Acquisition of Japanese quantity contrasts by L1 Cantonese speakers

Albert Lee

University of Tokyo

The Chinese University of Hong Kong

Peggy Mok

The Chinese University of Hong Kong

Department of English, University of Tokyo, Meguro-ku, Tokyo 153-8902, Japan; Department of Linguistics and Modern Languages, Leung Kau Kui Building, The Chinese University of Hong Kong, Shatin, Hong Kong.

The authors thank Prof. Yukari Hirata (Colgate) for invaluable comments on some of the findings presented below and Prof. Chun-Fat Lau (Xiamen) for advice on Hong Kong Hakka phonology.

Correspondence should be directed to Peggy Mok by email at peggymok@cuhk.edu.hk.

Abstract

This paper explores the acquisition of Japanese vowel and consonant quantity contrasts by Cantonese learners. Our goal is to examine whether transfer from L1 is possible when L1 experience is phonemic but restricted to a small set of sounds (short vs. long vowels) and when the experience is non-phonemic, derived only at morpheme boundaries (short vs. long consonants). We recruited 20 Cantonese learners (beginner and advanced learners) and 5 native speakers of Japanese, who produced target stimuli varying in consonant and vowel quantity framed in a carrier sentence. The resultant data were converted into several durational ratios for analyses. Results showed that both the beginners and advanced learners were able to distinguish between short vs. long vowels and consonants in Japanese, but only the native speakers enhanced the contrasts in slower speech. It was also found that in most cases the learners were able to lengthen the vowel before a geminate (i.e. long consonant), a secondary cue to Japanese consonant quantity known to be rare across languages. These results are discussed in terms of current theories of second language acquisition.

By now there is a vast body of literature on the acquisition of second language phonology (see Best & Tyler, 2007; Strange & Shafer, 2008 for a review). It is generally assumed that the reconfigured perceptual system as a result of L1 acquisition acts as a filter when processing L2. Currently prevailing theories of L2 speech perception are generally based on this view, including the Native Language Magnet model (NLM, Kuhl, 2000; Kuhl et al., 2008; Kuhl & Iverson, 1995), the Speech Learning Model (SLM, e.g. Flege, 1995), the Perceptual Assimilation Model (PAM, e.g. Best, 1995; PAM-L2, e.g. Best & Tyler, 2007), and the Second Language Linguistic Perception Model (L2LP, Escudero Neyra, 2009; van Leussen & Escudero Neyra, 2015). The direct implication of these theories is that L2 speech sounds are mapped onto the L1 phonetic categories that are acoustically or articulatorily similar. The classical theories based on the 'L1 category filter' insight have elegantly captured the difficulties in learning non-native speech sounds commonly encountered by L2 learners, such as the r/l distinction for Japanese learners of English (Flege, Takagi, & Mann, 1995).

A recent focus in the L2 literature is the role of features in transfer. In this line of research, L1 transfer is deemed to take place at the featural level, rather than the phonemic level. In other words, 'an L2 contrastive category will be difficult to acquire if it is based on a phonetic feature not exploited in the L1 to signal phonological contrast' (McAllister,

Flege, & Piske, 2002, p. 231). For example, in McAllister et al. (2002), native speakers of American English, Latin American Spanish and Estonian participated in a production test and a perception test to have their mastery of Swedish quantities assessed. All subjects were L2 learners of Swedish who had lived in Sweden for over 10 years and reported to use Swedish frequently. Results showed that the Estonian subjects, whose L1 has quantity distinctions based on duration, performed much like the Swedish controls. As expected, the English and the Spanish subjects, who had no comparable quantity distinctions in their L1, performed less well. Interestingly, the English subjects showed slightly better performance than their Spanish counterparts despite the absence of pure duration-based quantity contrasts in English. The short vs. long distinction in English vowels is marked by both vowel quality as well as duration. Whereas duration is not the only cue to English vowel quantity, listeners appear to be able to identify a vowel on the basis of duration (Whalen, 1989). As such, the partial use of the temporal dimension in English was deemed the reason for the better performance of the English subjects in McAllister et al (2002), compared to the Spanish subjects who did not make use of the same dimension.

Other studies that support the feature hypothesis include Brown (2000) and Pajak and Levy (2014). In particular, Pajak and Levy (2014) compared Korean, Vietnamese, Cantonese and Mandarin listeners in an AX discrimination task using Polish-like nonce

words that contrasted in consonant length. All participants also spoke English as L2. In these languages (except Mandarin), phonemic quantity is contrastive but informative to different degrees. Korean has length contrasts in all vowels (1999)[1], while long consonants both occur underlyingly in the lexicon and are 'derived' (in the sense that a sequence of two identical consonants occur on either side of a morpheme boundary, also à la Kubozono, 2017). In Vietnamese, length is contrastive in two sets of vowels. In Cantonese, there are vowel quantity contrasts to which duration is one cue alongside vowel quality. The authors predicted that either all length-experienced participants would pattern together, in which case Korean, Vietnamese, Cantonese >> Mandarin, or that a gradient pattern would emerge based on how informative duration is in each language, i.e. Korean >> Vietnamese, Cantonese >> Mandarin. However, results showed that both Korean and Vietnamese speakers outperformed Cantonese speakers, who in turn performed better than their Mandarin counterparts, i.e. Korean, Vietnamese >> Cantonese >> Mandarin. The difference between Vietnamese and Cantonese participants was hypothesized to be due to the fact that duration is but one cue to length contrasts in Cantonese whereas in Vietnamese duration plays a bigger role.

[1] Note, however, that currently a vowel length contrast merger is taking place (Kang, Yoon, & Han, 2015).

While Pajak and Levy's results would lead to the view that the use of duration in vowel quantities can be transferred to L2 consonant quantities, it is equally possible that the Cantonese participants' performance was due to the fact that there are derived geminates in their L1; Tsukada et al's (2014) findings would suggest the latter is the case. In a two-alternative forced-choice AXB task, they showed that compared to American English, Thai and Japanese listeners, native Italian listeners performed the worst in identifying Japanese vowel length, despite their heavy use of duration in L1 consonantal quantity contrasts. Meanwhile, that American English listeners performed better is consistent with the fact that there are short vs. long vowels in English to which duration is one cue. If L1 consonantal quantity contrasts do not benefit the acquisition of L2 vocalic quantity contrasts (i.e., Italian), it follows that the performance of the Cantonese participants in Pajak and Levy (2014) should be attributed to the derived geminates in Cantonese, rather than to the partial use of vowel quantity contrasts. By implication, one would expect that Cantonese L2 learners can easily acquire short vs. long consonants in Japanese. More details of Cantonese phonology will be introduced in the following sections.

The main goal of this paper is to contribute to the category vs. feature dialogue by looking at the production of Japanese phonemic quantities by Cantonese learners. Japanese has both vowel and consonant quantity contrasts. While secondary cues abound, local

duration is unarguably the primary cue to quantities in Japanese. That Japanese relies on duration to signal quantity contrasts for both consonants and vowels makes it the perfect testing ground for our research question, namely L1 benefits in the durational dimension. Like in English, Cantonese has short vs. long vowel distinctions which are not solely based on duration, as well as derived geminates when a stop coda is immediately followed by a homorganic obstruent. The performance of our participants will thus shed light on the benefits of L1 transfer where the target phonetic dimension (duration) is used only to a limited extent. Our results will also hinge upon issues of L1 benefits in secondary cues (to vowel identity) and the possible influence of typological tendency on L2 production (of geminates). In the rest of this Section our research questions will be motivated in more detail.

