.1.1. Speakers

The experimental stimuli and procedure were identical to those used in our previous research ([17]). 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 CV/CVCV/ tokens contained singleton (n = 96) or geminate (n = 96) consonants intervocically (underlined and bolded). As voiced geminates are limited in Japanese ([18, 19]), 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/) and 2.8 for velars (/k/), respectively. These durational values are in good agreement with what has been reported in previous research [13].

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

<table>
<thead>
<tr>
<th>Singleton</th>
<th>Geminate</th>
</tr>
</thead>
<tbody>
<tr>
<td>/t/</td>
<td>keta ‘unskilled’</td>
</tr>
<tr>
<td>kato ‘transient’</td>
<td>katto ‘cut’</td>
</tr>
<tr>
<td>mato ‘wait’</td>
<td>matto ‘waiting’</td>
</tr>
<tr>
<td>oto ‘sound’</td>
<td>otto ‘husband’</td>
</tr>
<tr>
<td>sate ‘well, then’</td>
<td>sate ‘leaving’</td>
</tr>
<tr>
<td>wata ‘cotton’</td>
<td>watta ‘broke’</td>
</tr>
<tr>
<td>/k/</td>
<td>eke ‘open’</td>
</tr>
<tr>
<td>haka ‘grave’</td>
<td>hakka ‘mint’</td>
</tr>
<tr>
<td>lka ‘below’</td>
<td>llka ‘lesson one’</td>
</tr>
<tr>
<td>kako ‘past’</td>
<td>kakko ‘parenthesis’</td>
</tr>
<tr>
<td>saka ‘slope’</td>
<td>sakka ‘author’</td>
</tr>
<tr>
<td>shike ‘rough sea’</td>
<td>shikke ‘humidity’</td>
</tr>
</tbody>
</table>

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 ______ carrier’). The pace of presentation was managed manually 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

Three groups of native Vietnamese speakers participated in this study. The first group of 12 (6 males, 6 females, mean age = 21.0 years, sd = 2.3) consisted of university students in Ho Chi Minh City and had no Japanese experience. Seven of them were born in Ho Chi Minh City and the rest in other locations in the South region of the country (e.g., Ben Tre, Dong Nai, Kien Giang, Tay Ninh, Tien Giang). The next two groups, one living in Ho Chi Minh City, Vietnam (7 males, 10 females, mean age = 21.5 years, sd = 4.4) and the other in Tokyo, Japan (5 males, 8 females, mean age = 28.6 years, sd = 5.1) were learners of Japanese. The 13 learners in Japan had all passed the JLPT at N1 and were considered highly advanced learners of Japanese. Their mean length of residence (LOR) in Japan was 8.4 years (sd = 3.4) at the time of participation. They started learning Japanese at the mean age of 17.4 years (sd = 4.1). They participated in the study at a university in Tokyo. Broadly speaking, their places of origin in Vietnam were 7 from North, 1 from Central and 5 from South. Compared to the nominally advanced learners included in previous research [12, 14], it is expected that the advanced learners in the present study were more experienced as well as immersed in the Japanese-speaking environment.

The 17 learners in Vietnam were students at a university in or around Ho Chi Minh City and were at the JLPT N3 level. Except for 2 participants, all were studying Japanese in the same program at the same university and were considered (pre-)intermediate learners. Nine of them were born in Ho Chi Minh City and the rest in the South/Central regions of the country (e.g., Ben Tre, Dong Nai, Gia Lai, Khanh Hoa, Kien Giang, Nha Trang, Quang Nam). None of them had lived overseas for an extended period of time (i.e., more than 6 months).

The last and a control group consisted of 10 (2 males, 8 females) native speakers of Japanese 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 LOR 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 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 [20]. 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’-‘yoka’, ‘yoka’-‘yokka’), or B (e.g., ‘sotto’-‘sotto’-‘sotto’, 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 [21]. 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 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 4.3.1 for statistical analyses and data visualization reported below [22]. The packages used include ez [23] and tidyverse [24].

