Research Note

second language research

Second Language Research
1–16

DOI: 10.1177/02676583221128530
journals.sagepub.com/home/slr
SAGE

False geminates as an effective transitional strategy for Cantonese learners of Japanese

Albert Lee
The Education University of Hong Kong, Hong Kong

Xiaolin Li
Peggy Mok
The Chinese University of Hong Kong, Hong Kong

Abstract

This article revisits Lee and Mok (2018) and examines how the Cantonese learners in the study produced second language (L2) Japanese short vs. long consonants which are absent in their first language (L1). Specifically, our goal is to find out whether these learners were substituting real geminates (i.e. long consonants) with the improvised strategy of an unreleased stop coda + homorganic initial consonant sequence (like in the phrase *cat tail*), which would not have been detectable with the durational ratios in the original study. We analysed the mean intensity of the words *sassa*, *sesse*, and *sosso*, to investigate whether the learners were producing a [t̚]+/s/ cluster, presumably drawn from their L1. The results showed that the beginner group were indeed using this strategy, whereas the advanced learners were largely producing a genuinely geminated /s/ akin to the native speakers. The use of this transitional strategy was also speech-rate dependent, with more cases of /t/-insertion in slower speech for both learner groups. We conclude that (1) although having L1 false geminates does not enable beginners to readily acquire genuine L2 geminates, the latter can be learned after enough exposure, and (2) during this transition, Cantonese learners can draw on L1 phonotactic knowledge to improvise creative and effective strategies to attain L2 durational targets.

Keywords

Cantonese, consonant quantity, Japanese

Corresponding author:
Peggy Mok, Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, Hong Kong.
Email: peggymok@cuhk.edu.hk

I Introduction

One line of research in second language (L2) phonological acquisition asks whether learners' difficulty in learning L2-specific phonetic contrasts is affected by L1–L2 differences at the level of discrete sound categories or continuous phonetic dimensions (or 'features') (see review in Jiang, 2018: Section 2.5). For example, does having a /t/–/d/ contrast in L1 help learners acquire a new /p/–/b/ contrast in L2, as the primary acoustic cue to both contrasts is voice onset time (VOT, Flege and Port, 1981)? A further question is whether the ability to exploit a phonetic dimension is transferrable from one syllable constituent to another, e.g. applying L1 VOT in initial consonants to L2 codas (e.g. in an ambisyllabic position). The answers to these questions have strong implications both for theories that account for L2 phonological acquisition and for the teaching and learning of new sounds in an L2.

1 The category vs. feature debate

Many of the currently prevailing models of L2 phonological acquisition, including the Native Language Magnet Model (e.g. Kuhl, 1992, 1994), the Perceptual Assimilation Model (e.g. Best, 1995), and the Speech Learning Model (e.g. Flege, 1995), mainly consider phonological processing and development at the segmental level (for a recent summary of these models, see Chang, 2019). These models largely draw their support from evidence of phoneme categories in L1 shaping the acquisition of L2 categories. However, as categories comprise phonetic features, logically speaking, evidence of categorical transfer does not rule out the possibility of L1 influence at the featural level.

Other models consider features as the unit that affects perception. By extension, if a speaker's L1 sound system has a particular feature (e.g. quantity), this feature will facilitate the perception of that non-native contrast, irrespective of whether the particular segment is present in the L1. Note that the term 'feature' is used in different senses in this line of research: in some studies (e.g. Hancin-Bhatt, 1994) it refers to traditional distinctive features à la Chomsky and Halle (1968) and in others (e.g. McAllister et al., 2002) it is taken to refer to continuous phonetic dimensions; in this study we use 'feature' in the latter sense.

One relevant study is McAllister et al. (2002), which evaluated L2 learners' ability to use duration as a feature to cue phonemic quantity. This feature is needed to produce and perceive vowel distinctions in Swedish. Participants were speakers of English and Spanish who had lived in Stockholm for more than 10 years, as well as native speakers of Estonian (which has a three-way vowel quantity contrast; Asu et al., 2009) and Swedish as controls. The results showed that while Estonian controls performed much like native Swedish speakers, the learner groups did not perform as well. In addition, the native English speakers outperformed their Spanish speaker counterparts, echoing the fact that duration is partially used to cue lexical stress contrasts in English (alongside vowel quality and intensity; Fry, 1955). McAllister et al.'s findings lend support to the featural transfer hypothesis, and demonstrate that the degree to which a phonetic feature (in this case duration) is used in L1 can influence how well it can be exploited for L2 (see also other relevant studies such as Lee et al., 2006; Tsukada et al., 2014, 2018).