Phonetics of Japanese quantities

In Japanese, both consonants (e.g. *kita* 'came' vs. *kitta* 'cut') and vowels (e.g. *kita* 'came' vs. *kiita* 'heard') contrast in quantity[2]. Cues to the short vs. long vowel distinction

[2] Note that in most theories of formal phonology long vowels and geminates have structurally different representations. See Labrune (2012) for a recent introduction in the Autosegmental framework and Yoshida (1990) for a treatment in 'Standard' Government Phonology. However, since the focus of the present study is on the phonetics of quantity, the formal issue of syllable structure will not be addressed.

include vowel and word duration (Hirata, 2004), formant frequency (Hirata & Tsukada, 2009) as well as f_o contour (Takiguchi, Takeyasu, & Giriko, 2010), whereas the long (geminate) vs. short (singleton) consonant distinction is associated with closure duration (Han, 1992), duration of the vowels surrounding the closure (Han, 1994) and apparently intensity and f_o range (Ofuka, 2003).

For vowels, Hirata (2004) looked at the production of four native speakers speaking at various speech rates and compared the effectiveness of a range of measurements (absolute duration vs. duration ratios). Some of these durational ratios will be taken as reference values in this paper (summarized in Table 2 and Table 7). For non-durational cues, Takiguchi et al (2010) reported that f_o contour affected perception only when durational cues were ambiguous and that the effects of rising and falling contours were asymmetric. Specifically, native Japanese listeners tended to perceive a vowel as long when hearing a falling f_o contour compared to hearing a level contour. At the segmental level, Hirata and Tsukada (2009) found that long vowels occupied the peripheral portion of the F1~F2 space whereas short vowels were found in the inner regions, but this difference became less distinct in slow speech.

For consonants, Han (1992) observed a Closure Duration Ratio (singleton:geminate) of 1:2.8 (or 1:2.4 in Toda, 2003), together with a shorter VOT for geminates. For non-local

durational cues, Han (1994) and Idemaru and Guion (2008) observed that the vowel before a geminate (V1) is longer than that preceding a corresponding singleton, whereas the vowel following a geminate (V2) is shortened. This pre-geminate lengthening of V1 is interesting as it violates a typological tendency that vowels are shorter before a long consonant (Maddieson, 1985) . Kingston et al (2009) found that Japanese listeners tended to judge a consonant as 'long' if the preceding vowel was longer, whereas the opposite was true for Norwegian and Italian listeners. Surprisingly, English listeners, who have no underlying geminates in their L1, showed the same pattern as Japanese listeners. As both Cantonese and English have derived geminates, it would be interesting to see whether Cantonese learners of Japanese can acquire pre-geminate lengthening against this typological tendency.

Previous work on L2 acquisition of Japanese quantities

The acquisition of Japanese quantity by L2 learners has been extensively studied, with Mandarin, American English and Korean learners among the most investigated so far (see Hirata, 2015 for a comprehensive review). In works looking at perception, considerable cross-study variation is observed. For example, in Hirata and Lambacher (2004), the presence of a carrier sentence was found to help distinguish long vs. short vowels; while Motohashi-Saigo and Hardison (2009) found no such effect. As expected, learners whose L1 does not have quantity contrasts encounter difficulty identifying

Japanese quantities (e.g. Kurihara, 2004 where Chinese learners tend to judge vowels as long). There are also other factors that affect L2 learners' perception, such as lexical pitch accent (Minagawa & Kiritani, 1996; Minagawa, Maekawa, & Kiritani, 2002), nature of the experimental task (Tsukada, 2011) and training (Motohashi-Saigo & Hardison, 2009; Tajima, Kato, Rothwell, Akahane-Yamada, & Munhall, 2008).

Studies looking at the production of L2 Japanese quantities offer a different perspective on the problem. On the one hand, learners encounter difficulty producing these contrasts, showing the effect of L1 category filter. For example, Kurihara (2005) found that Chinese beginner learners tended to shorten long vowels, whereas advanced learners tended to lengthen short vowels; but both groups tended to erroneously lengthen singleton consonants. On the other hand, in production experiments where explicit instructions to make quantity distinctions were given, learners appeared to be able to use duration to contrast quantities at least to some extent. Han (1992) found that American learners' closure duration ratio of Japanese singleton vs. geminate consonants was 1:2, compared to 1:2.8 in the case of native Japanese speakers (or 1:2.4 in Toda, 2003). Even though there was a gap between the American learners and native speakers as suggested by these ratios, insofar as the distinction between long and short is concerned the learners' production was satisfactory. Such a mixed picture painted by only a handful of previous studies calls for

further investigation into this phenomenon in a language that is different from English, Mandarin or Korean.

The Japanese speech of Cantonese L2 learners is an understudied topic, with few systematic studies available. Lai (1999) provided a mainly qualitative account of Cantonese vs. Japanese prosody, forming the basis of subsequent works on this subject. She suggested that Cantonese learners of Japanese would erroneously lengthen short syllables because all Cantonese syllables are long. Sagayama (2010) was the first comprehensive production study looking at Cantonese L2 learners of two proficiency levels. All of the six speakers in Sagayama (2010) were from the same class, but were put into two groups based on their pronunciation. Using measures of central tendency, Sagayama (2010) observed random production in the less native-like group (N = 3) and good but hyper-corrected production in the more native-like learners (N = 3). Building on the foundation of these previous works, the present study will revisit Cantonese learners' production with control over subjects' proficiency in terms of year groups as well as speech rate. Including speech rate in our design may reveal useful insights into the learners' production given the known effect of speech rate on foreign accentedness and comprehensibility ratings in perception (Munro & Derwing, 2001). In-depth statistical analyses will also be conducted to illuminate any interactions between factors.

Although both consonant and vowel quantities are cued by duration in Japanese, there are differences in how duration is exploited to cue the two types of contrasts. Some of these strategies, as discussed above, are used also in Cantonese (e.g. vowel duration, Kao 1971) while others are not (e.g. pre-geminate lengthening of vowels). The different uses of duration in these contrasts may give us useful insights into our research questions. As mentioned above, one secondary cue to Japanese geminates is a lengthened V1, which however violates a universal trend reported in Maddieson (1985). We thus also intend to examine whether the typologically anomalous nature of Japanese geminates would render it less successfully acquired by Cantonese learners, compared to vowel quantities. Since we seem to have a good source of L1 transfer for both consonantal (cf. Pajak & Levy, 2014) and vocalic (cf. McAllister et al., 2002) quantities, if the former turns out to be less successfully acquired, the discrepancy could stem from this typological tendency, pre-geminate V1 lengthening may in some way be a hard-to-acquire phonetic feature.