3.1. Overall

The overall mean discrimination accuracy was 67%, 80%, 91% and 99% for the non-learner group, two groups of learners (Vietnam, Japan) and the control Japanese group, respectively (Figure 1). The Japanese group was at near ceiling with little individual variation. The two groups of learners clearly outperformed the non-learners in perceiving Japanese consonant length contrasts. Regardless of Japanese experience, all Vietnamese groups were apparently much more variable than the Japanese group. While none of the 12 non-learners reached the range of discrimination accuracy set by the L1 Japanese group (96 - 100%), at least one of the participants in the learner groups reached the L1 Japanese range (i.e., 1 out of 17 (0.6%) in Vietnam and 5 out of 13 in Japan (38%), respectively).

One-way analysis of variance (ANOVA) with Group (non-learner, learner in Vietnam, learner in Japan, L1 Japanese) reached significance \( F(3, 48) = 23.6, p < .001, \eta^2_p = .60 \). According to the post-hoc t-tests, all between-group differences were statistically significant \( t(11.3 - 26.6) = -5.9 \) – 8.7, \( p < .05 \) as clearly shown in Figure 1.

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

Next, we examined if the overall pattern of results remained the same for different types of sounds, i.e., stop length contrasts at different places of articulation. Table 2 and Figure 2 show length discrimination accuracy for the four groups of participants when the place of articulation (Alveolar, Velar) of the target token (i.e., X in the AXB sequence) was taken into consideration.

![Figure 2: The distributions of length discrimination accuracy (%) by four groups of participants for trials differing in the place of articulation (Alveolar, Velar) of the target token. The light lines connect individual participants’ scores.](image)

The pattern of place effect differed for different groups. As shown in Table 2 and Figure 2, the two groups in Vietnam were differentially affected by the place of articulation when the target token was geminate (non-learner: 70% for Alveolar vs 64% for Velar, learner in Vietnam: 82% for Alveolar vs 78% for Velar) as opposed to when it was singleton (non-learner: 67% for Alveolar vs 69% for Velar, learner in Vietnam: 80% for both Alveolar and Velar). This pattern was absent in the other two groups with greater Japanese experience. Specifically, the advanced learners in Japan diverged from the participants in Vietnam and resembled the L1 Japanese speakers by not showing the differing effect of the place of articulation.

Table 2: Mean length discrimination accuracy (%) by four groups of participants for trials differing in the place of articulation (Alveolar, Velar) of the target token. Standard deviations are in parentheses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Geminate</th>
<th>Velar</th>
<th>Alveolar</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-learner</td>
<td>70 (12.2)</td>
<td>64 (15.4)</td>
<td>67 (15.3)</td>
<td>69 (11.2)</td>
</tr>
<tr>
<td>Learner (Vietnam)</td>
<td>82 (13.3)</td>
<td>78 (13.6)</td>
<td>80 (13.0)</td>
<td>80 (11.1)</td>
</tr>
<tr>
<td>Learner (Japan)</td>
<td>95 (5.0)</td>
<td>88 (9.8)</td>
<td>93 (8.1)</td>
<td>91 (8.6)</td>
</tr>
<tr>
<td>Japanese</td>
<td>99 (1.9)</td>
<td>99 (2.5)</td>
<td>100 (0.6)</td>
<td>98 (2.6)</td>
</tr>
</tbody>
</table>