2 Vowel vs. consonant quantity: Asymmetrical effects on L2

A further question is whether the aforementioned facilitative effect of an L1 feature on L2 acquisition is constituent specific: does having cues to a vowel quantity contrast in L1 help the learner acquire a consonant quantity contrast in L2? Altmann et al. (2012) reported asymmetrical effects of L1 vowel vs. consonant quantity contrasts on non-native quantity perception. They found that Italian listeners (who have only consonant quantity contrasts in L1) perceived non-native German vowel quantity contrasts as well as they did L1 consonant quantity contrasts; but German learners (who have only vowel quantity contrasts in L1) of Italian did not perceive Italian consonant quantity contrasts equally well. Their findings appear to suggest that the influence of L1 quantity contrasts may be constituent specific (i.e. L1 vowel quantity contrasts do not necessary help perceive new L2 consonant quantity contrasts), and asymmetric between vowels and consonants.[1]

3 Lee and Mok (2018)

The studies reviewed above have revealed that the relationship between L1 and L2 quantity contrasts is much more complex than the original feature hypothesis (McAllister et al., 2002). The next question was whether L1 quantity contrasts restricted to only a small subset of sound pairs would (partially) facilitate the acquisition of new L2 quantity contrasts; to this end Lee and Mok (2018) tested Cantonese learners of Japanese.

In Japanese, both consonants (e.g. *kita* 'came' vs. *kitta* 'cut') and vowels (e.g. *kita* 'came' vs. *kiita* 'heard') contrast in quantity. Cues to the short vs. long vowel distinction include vowel and word duration (Hirata, 2004), formant frequency (Hirata and Tsukada, 2009) as well as f_o contour (Takiguchi et al., 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 intensity and f_o range (Ofuka, 2003). For a recent review of the phonetics of Japanese geminates, see Kawahara, 2015. In Cantonese, there are short vs. long vowels but only to a very limited extent. Only the low vowels (/aː/ vs. /ɐ/) contrast in quantity phonemically (e.g. /kaːi/ 街 'street' vs. /kɐi/ 雞 'chicken'), but they also differ in vowel quality (Kao, 1971).[2] For consonants, there are no underlying geminates in Cantonese, but there are the 'cat tail' type false geminates – derived at syllable boundaries, e.g. /pʰaː.kʰɵy/ 怕佢 'afraid of him' vs. /pʰaːk.kʰɵy/ 拍佢 'tap him (e.g. on the shoulder)' – because Cantonese allows an unreleased stop coda in its syllable structure (for English, see Oh and Redford, 2012).

Lee and Mok (2018) explored the production of L2 Japanese quantity contrasts by 20 Cantonese learners vs. native Japanese speakers. The learners consisted of beginners (first-year undergraduate students three months into their BA Japanese Studies programme) and advanced learners in the final year of the same programme who had studied in Japan for one year. Based on the analysis of a series of durational ratios, they concluded that both groups of learners were able to distinguish between short vs. long vowels and consonants. These findings were taken to suggest that (1) the use of duration in a restricted subset of vowels suffices to facilitate the use of duration in L2 vowel quantity contrasts even for beginners, and (2) false geminates may allow Cantonese learners to correctly produce Japanese consonant quantity contrasts. Alternatively, (2) could also be interpreted to mean that (3) the restricted use of duration in L1 vowel quantity contrasts can benefit the acquisition of L2 consonant quantity contrasts.

Whether (2) and/or (3) are true is important as they have theoretical implications for the nature of transfer in L2 phonological acquisition. However, upon closer examination of the recordings in Lee and Mok (2018), the authors detected non-native-like pronunciation by the beginners that was suggestive of a consonant cluster [t̚]+/s/ as an alternative strategy for long /s/, which could not have been identified by durational ratios. Specifically, for many utterances involving a geminated /s/, we detected silence before frication but without noticeable affrication in between. That is to say, it is possible that the beginners in Lee and Mok (2018) were not producing real geminates; rather, they were producing a homorganic unreleased stop coda + initial consonant sequence to achieve the target long constriction: a phonotactic strategy imported from their L1. If this was the case, both (2) and (3) need to be reconsidered as these beginners were not producing the consonant quantity contrasts in the same way as native Japanese speakers do.

The /s/ subset of the stimuli provides a perfect window to test this. To determine whether speakers are producing real or false geminates, stops would not be ideal as both would equally yield lengthened stop constriction. On the other hand, because Cantonese does not allow fricatives in the coda position, geminated /s/ would become a consonant cluster [t̚]+/s/ (the former being a homorganic unreleased stop), which can provide a direct diagnosis of the problem at hand.