Phonology of Cantonese

Cantonese is relevant to the study of L2 quantity contrasts because there are short vs. long consonants and vowels but only to a very limited extent, making it an interesting test case for studying the transfer of L1 benefits. According to Yip (1993, p. 265), in a Cantonese syllable '(c)odas may be… unreleased stops (p, t, k). Open syllables always have

long tense vowels... The low back vowels contrast in length (or tenseness). All other vowels have long and short allophones, conditioned by choice of coda consonant in complex ways.' There are vowel pairs (e.g. /a:i/ vs. /ɐi/) that contrast in length (e.g. /ka:i/ 街 '*street* vs. /kɐi/ 雞 '*chicken'*), but they also differ in vowel quality (Kao, 1971). Hence there are no true minimal vowel contrasts based on duration only in Cantonese. For consonants, although there are no underlying geminates in Cantonese, there are the 'cat tail' type derived geminates (e.g. /p^ha:.k^hɵy/ 怕佢 *'afraid of him'* vs. /p^ha:k.k^hɵy/ 拍佢 *'tap him (e.g. on the shoulder)'*) given Cantonese allows an unreleased stop coda in its syllable structure. Comparable examples of 'cat tail' geminates in English include *mi**dd**ay* and *oran**ge j**uice*. These partial uses of quantity contrasts beg the question of whether Cantonese speaking learners of Japanese could distinguish *kita* vs. *kitta* vs. *kiita*.

Although our learners also speak Mandarin and English, here we treat Japanese as an L2 (instead of L3, i.e. third and subsequent languages). Table 1 summarizes the phonetic cues to quantity contrasts in the languages spoken by the participants (i.e. L1: Cantonese, L2: English, Mandarin, Japanese). In terms of the overall direction of transfer, Cantonese and English (Roach, 2004) are consistent for both vocalic and consonantal quantities, whereas Mandarin makes no quantity distinction at both phonemic and phonetic levels. Thus we can assume that the multilingual background of the participants would not

confound our findings. As speaking more than two languages is increasingly the norm, our learners represent an ecologically realistic case where learners have extensive prior experience with foreign language learning.

Insert Table 1 about here

Research questions

Here we seek to test several hypotheses. Firstly, although Sagayama (2010) found that her less native-like speakers produced short vs. long vowels randomly whereas her more native-like speakers tended to exaggerate the contrast, her study was based on the same group of students put into two categories by the experimenter herself, thus not comparable to the two proficiency groups in the present study. On the other hand, since the English participants in McAllister et al. (2002) benefit from the partial use of duration in their L1 vocalic quantity contrasts, it is reasonable to assume that Cantonese learners can make use of the same L1 knowledge in their acquisition of L2 Japanese vowel categories. We thus hypothesize that our **(H1a) Beginner group will show evidence of some ability to distinguish Japanese short vs. long vowels** whereas our **(H1b) Advanced group will distinguish Japanese short vs. long vowels more similarly to native speakers**. Support

for H1a would serve as direct evidence for L1 transfer where a phonetic dimension is used only to a very limited extent (i.e. one of multiple cues and used in only a subset of vowels), whereas support for H1b would show that this L2 phonetic dimension is ultimately learnable through such means as classroom instruction and immersion in Japan. For consonantal quantities, since Cantonese listeners were found to be sensitive to non-native consonant length contrasts in Polish nonce-words (Pajak & Levy, 2014), we hypothesize that our **(H2a) Beginner group will show evidence of some ability to distinguish Japanese short vs. long consonants** whereas our **(H2b) Advanced group will distinguish Japanese short vs. long consonants more similarly to native speakers**. Support for (H2a) will strengthen the view that non-phonemic use of duration in L1 (i.e. derived geminates) can be transferred to the acquisition of L2 categories. Further, based on Kingston et al's (2009) observation about English listeners' response to the duration of pre-geminate vowel duration, we hypothesize that **(H3) our learners will be able to lengthen V1 before a geminate**. Failure to replicate (H3) would lead to the conclusion that pre-geminate lengthening of V1 is hard to acquire, possibly related to the typological tendency reported in Maddieson (1985).

Methods

Speakers and materials

We conducted a production study with five native speakers of Japanese as controls (three male, mean age = 31.0, SD = 10.6), 10 advanced learners (two male, mean age = 21.2, SD = .42) in their final year of the BA Japanese Studies programme at the Chinese University of Hong Kong and 10 beginners (three male, mean age = 18.2, SD = .42) who were in their first year of the same programme. It was not possible to recruit more learners as the annual intake of the degree programme was only about 20 students and the beginner participants were required to be genuine beginners. The Advanced group had stayed in Japan for one year as exchange students; otherwise none of the learners had any experience living in a foreign country. Both learner groups were native speakers of Hong Kong Cantonese, speaking English and Mandarin as L2. The learners' English proficiency all reached the admissions requirement of the Chinese University of Hong Kong (e.g. IELTS 6.0 / TOEFL iBT 80 / HKDSE Level 3), whereas for Mandarin there is no uniform score to objectively measure their proficiency. While successfully controlling for proficiency in terms of formal instruction input (i.e. year group), admittedly some variations in the subject pool had to be tolerated. Some learners started their degree programme without any knowledge of Japanese, while others had some knowledge of the *hiragana* syllabary. One learner was a parallel bilingual in Cantonese and Hakka (which has derived geminates like Cantonese but no vocalic quantity contrasts). Another one attended an international school

and self-identified as a near-native speaker of English. Otherwise, the learners in the two groups had relatively uniform language backgrounds. All participants reported no history of speech and hearing impairment. Other information of the participants can be found in Appendix A. A version of this data set was reported in Lee and Mok (2016).

During the experiment we noticed that the pronunciation of two subjects (B8 and B9) in the Beginner group was unusually accurate. They later admitted that they had learnt Japanese prior to their degree study (having respectively passed the N2 and N1 levels of the Japanese Language Proficiency Test), despite our requirement that speakers in the Beginner group should be genuine beginners in their first year of the programme. For this reason, we reclassified these two speakers as Advanced, leaving us with 12 subjects in the Advanced group, eight in Beginner group and five in Native. Most other learners in the Beginner group reportedly had no knowledge of Japanese or at best just some knowledge of *hiragana* when they entered university.

Given the known differences between real words and non-words in durational variability reported in Hirata (2004), and that non-words have not been investigated in the speech of Cantonese learners (Lai, 1999; Sagayama, 2010), in the present study both word types were included for the sake of comprehensiveness. Examining non-words also allows further verification of whether the learners can generalize their ability to distinguish

between long and short sounds to words that they have not encountered. A total of 27 (quasi-)real Japanese words and 18 non-words were used as stimuli (see Appendix B). They contrast in vowel and consonant quantity (CVCV, CVVCV, CVCCV). All real words were displayed in the *kana* syllabaries (*hiragana* or *katakana*) as well as *kanji* characters where applicable while non-words were presented in *katakana*. The writing system of modern Japanese comprises two types of characters, namely logographic *kanji* characters which were adopted from Chinese characters, and moraic *kana* characters which in turn consist of two syllabaries: *hiragana* and *katakana*. To obtain true minimal triplets, infrequent words, some place names and personal names had to be used. Likewise, a small part of the non-words could also be construed as meaningful by some native speakers. Following Beckman (1982a, 1982b), the effect of lexical pitch accent on duration was deemed insignificant and thus was not controlled in our stimuli.