Three-way ANOVA with Group (non-learner, learner in Vietnam, learner in Japan, L1 Japanese), Length (Geminate, Singleton) and Place (Alveolar, Velar) reached significance.
for the main effects of Group $[F(3, 48) = 23.1, p < .001, \eta_G^2 = .54]$, Place $[F(1, 48) = 10.9, p < .01, \eta_G^2 = .013]$ and an interaction between Length and Place $[F(1, 48) = 10.4, p < .01, \eta_G^2 = .010]$. The overall pattern of between-group differences seen in Figure 1 was retained in Figure 2 with the L1 Japanese group being the most and the non-learner group the least accurate. While all between-group differences reached significance as shown above for the overall results (Figure 1), when the place effect was taken into account, the advanced learners significantly differed from the L1 Japanese speakers only for the trials in which the velar stops occurred in the target position ($\eta(200) = 2.9, p < .05$) (Advanced learners: 89% vs L1 Japanese speakers: 98%). As for the two-way interaction between the Length and Place factors, the slight influence of place was limited to the target geminate. The participants were more accurate when the target token was alveolar geminate (86%) than when it was velar geminate (81%). The Place effect was non-existent for the target singleton (84% for both alveolar and velar).

4. Discussion
This study examined the perception of short/singleton vs long/geminate consonants in Japanese by three groups of native Vietnamese speakers differing in Japanese experience and a control group of native Japanese speakers. One of our motivations for conducting this study came from the fact the number of Vietnamese learners of Japanese in and out of Japan has been dramatically increasing in recent years [6]. Despite this, information regarding their acquisition/processing of Japanese sounds has not yet reached a wider audience in the field.

Another reason is based on the cross-linguistic phonetic differences between Japanese and Vietnamese which include that, unlike Japanese, Vietnamese does not use consonant length contrastively and the role of vowel length is also limited in this language [4, 5]. We reasoned that the cross-linguistic phonetic characteristics would make it difficult for Vietnamese speakers to perceive Japanese singleton/geminate contrasts at least initially.

Our main findings were: 1) clear and expected between-group difference based on Japanese experience and 2) different influence of the consonant type (i.e., place of articulation) on the discrimination of the singleton/geminate length contrasts. Despite the lack of consonant length at the word level in their L1 Vietnamese, the advanced learners were more successful in their perception of Japanese singleton/geminate contrasts and resembled the native Japanese speakers to a greater extent than the Vietnamese speakers who were less experienced in Japanese. It needs to be pointed out that, whilst not completely native-like, the advanced learners’ performance was closer to that of the native Japanese speakers than to that of the Vietnamese speakers with less or no experience in Japanese. Undoubtedly, the advanced learners in this study benefited from L2/FL Japanese learning unlike the learners in some previous research [3, 4, 16]. The slight but persistent difference between the advanced learners and Japanese speakers suggests genuine difficulty of Japanese consonant length.

While we observed clear between-group differences in the perception of Japanese consonant length, some limitations regarding characteristics of the learners need to be acknowledged. Specifically, extensive L2/FL experience and advanced proficiency were combined in this study. In other words, the learners in Japan had the greater advantage compared to the learners in Vietnam by living in Japan for an extended period of time (more than 8 years) and being at the advanced proficiency level. Thus, their chance of success was expectedly high. It would be necessary to separate these factors to understand if and how experience and proficiency may explain the pattern of results.

Another procedural limitation was that we employed the discrimination task only, because we included participants who were naive to Japanese, for whom length categories for consonants did not exist. However, identification of vowel and consonant length is a necessary skill for learners who need to learn to categorize length contrasts to efficiently communicate in Japanese. To gain an insight into how Vietnamese speakers process and acquire this essential skill, their perceptual assimilation patterns of Japanese to Vietnamese sounds need to be empirically established.

5. Conclusions
A clear difference between the two learner groups in Japan and Vietnam who shared the L1, i.e., Vietnamese, but differed in their experience with Japanese demonstrates learnability of Japanese consonant length beyond early childhood in an immersion setting. We observed that it is possible for some non-native learners to perceive difficult L2/FL sounds as accurately as native speakers in certain contexts. At the same time, a non-negligible difference between the highly advanced learners and native Japanese speakers confirms genuine difficulty of Japanese consonant length, which was reported in numerous previous studies.

6. Acknowledgements
Portions of this study were presented at the Second International Conference on Tone and Intonation (TAI 2023) in Singapore. Part of the data collection 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 constructive comments and suggestions for improving this work.

7. References