II Research questions

Here we seek to examine whether Cantonese learners' production of Japanese geminated /s/ is genuinely lengthened or is a [t̚]+/s/ cluster, i.e. only using the allowable Cantonese combination to fulfil the required duration. While a CVsV (singleton) word and a CVs:V (genuine geminate) word would only differ in terms of the duration, but not intensity, of the second consonant (C2), a false geminate CVt.sV would have lower corresponding C2 intensity than CVsV because a [t]–/s/ cluster contains a period of stop closure with no acoustic energy. Geminated stops would not allow such comparison as complete closure means zero intensity in both quantity conditions (i.e. CVt.tV for false geminate stop vs. CVt:V for genuine geminate stop).

We hypothesize that for Cantonese learners of Japanese, mean C2 intensity would be greater for singleton /s/ than for geminate /s/ (hypothesis 1). This is based on the assumption that they were unable to produce real geminates, but produced a [t̚]–/s/ cluster as an alternative strategy. For native Japanese speakers, who are capable of producing genuine geminates, we expect that singleton and geminated /s/ do not differ significantly in C2 mean intensity.

Moreover, if the learners did produce [t̚]–/s/ clusters instead of /s.s/, a further question would be how much they did so at different speaking rates. Lee and Mok (2018) found that the learners deviated from the native speakers more in slower speech, in which the actual durational differences between native and learners' production were larger given longer average syllable duration. In addition, decreased fricative duration can affect listeners' ability to correctly identify fricatives (Jongman, 1989), hence there is less room for /t/-insertion in faster speech for [t̚]–/s/ clusters. Consequently, we further hypothesize that there would be more cases of /t/-insertion in slower speech (hypothesis 2).

III Method

1 Speakers and materials

To test the aforementioned hypotheses, we reanalysed a subset of the production data in Lee and Mok (2018). This subset contained three non-word roots, namely *sasa*, *sese*, and *soso* (3 roots × 3 quantity conditions [CVCV, CVVCV, CVCCV] × 25 speakers × 3 speaking rates × 3 times = 2,025 utterances), of which the mean intensity of C2 was the dependent variable. There were five native speakers of Japanese as controls (2F + 3M, mean age = 31.0, *SD* = 10.6), 12 advanced learners (10F + 2M, mean age = 20.7, *SD* = 1.3) in their final year of the BA Japanese Studies programme at a university in Hong Kong and eight beginners (5F + 3M, mean age = 18.3, *SD* = .46) who were in their first year of the same programme. Both learner groups were native speakers of Hong Kong Cantonese, speaking English and Mandarin[3] as L2. The Beginner group reportedly had no knowledge of Japanese or at best just some knowledge of *hiragana* when they entered university. For more details of the speakers, please refer to Lee and Mok, 2018.

2 Procedures

Recording took place in a quiet room on campus, using a Zoom H2n voice recorder, placed on a desk approximately 30 cm in front of the participant. Stimuli were presented on a computer screen using a Javascript-based sentence randomizer. The speakers were briefed about the experimental task and granted their written consent before recording commenced. They were to say the target words in the carrier sentence *Kore-wa* XX *desu* ‘This is XX’. Utterances were collected over randomized blocks at three speaking rates (in order: normal ⇒ slow ⇒ fast), by asking the speakers to speak obviously more slowly in the slow production and obviously faster in the fast production, both relative to the normal speech rate. They were instructed to use the high–low pitch accent pattern. Within each block, each target word appeared three times. In this subset, there was no outlier and all data were retained for analysis.

Speech data were manually labelled by the segment (consonants and vowels) using FormantPro (described in Cheng and Xu, 2013; Chiu et al., 2015), following the principles in Turk et al. (2006). For all the disyllabic target words, four segments (henceforth $C_1V_1C_2V_2$) were labelled. Subsequently, for each labelled interval FormantPro extracted the duration and mean intensity values.