Procedures

Recording took place in a quiet room at the Chinese University of Hong Kong, using a Zoom H2n voice recorder. Stimuli were presented on a computer screen using a Javascript-based sentence randomizer. Speakers were briefed about the experimental task and granted their written consent before recording commenced. Speakers were to say the target words in the carrier sentence *Kore-wa* XX *desu* 'This is XX'. Utterances were

collected over six randomized blocks, namely Real Word normal⇒Slow⇒Fast⇒Non-word normal⇒Slow⇒Fast. Speech rate can be controlled in either (near-)absolute or relative terms, and in this paper we opted for the latter (also cf. Hirata 2004). Participants were instructed to speak obviously more slowly in the slow production, and obviously faster in the fast production, both relative to the normal speech rate. Had it been controlled in absolute terms, say, by imitation or following a metronome, speakers' attention would have been distracted to adhering to the precise speeds which may incur unnaturalness in their speech production. For the non-word blocks, speakers were instructed to use the high-low accent pattern. Within each block, each word appeared three times. Altogether, 15 roots (9 for real words and 6 for non-words) × 3 quantities × 3 speech rates × 3 repetitions × 25 speakers (5 native + 8 Beginner group +12 Advanced) = 10,125 utterances were collected. Two utterances were discarded due to mispronunciation, leaving us with 10,123 utterances for acoustic analysis. No other data were removed as outliers in subsequent statistical analyses.

Speech data were manually labeled by the segment (consonants and vowels) using FormantPro (described in Cheng & Xu, 2013; Chiu, Fromont, Lee, & Xu, 2015). It is a Praat (Boersma & van Heuven, 2001) script for extracting formant trajectories, as well as intensity and duration values. Since all target words were disyllabic, four segments

(henceforth $C_1V_1C_2V_2$) were labelled. Vowel boundaries were located at the onset and offset of periodicity in the waveform; when preceded by a nasal consonant (i.e. /m/ or /n/), the left edge of the vowel is where abrupt spectral changes associated with closure release were observed. Subsequently, for each labelled interval FormantPro extracted the duration and mean intensity values as well as time-normalized formant values. Then the extracted duration values were converted into several duration ratios used in previous studies, summarized as follows:

Insert Table 2 about here

Results

Average syllable duration

First the mean syllable duration of all target words was checked to assure that speakers' speech rates differed according to the appropriate mode. Figure 1 showed that in all speaker groups, average syllable duration was the shortest in fast speech and the longest in slow speech. A two-way ANOVA was performed with Group (Advanced, Beginner, Native) and Rate (Fast, Normal, Slow) as fixed factors. There were significant main effects of Group ($F(2,3366) = 109.0$, $p < .001$) and Rate ($F(2,3366) = 2161.0$, $p < .001$) as well as a

significant interaction between them ($F(4,3366) = 2.6$, $p = .033$). Post-hoc Bonferroni tests confirmed that Fast speech had shorter syllable duration than Normal speech, which in turn was shorter than Slow speech in mean syllable duration (all $p < .001$). It is thus safe to conclude that for all speaker groups, any significant effects of speech rate observed in subsequent analyses are reliable. Overall, the average syllable duration of the Native group was 16 ms shorter than the Advanced group, whose syllable duration in turn was shorter than that of the Beginner group by 18 ms. All speaker groups were significantly different from one another in post-hoc Bonferroni tests (all $p < .001$).

Insert Figure 1 about here

Short vs. long vowel

Following Hirata (2004), here we compared short vs. long vowels in terms of V1 Duration Ratio, Word Duration Ratio and Vowel-to-Word Duration Ratio. In our data, V1 Duration Ratios of the Native, Advanced and Beginner groups were respectively 1:2.24, 1:2.01 and 1:1.92. A V1 Duration Ratio greater than 1:1 means that long vowels are longer than short vowels. There is also the 1:2.51 line in Figure 2 for reference, which is the value of the same ratio reported by Hirata (2004) for accented vowels (or 1:2.22 for unaccented

vowels in her study). As is clear from this diagram, for all speaker groups V1 Duration Ratio far exceeded the 1:1 threshold (grand mean = 1:2.03, SD = .58), suggesting that everyone, including learners in the Beginner group, was able to distinguish between long and short vowels. In addition, speech rate appears to affect V1 Duration Ratio in the Native group but not in the learner groups. As the native speakers moved from fast speech to slower speech, V1 Duration Ratio increased; but this pattern was not consistently observed in either of the learner groups, especially for non-words. All individual speakers exceeded the 1:1 threshold in all speech rate and word type conditions (range 1.17~3.10).

Insert Figure 2 about here

The same holds true for Word Duration Ratio (Figure 3). Here if the ratio exceeds 1:1, a CVVCV word is longer than a CVCV word. The 1:1.4 reference is adapted from Hirata (2004), where the Word Duration Ratio of CVCV:CVVCV is 2:2.7~2.95 (i.e. ~2:2.8, and halved for better comparability with other duration ratios, thus 1:1.4). In our data, the mean Word Duration Ratios were 1:1.34 for the Native group, 1:1.27 for the Advanced group and 1:1.24 for the Beginner group. Hence, for all speaker groups the duration of CVVCV words was longer than CVCV words. For native speakers, again, slow speech had

the effect of enhancing the short vs. long contrast, but the same effect was not observed in the learner groups. All individual speakers exceeded the 1:1 threshold in all speech rate and word type conditions (range 1.07~1.66).

Insert Figure 3 about here

Linear regression analysis revealed a significant positive correlation between mean syllable duration and V1 Duration Ration ($r = .392$, $p < .001$) and Word Duration Ratio ($r = .236$, $p < .001$) for native speakers, confirming that as one speaks slower the contrast between short vs. long vowels becomes greater. For the learner groups, mean syllable duration was not significantly correlated with V1 Duration Ratio whereas Word Duration Ratio was inversely correlated with mean syllable duration for both the Advanced group ($r = -0.179$, $p < .001$) and the Beginner group ($r = -0.123$, $p = .020$). This shows that for the learners, short vs. long vowels tended to become less distinct in terms of word duration in slower speech. In other words, although the learners successfully distinguished short vs. long vowels, they were using a strategy different from that of the native speakers across speech rates.

Further analyses were conducted using mixed-effects models with crossed random effects for subjects and items using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015, version 1.1-12) of R (R Core Team, 2016, version 3.3.1). The analyses included a treatment-coded fixed effect of Rate (baseline = Normal), Helmert-coded fixed effect of Group (Advanced = '-1/3,1/2', Beginner = '-1/3,-1/2', Native = '2/3,0'), and a deviation-coded fixed effect of WordType. All the interactions of these main effects were also included. Random effects were modelled using a maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). This included random intercepts for subjects and items, by-subject random slopes for WordType and Rate and by-item random slopes for Group. Log-likelihood tests using the *anova()* function in R revealed that removing any fixed factors would lead to significantly worse model fit. Models were fitted using a maximum likelihood technique. A fixed effect was considered significant if the absolute value of the *t*-statistic was greater than or equal to 2.0 (Gelman & Hill, 2007). The raw data were right-skewed and consequently log-transformed. Shapiro-Wilk tests showed that the transformed data were normally distributed ($p > .05$).

The first model (Table 3) tested how V1 Duration Ratio changed in different conditions. This model revealed that native speakers tended to have a greater average V1 Duration Ratio than the learners ($\beta = .327$, SE = .174, $t = 1.88$), although their difference

was only marginally significant. On the other hand, the two learner groups were not significantly different from each other (t = .48). Fast speech had a smaller V1 Duration Ratio than normal speech (β = -0.126, SE = .038, t = 3.36), which in turn had a smaller ratio than slow speech (β = -0.086, SE = .038, t = 2.26). Compared to the learners, the Native group had significantly greater V1 Duration Ratios in normal than in fast speech (β = -0.286, SE = .089, t = -3.22) and in slow speech compared to normal speech (β = .196, SE = .090, t = 2.18).

The second model had Word Duration Ratio as the dependent variable but was otherwise identical to the first one. Unlike in V1 Duration Ratio, normal speech rate had the greatest Word Duration Ratio in general, compared to fast (β = -0.031, SE = .010, t = -2.98) and slow (β = -0.024, SE = .011, t = -2.22) speech. The main effect of Group was non-significant. Compared to the learners, native speakers had significantly greater Word Duration Ratios in slow than in normal speech (β = .061, SE = .025, t = 2.41). Taken together, the slow speech of native speakers saw both greater V1 Duration Ratio and Word Duration ratio (see Figure 2 and Figure 3), suggesting that they were enhancing the quantity contrasts in slower speech. The same pattern was not observed in any of the learner groups. Finally, the lack of difference between Advanced and Beginner groups in both duration

ratios suggests learners in the Advanced group may not be better than their peers in the Beginner group.

Insert Table 3 about here

To find out if/how the learners are distinguishing between long and short vowels with respect to neighbouring segments of the same word, we considered a final measurement, namely Vowel-to-Word Duration Ratio (Hirata, 2004). As shown in Figure 4, for all speaker groups, mean Vowel-to-Word Duration Ratio was greater for CVVCV words than for CVCV, although for the learners there was greater overlap between the two quantities. Like in the global measurements above, Figure 4 showed clear evidence of contrast enhancement in native speakers' slower speech, with the ratio in CVVCV becoming greater as one spoke slower. With CVCV words, the three groups did not appear to be different. The two reference values 0.29 and 0.49 came from Hirata (2004), representing respectively accented short and long vowels. All in all, the above measurements point to the fact that although the learners could clearly distinguish short vs. long vowels in their production, it is only the native speakers who enhanced the contrasts in slower speech.

Insert Figure 4 about here

Singleton vs. geminate consonants

The production of singleton vs. geminate consonants was analyzed in terms of the duration ratio of C2 (C:CC) as well as that of surrounding vowels (i.e. V1 and V2). In Figure 5, the 1:1 Closure Duration Ratio threshold means that singleton and geminate consonants are equal in C2 duration. The 1:2.8 reference was taken from Han (1992). Our native speakers were much closer to the 1:2.8 reference (mean = 2.37, SD = .55) than were the learners (Advanced mean = 1.69, SD = .62; Beginner group mean = 1.76, SD = .63). The contrast-enhancing effect of slow speech was obvious in the native speakers but unclear for the learner groups, especially in non-words. Interestingly, Closure Duration Ratio turned out to be much larger in non-words than in real words for both Advanced (respectively 1:1.98 and 1:1.50, S.D. = .62) and Beginner groups (respectively 1:2.05 and 1:1.57, S.D. = .63) but not for the native speakers (respectively 1:2.41 and 1:2.35, S.D. = .55). All individual speakers exceeded the 1:1 threshold in all speech rate and word type conditions (range .99~3.58), except A2 whose Closure Duration Ratio was .99 for non-words at normal rate.

Insert Figure 5 about here

Linear regression analysis showed a significant positive correlation between mean syllable duration and Closure Duration Ratio (r = .537, p < .001) for native speakers, confirming that the contrast between singleton vs. geminate consonants became greater in slower speech. For the Advanced group, mean syllable duration was weakly correlated with Closure Duration Ratio (r = .088, p = .042) whereas for the Beginner group the same correlation was non-significant. Like with vowel quantity contrasts, although the learners were able to distinguish short vs. long consonants, they used a strategy different from that of the native speakers across different speech rates.

The data were highly right-skewed (max. value = 6.61, skewness = 1.265, SE = .073), which echoed Vance's (1987, p. 71) remark about Closure Duration Ratio: 'as long as the average duration of a geminate stop is significantly longer than twice that of a single stop we can maintain the claim that moras are isochronous'. While the upper limit of this singleton-to-geminate ratio appears to be quite flexible, it also means that the normality assumption of *lmer()* was violated. Shapiro-Wilk tests showed that these data were non-normally distributed even after transformation (log, square and cube root, reciprocal). As a

result, here we performed a generalized linear mixed-effects model with Gamma distribution (log link) instead.

The final model in Table 4 was built by removing non-significant fixed effects from the most complex model. The significance of each fixed effect was determined using *anova()*, by comparing a model with the fixed effect in question and a model without. The final model contained the fixed effects of Group, Rate, WordType, and the interaction between Rate and WordType. Random effects included intercepts for subjects and items, by-subject random slopes for WordType and Rate and by-item random slopes for Group. The coding of variables was the same as in the models in Table 3.

Results showed that the learner groups had a significantly smaller Closure Duration Ratio than the Native group (β = .345, SE = .118, t = 2.93) whereas the two learner groups were not significantly different from each other. On the whole, Closure Duration Ratio was greater in slow speech compared to normal speech (β = .100, SE = .031, t = 3.18). For fast speech, Closure Duration Ratio was greater in non-words than in real words (β = -.101, SE = .045, t = -2.24). It is interesting to note that the interaction between Group and Rate was non-significant, which echoed Figure 5 where learners seemed to be also consistently enhancing the short vs. long contrast in slow speech at least in real words.

Place of articulation appeared to affect C2 duration ratio too. Table 5 showed that in the present study, for all speaker groups, /t/ had a greater C2 duration than /k/, like the native speaker group in Han (1992). It is worth pointing out that this included our learner groups as well, unlike the American English speakers in Han's data, who did not manifest such a pattern.

Insert Table 5 about here

Next, the effect of consonant quantity on the duration of the preceding V1 was examined, following Han (1994) and Idemaru and Guion (2008). Han (1994) reported that V1 was 11% longer (see the 1:1 threshold and the 1.11 reference in Figure 6) before and V2 was 9% shorter after a geminate. Like in previous studies, our native speakers lengthened V1 before a geminate, whereas the learner groups did not do so consistently (see Discussion on the effect of speech rate). For example, in non-words spoken at slow speed, V1 was even shorter before a geminate for both learner groups. Examination of individual data revealed that many participants (three beginners and six advanced learners) deviated from the 1:>1 norm in terms of grand mean. Of these speakers, one (B4) failed to lengthen V1 before a geminate across all speech rate × word type conditions, while others managed

to do so at least in some contexts. Only four learners (Advanced: A2, A5, B9; Beginner: B7) consistently lengthened V1 across all speech rate × word type conditions. Finally, slow speech did not seem to enhance quantity contrast in terms of V1 Duration Ratio (C:CC) even for native speakers.