IV Results

1 Intensity trajectories

We set out to examine whether Cantonese learners’ production of geminated /s/ is genuinely lengthened or is a [t̚]–/s/ cluster. To this end, we compared the intensity of C2

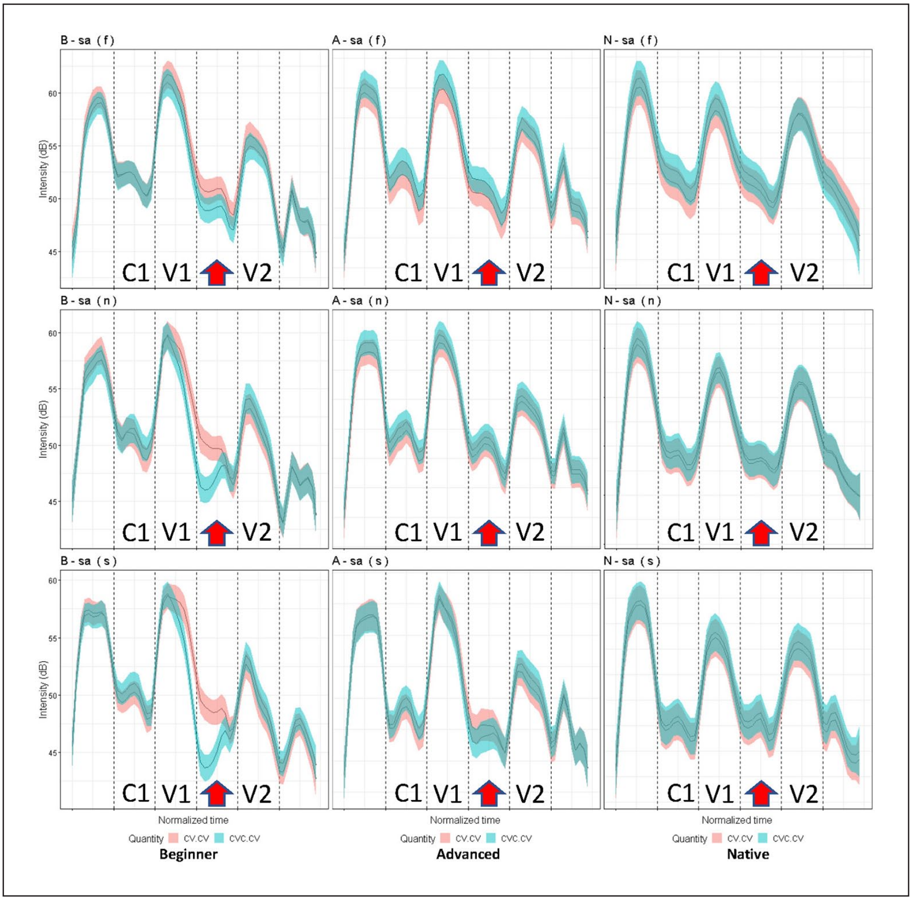

Figure 1. SS ANOVA plots showing mean intensity profiles of *sasa* vs. *sassa* produced by beginners (N = 8), advanced learners (N = 12), and native speakers (N = 5).
Notes. X-axis shows normalized time (10 measurements / segment). Vertical dashed lines indicate segment boundaries. Calculation of 95% Bayesian confidence intervals (i.e. thickness of ribbon widths) were calculated following Davidson (2006). Where two ribbons do not overlap, they may be considered significantly different at that particular point in time. From the top to the bottom rows are fast, normal, and slow speaking rates. The red arrows indicate the segment of interest (C2) (colour online).

produced by the speaker groups under different quantity conditions. As an example, Figure 1 shows mean intensity trajectories of *sasa* (CVCV) vs. *sassa* (CVCCV) produced by all speakers at different speaking rates. The right panels reveal that intensity at C2 did not differ between any quantity conditions for the native speakers (the Bayesian 95% confidence intervals overlap); the same was true for the advanced learners as indicated by the middle panels. For the beginners, however, the pink and lime contours are separated in the Normal and Slow panels, suggesting that intensity at C2 for *sassa* was significantly lower than that for *sasa*. Note that for all other segments in these target sentences, intensity was not significantly different between *sasa* and *sassa* for any speaker group.

Table 1. Model summary (beginners): Intensity ~ Quantity + Rate + (Quantity + Rate | Speaker).

	Fixed			Random
				Speaker
Parameters	β	SE	t	SD
(Intercept)	3.849	0.047	82.246***	.032
CVVCV	.032	.022	1.445	.024
CVCV	.052	.023	2.264*	.025
Fast	.033	.016	2.091*	.018
Slow	−.054	.013	−4.316***	.013
CVVCV × Fast	.002	.009	.221	
CVCV × Fast	−.013	.009	−1.397	
CVVCV × Slow	.032	.009	3.568***	
CVCV × Slow	.028	.009	3.063**	

Notes. Pr(>|z|) < .05 = '*'. < .01 = '**'. < .001 = '***'.