Insert Figure 6 about here

We fitted another generalized linear mixed-effects model with Gamma distribution (log link) to V1 Duration Ratio. All fixed effects were the same as the model for Closure Duration Ratio above, whereas for random effects we only included intercepts for subjects and items; more complex random effects structures led to non-convergence of the model. Table 4 showed that for V1 Duration Ratio the difference between the Native group and the learners was marginally significant (β = .127, SE = .069, t = 1.84) whereas the two learner groups were not significantly different from each other. Unlike any other duration ratios discussed above, in general V1 Duration Ratio was smaller in slow speech than in normal speech (β = -.107, SE = .013, t = -7.98). In slow speech, V1 Duration Ratio was significantly greater in real words than in non-words (β = .080, SE = .027, t = 2.99). All in all, V1 Duration Ratio manifests an opposite pattern to Closure Duration Ratio.

For V2 Duration Ratio, Table 6 showed that our native speakers always shortened V2 after a geminate, as did the Beginner group (all 1:<1); whereas the Advanced learners' production was a mixed picture. We fitted a generalized linear mixed-effects model with the same fixed and random effects structure as the Closure Duration Ratio model (Table 4), and found that this time the Native group did not differ significantly from the learner groups, who in turn differed from each other marginally significantly (β = .071, SE = .041, t = 1.74). The difference in V2 duration between a singleton and a geminate C2 was smaller in slow speech than in normal speech (β = -.050, SE = .022, t = -2.28).

Insert Table 6 about here

Insert Table 7 about here

Discussion

Overall production performance

The present study has yielded a range of evidence to show that Cantonese-speaking learners of Japanese were able to distinguish between vocalic and consonantal quantities,

albeit using a different strategy from that of their native speaker counterparts. **H1a** ('Beginner group will show evidence of some ability to distinguish Japanese short vs. long vowels') and **H2a** ('Beginner group will show evidence of some ability to distinguish Japanese short vs. long consonants') were therefore supported. However, they differed from the native speakers with smaller ratio values and by failing to enhance quantity contrasts in slower speech. Moreover, it was also observed that while the Beginner group made a clear distinction between short vs. long phonemes, the Advanced learners were not remarkably better than they were in terms of demonstrating native-like duration ratios, thus rejecting **H1b** ('Advanced learners will distinguish Japanese short vs. long vowels more similarly to native speakers') and **H2b** ('Advanced learners will distinguish Japanese short vs. long consonants more similarly to native speakers'). Taken together, it appears that the learners have acquired the quantity distinctions but have not fully developed the acoustic targets, specifically they have not mastered the control of duration in different speech rate conditions.

For short vs. long vowels, the learners showed a smaller V1 Duration Ratio (1:2.01 for Advanced and 1:1.93 for Beginner) than the native speakers (1:2.24), but the two learner groups did not differ from each other significantly. We also replicated the contrast-enhancing effect of slow speech on vowel duration and word duration ratios in the native

speakers (Hirata, 2004), which was absent in both learner groups. With regards to singleton vs. geminate consonants, the learners showed a smaller Closure Duration Ratio (1:1.69 for Advanced and 1:1.76 for Beginner) than the native speakers (1:2.37). Compared to 1:2 observed in the fluent American learners in Han (1992), our Cantonese Advanced learners did not seem to be any better. Again, there was a contrast-enhancing effect of slow speech on Closure Duration Ratio in the native speakers, but not in the learner groups in general.

Although the learners' acquisition of quantity contrasts was not perfect in the sense that their slow production deviated from the native speakers significantly, their ability to tell apart short vs. long sounds was clear and undeniable. If it is indeed the case that the production of segments is capped by perception, as posited by SLM, our results would predict that Cantonese learners' perception of Japanese quantity contrasts would be quite accurate. However, a pilot study (Liu & Hirata, 2016) using duration- and f_o-manipulated stimuli showed that Cantonese learners perceived Japanese vowel lengths only gradiently (no accuracy data were reported). A follow-up experiment is under way to test this prediction further.

The good performance of the Beginner group was unexpected. For both consonantal and vocalic quantities they were evidently able to produce short vs. long sounds differently. Since they were only three months into their degree programme, it seems reasonable to

attribute their performance to L1 (and perhaps L2) transfer. In Cantonese there is only one monophthong pair (i.e. /ɐ/ vs. /a:/ which can appear as part of numerous rhymes) that uncontroversially contrast in length as well as the 'cat tail' geminates. Our learner groups' ability to distinguish between the quantities is thus likely transferred from these partial uses of duration in their L1, much like the American English participants in McAllister et al (2002). Our data thus suggest that, in this case, facilitative L1 transfer is based on phonetic features (e.g. McAllister et al., 2002) rather than on actual phonemes. That is, the use of duration as a cue to only a subset of vowels in Cantonese seems already enough to help learners distinguish quantity conditions in different L2 vowels. Our results also point to the fact that learners can benefit from their L1 even if the phonetic dimension in question is not used phonemically. That is, Cantonese has no underlying geminates but the derived geminates may have helped our learners acquire Japanese geminates.

The next logical question is whether the good performance in both types of quantity contrasts may have come only from the phonological use of duration in vowels. As reviewed in the Introduction, although the performance of Cantonese participants in Pajak and Levy (2014) could logically be attributed to their L1 experience in both consonantal and vocalic quantities, the findings in Tsukada et al (2014) suggest that the use of duration in L1 consonantal quantities does not get transferred to L2 vocalic quantities, no matter

how heavily duration is used as a cue in L1. It then follows that our learners' ability to contrast short vs. long Japanese vowels should be attributed to the partial use of duration in a small set of short vs. long vowels in Cantonese, whereas their ability to contrast short vs. long Japanese consonants could be due to the presence of derived geminates in their L1.

It is unclear why the Advanced and Beginner groups did not differ from each other, even after we reclassified the two more experienced learners as 'Advanced'. Since our Beginner group had only received three months of formal instruction in the classroom, perceptual learning should still be ongoing (cf. Best & Tyler, 2007, who suggest that the cut-off should be 6-12 months). Meanwhile, the Advanced group's production did not become significantly more native-like than the Beginner group's even after two years of intensive language training and one year of immersion in Japan. In this sense, it is as if the challenge that Cantonese learners face as beginners persists through their proficiency curve and remains even after they have become much better speakers. It is possible that the Beginner group had stopped improving prior to university as a result of extensive exposure to Japanese in Hong Kong as naive listeners (e.g. film, manga, J-pop, TV drama). This is because for them to have chosen Japanese as their major, likely they had developed their interest in the language through extensive exposure prior to formal classroom training. In fact, in post-experiment interviews all participants indicated that in their spare time they

would do at least one of the following: watch Japanese *anime* / TV drama, read Japanese *manga*, and listen to Japanese songs. If exposure to Japanese as naive listeners does count, Best and Tyler's 6-12 month clock must have started ticking before our subjects started actively studying the language. A follow-up study comparing naive Cantonese speakers and learners can test this hypothesis. Alternatively, it is possible that after the contrast is acquired and there is no problem in communication, the need to fully develop the acoustic target is no longer the same as that to acquire the distinction in the first place. In that case, their not approaching native-like duration ratios is not to be seen as a lack of 'improvement'. This is in line with the view in Munro (2008) that L2 pronunciation should focus on ensuring communication with interlocutors (whether native or non-native), instead of native-likeness, which in itself is hard to define. Last but not least, perhaps simply having a larger sample size could solve this puzzle, although recruiting Cantonese-speaking learners of Japanese has been more challenging than, say, learners of English.

Effect of speech rate

Two observations are interesting with regards to speech rate. The first is that our learners failed to enhance contrasts in slower speech like their native counterparts did. In both vocalic and consonantal contrasts, our learners deviated more from native speakers in slow speech. Earlier work on articulation (Gay, 1981) showed that gestures are reorganized

under different speech rates and the change in segment duration in various speech rate conditions does not occur linearly across phonetic segments. Although our learners did show different duration ratios across speech rate conditions, these changes were not systematic like those in the native speakers. That our learners appeared to deviate more from native speakers when speaking slowly might possibly be due to their lack of practice in unnaturally slow speech. However, at least at the beginning of their study students are usually first exposed to teachers' canonical slow production; in that case why were our learners not better at slow speech instead? A likely explanation would be that compared to fast speech, speaking slowly requires higher precision in controlling relative segment duration; longer word duration means higher chances of being inaccurate.

Secondly, our data suggest that for production, quantity distinction is harder to master in slower speech, while the opposite is true for perception (Hirata, Whitehurst, & Cullings, 2007). This observation has implications for language teaching. Whereas in slow production, longer duration leaves more room for imprecision for learners to produce, perceiving slower speech means more time to process the clearer contrasts from native speakers. That learners' struggle with different speech rates for production vs. perception thus reminds us that it is not ideal to expose them to input of only one speed, e.g. slow speech.

Pre-geminate lengthening

With regards to pre-geminate lengthening (cf. Maddieson, 1985), Figure 6 suggests that the learners were only lengthening their V1 in some conditions, unlike their native peers who consistently did so across all speech rate and word type conditions (thus partially supporting **H3**: 'our learners will be able to lengthen V1 before a geminate'). In some cases, the Advanced learners were lengthening V1 less than the Beginner group as if their pronunciation had deteriorated. It appears that the learners performed V1 lengthening better at normal speech rate than slow speech rate, better in real words than non-words. Then in the most challenging condition, namely non-words in slow speech, the learners shortened V1 instead, somehow conforming to the typological tendency per Maddieson (1985).

The source of discrepancy between real words and non-words is unclear. One conceivable explanation might be the fact that in the real word condition speakers were presented with the *hiragana* syllabary as well as *kanji* characters, the latter of which are familiar to our learners. The fact that, in the non-word condition, they were presented only with the *katakana* syllabary which is absent in their L1 and to which the Beginner group would be less accustomed might contribute to the different durational patterns in their production. It would be interesting to verify if orthography does have an effect on the production of phonemic quantities.

Other implications

With regards to the feature hypothesis, that our learners were able to distinguish short vs. long vowels in Japanese despite the limited use of duration in Cantonese vowel quantity contrasts clearly shows that facilitative L1 transfer occurs at the feature level. Our results would agree with Pajak and Levy's proposal that 'perceptual reorganization leads not only to perceptual sensitivity to specific L1 phonetic categories, but also to sensitivity induced by these higher-order generalizations' (2014:156). In other words, a native speaker's knowledge with respect to L1 transfer is hierarchical, not flat-level and category-by-category. By implication, the ultimate state of L2 phonology would consist of both the inventory of specific phonetic categories as well as a refined sensitivity and precision of encoding for the relevant phonetic dimensions that determine category contrasts (ibid.). See also other works relevant to feature redeployment such as Archibald (2009), Goad and White (2006) and Lardiere (2009).

Another potential implication concerns phonology. Given Tsukada et al (2014), the learners' success in distinguishing singleton vs. geminate consonants, as evidenced by the duration ratios, should probably be attributed to the derived geminates in Cantonese. If this is indeed the case, the resyllabification which accompanies this transfer would have theoretical implications for the formal representation of syllable structure. Whereas how

exactly geminates should be represented is a theory-internal question, that coda+initial sequences and geminates have different structures should go without saying. Having said that, it is logically possible that our learners actually did not produce any geminates in the experiment, but coda+initial sequences like in their L1, in which case the conclusion here would become that featural transfer only occurs when the phonetic dimension in question is used in an underlying contrast. An articulatory study is thus needed to verify the consonant part of our findings.

Caveats and limitations

Finally, some possible limitations of this study should be noted. The first issue is the imperfect homogeneity of the learner groups. In this study we have two homogenous groups of learners in the sense that they came from the same department and followed the same syllabi, but their Japanese proficiency both prior to university and at the time of testing was not identical as desired. In particular, two participants in the Beginner group who posed as genuine beginners had to be reclassified as advanced learners in the analysis. The result of such reclassification was that two of 12 participants in the Advanced group did not have any experience living in Japan. Although all individual speakers fell within the 1:>1 range in all of the duration ratios discussed (except pre-geminate lengthening of V1), individual variability (range of values) was considerable. Relatedly, the second issue

concerns sample size. In a degree programme of which the annual intake was about 20 students, there were not many with comparable proficiency who were also available to participate. Admittedly, for an L2 phonetic study 10 speakers in each group was not a large number and could have led some of the effects tested to be statistically non-significant. The third concerns the possible influence of task in the nature of the results. While our participants were able to distinguish between short and long vowels and consonants, they could have benefitted from being in a controlled reading-aloud context where they were able to attend to and control phonetic implementation, especially when the stimuli were presented in quantity-transparent orthography (i.e. *kana*). A less controlled task that is administered without the use of orthography would be ideal for verifying our findings. Last but not least, although our participants' L2s are not deemed to confound our results, it would have been ideal to verify our findings by comparing speakers with and without these languages. That said, in both China and Hong Kong it is difficult to find literate Cantonese speakers learning Japanese who have no knowledge of either English or Mandarin.

Conclusion

In this paper we have presented a series of durational data to compare the production of Japanese phonemic quantity contrasts by native speakers, advanced learners and beginner learners from Hong Kong. We set out to ask whether L1 transfer is possible

when L1 experience is restricted to a small set of sounds (i.e. vocalic quantity contrasts in Cantonese) and when L1 experience is non-phonemic (i.e. derived geminates in Cantonese), and our results suggest that both are true. The key findings are as follows: (i) both learner groups showed clear ability to distinguish short vs. long sounds, and the Advanced group did not seem to be any more native-like; (ii) only the native speakers enhanced the quantity contrasts in slower speech; and (iii) the learner groups showed evidence of pre-geminate lengthening of V1 only in some cases. Future research should verify these findings using a less controlled task that is administered without the use of orthography. It is hoped that our results will help us gain a better understanding of the nature of L1 transfer in L2 acquisition.

References

Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Second Language Research*, *252*(2), 231–233.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models using {lme4}. *Journal Of Statistical Software*, *67*(1), 1–48.

Beckman, M. E. (1982a). Effects of accent on vowel duration in Japanese. *Journal of the*

Acoustical Society of America, *71*, S23.

Beckman, M. E. (1982b). Segment duration and the "mora" in Japanese. *Phonetica*, *39*, 113–135.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementaries. In M. J. Munro & O.-S. Bohn (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Boersma, P. P. G., & van Heuven, V. J. J. P. (2001). Speak and unSpeak with PRAAT. *Glot International*, *5*(9/10), 341–347.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Oxford: Wiley-Blackwell.

Cheng, C., & Xu, Y. (2013). Articulatory limit and extreme segmental reduction in Taiwan Mandarin. *Journal of the Acoustical Society of America*, *134*(6), 4481–4495.

Chiu, F., Fromont, L. A., Lee, A., & Xu, Y. (2015). Long-distance anticipatory vowel-to-

vowel assimilatory effects in French and Japanese. In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, Scotland.

Escudero Neyra, P. R. (2009). The linguistic perception of similar L2 sounds. In P. P. G. Boersma & S. R. Hamann (Eds.), *Phonology in Perception* (pp. 152–190). Berlin: Mouton de Gruyter.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., Takagi, N., & Mann, V. A. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, *38*(1), 25–55.

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, *38*, 148–158.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, *22*(3), 243–268.

Han, M. S. (1992). The timing control of geminate and single stop consonant in Japanese: A challenge for nonnative speakers. *Phonetica*, *49*, 102–127.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the*

Acoustical Society of America, *96*(1), 73–82.

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, *32*(4), 565–589.

Hirata, Y. (2015). L2 phonetics and phonology. In H. Kubozono (Ed.), *Handbook of Japanese Phonetics and Phonology* (pp. 719–762). Berlin: Mouton de Gruyter.

Hirata, Y., & Lambacher, S. G. (2004). Role of word-external contexts in native speakers' identification of vowel length in Japanese. *Phonetica*, *61*(4), 177–200.

Hirata, Y., & Tsukada, K. (2009). Effects of speaking rate and vowel length on formant frequency displacement in Japanese. *Phonetica*, *66*(3), 129–149.

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates. *Journal of the Acoustical Society of America*, *121*(6), 3837–45.

Idemaru, K., & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association*, *38*(2), 167–186.

Kang, Y., Yoon, T.-J., & Han, S. (2015). Frequency effects on the vowel length contrast merger in Seoul Korean. *Laboratory Phonology*, *6*(3–4), 469–503.

Kao, D. L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.

Kingston, J., Kawahara, S., Chambless, D., Mash, D., & Brenner-Alsop, E. (2009).

Contextual effects on the perception of duration. *Journal of Phonetics*, *37*(3), 297–320.

Kubozono, H. (2017). Introduction to the phonetics and phonology of geminate consonants. In H. Kubozono (Ed.), *The phonetics and phonology of geminate consonants* (pp. 1–10). Oxford: Oxford University Press.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *97*(22), 11850–11857.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 979–1000.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect." In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, MD: York Press.

Kurihara, M. (2004). 中国語北方方言話者の日本語長音の知覚特徴 [Perception of Japanese long vowels by speakers of Mandarin]. *Tohoku University Journal of Linguistic Science [言語科学論集]*, *8*, 1–12.

Kurihara, M. (2005). 中国語北方方言話者の日本語長音と短音の産出について [On

the production of Japanese long vs. short vowels by speakers of Mandarin]. *Tohoku University Journal of Linguistic Science [言語科学論集]*, *9*, 107–118.

Labrune, L. (2012). *The phonology of Japanese*. New York, NY: Oxford University Press.

Lai, Y. W. (1999). *Prosody and prosodic transfer in foreign language acquisition: Cantonese and Japanese*. *PhD Thesis*. University of Hong Kong, Hong Kong.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, *252*(2), 173–227.

Lee, A., & Mok, P. K. P. (2016). Durational correlates of Japanese phonemic quantity contrasts by Cantonese-speaking L2 learners. In *Proceedings of the 8th International Conference on Speech Prosody (SP2016)* (pp. 597–601). Boston, MA.

Lee, H. B. (1999). Illustrations of the IPA: Korean. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Liu, C. Y., & Hirata, Y. (2016). Transfer of L1 tone knowledge to L2 vowel quantity contrasts as a secondary cue: The case of Cantonese learners of Japanese. *Journal of the Acoustical Society of America*, *140*(4), 3340.

Maddieson, I. (1985). Phonetic cues to syllabification. In V. A. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 203–221). Orlando, FL: Academic Press.

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258.

Minagawa, Y., & Kiritani, S. (1996). Discrimination of the single and geminate stop contrast in Japanese by five different language groups. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, *30*, 23–28.

Minagawa, Y., Maekawa, K., & Kiritani, S. (2002). 日本語学習者の長／短母音の同定におけるピッチ型と音節位置の効果 [Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers]. *Journal of the Phonetic Society of Japan*, *6*(2), 88–97.

Motohashi-Saigo, M., & Hardison, D. M. (2009). Acquisition of L2 Japanese geminates: Training with waveform displays. *Language Learning & Technology*, *13*(2), 29–47.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition*. Amsterdam: John Benjamins.

Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*(4), 451–468.

Ofuka, E. (2003). 促音/tt/の知覚：アクセント型と促音・非促音語の音響的特徴による違い [Perception of a Japanese geminate stop /tt/: The effect of pitch type and acoustic characteristics of preceeding/following vowels]. *Journal of the Phonetic Society of Japan*, *7*(1), 70–76.

Pajak, B., & Levy, R. (2014). The role of abstraction in non-native speech perception. *Journal of Phonetics*, *46*(1), 147–160.

R Core Team. (2016). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Roach, P. J. (2004). Illustrations of the IPA: British English—Received Pronunciation. *Journal of the International Phonetic Association*, *34*(2), 239–245.

Sagayama, J. (2010). *A comparison of the duration of special morae in the speech of native speakers and Cantonese learners of Japanese*. *MPhil Thesis*. University of Hong Kong, Hong Kong.

So, L. K. H., & Wang, J. (1996). Acoustic distinction between Cantonese long and short vowels. In *Proceedings of the 6th Australian International Conference on Speech Science & Technology (SST 1996)* (pp. 379–384). Adelaide, Australia.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hansen Edwards & M. L. Zampini (Eds.),

Phonology and second language acquisition (pp. 153–191). Amsterdam: John Benjamins.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. G. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America*, *123*(1), 397–413.

Takiguchi, I., Takeyasu, H., & Giriko, M. (2010). Effects of a dynamic F0 on the perceived vowel duration in Japanese. In *Proceedings of the 5th International Conference on Speech Prosody (SP2010)*. Chicago, IL.

Toda, T. (2003). 外国人学習者の日本語特殊拍の習得 [Acquisition of special morae in Japanese as a second language]. *Journal of the Phonetic Society of Japan*, *7*(2), 70–83.

Tsukada, K. (2011). The perception of Arabic and Japanese short and long vowels by native speakers of Arabic, Japanese, and Persian. *Journal of the Acoustical Society of America*, *129*(2), 989–998.

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

van Leussen, J.-W., & Escudero Neyra, P. R. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*(1000).

Vance, T. J. (1987). *An introduction to Japanese phonology*. New York, NY: State University of New York Press.

Whalen, D. H. (1989). Vowel and consonant judgments are not independent when cued by the same information. *Perception & Psychophysics*, *46*(3), 284–292.

Yip, M. J. W. (1993). Cantonese loanword phonology and Optimality Theory. *Journal of East Asian Linguistics*, *2*(3), 261–291.

Yoshida, S. (1990). A Government-based analysis of the "mora" in Japanese. *Phonology*, *7*(2), 331–351.

Zee, Y. Y. E. (1991). Illustrations of the IPA: Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*, *21*(1), 46–48.