The full set of C2 mean intensity data (N = 2,025) were analysed using generalized linear mixed models with R (ver. 3.6.3) package *lme4* (Bates et al., 2015 ver. 1.1-23). The three-level categorical factors (quantity and speaking rate) both followed a treatment coding structure with a baseline level: for quantity, baseline = CVCCV; for speaking rate, baseline = normal.

Separate models were fitted for each speaker group. The best models were chosen following a model-comparison approach, in which factors were added incrementally to a base model, and improved fit was assessed using the likelihood ratio test (*anova()*). Interaction terms were assessed by comparing models containing all main effects and interaction terms with models in which the critical interaction term was removed. For all models, we included the most complete random structure that led to successful convergence.

Table 1 shows the model summary for the Beginner group. The model included main effects of quantity and speaking rate and their interaction, by-speaker intercepts, as well as by-speaker random slopes for quantity and speaking rate. The results showed that mean C2 intensity of CVCCV (geminates) was significantly lower than CVCV (β = .052, SE = .023, t = 2.264); the difference between CVVCV and CVCCV, however, was non-significant (β = .032, SE = .022, t = 1.445). Meanwhile, C2 had significantly greater intensity in fast than in normal speech (β = .033, SE = .016, t = 2.091), and in normal than in slow speech (β = −.054, SE = .013, t = −4.316). The interaction terms reflect that the mean intensity difference between CVVCV and CVCCV (β = .032, SE = .009, t = 3.568) and that between CVCV and CVCCV (β = .028, SE = .009, t = 3.063) are greater for slow than for normal speech rate. For both native speakers ($\chi^2(2)$ = 1.112, p = .573) and advanced learners ($\chi^2(2)$ = 1.729, p = .421), the effect of Quantity was non-significant.

2 Manual inspection of stop closure

Utterances with an inserted /t/ were manually culled by two phoneticians (including the first author) who spoke Japanese as an L2 (achieving the highest Level 1 and Level N2 in the Japanese Language Proficiency Test, respectively). In deciding whether /t/ was inserted, they checked for silence (absence of energy; see Figure 2) in the spectrogram which indicates stop closure. Altogether they identified 230 utterances (42.6% of 540 geminates produced by learners) with an inserted /t/.

We used a decision tree to determine predictive factors of /t/-insertion by the participants. The decision tree was constructed using IBM SPSS 26 (CHAID algorithm). The training set consisted of 80% of the data. Five factors, namely (1) speaker ID, (2) speaker group, (3) speaking rate, (4) average syllable duration, and (5) surrounding vowels, were incorporated to train the decision tree model for prediction. The accuracy of the decision tree was 79.6% (78 hits of the test set, N = 98).

Figure 3 shows that both speaker ID and speaking rate were predictive factors of /t/-insertion, whereas the identity of surrounding vowels was not. The left child node under the root note corresponds to the speakers who tended not to insert /t/, with 8 advanced learners and 3 beginners; the right child node represents the speakers who tended more to insert /t/, with 5 beginners and 4 advanced learners. The beginners (M = 57.4%, SD = 28.6%) inserted /t/ more often than the advanced learners (M = 32.7%, SD = 24.3%) did. Meanwhile, the terminal nodes combined suggest that faster speech was predicted to yield fewer /t/-insertion cases than otherwise.

3 Relative duration of /s/

As weaker C2 intensity (see Figure 1) can indicate either false geminates or generally weakened frication, we further examined the relative duration of /s/ in C2 of all speaker groups (including both utterances with and without /t/-insertion). Figure 4 shows that for the native speakers, /s/ occupied the entire duration of C2; for the advanced learners it occupied 97.2% (SD = 9.8%); for the beginners it occupied 92.2% (SD = 17.2%). The difference between advanced and beginner learners was significant according to Welch's t-test, $t(11.756) = 2.848$, $p = .007$ (one-tailed). This shows that the beginners' production indeed contained a greater proportion of stop closure without frication. Furthermore, greater variability of relative /s/ duration was observed in slower speech in general, with 7.3% SD for fast speech, 10.6% for normal speech, and 16.4% for slow speech.

V Discussion

1 Summary of findings

We set out to examine whether Cantonese learners' production of long /s/ in Lee and Mok (2018) was genuine or a [t˺]–/s/ cluster. Table 1 and Figure 1 showed that for the beginners mean C2 intensity in the geminate condition was lower than that in the singleton condition, consistent with the view that there was a period of complete closure during gemination, i.e. a [t˺]–/s/ cluster. For the advanced learners and the native speakers, mean C2 intensity was not significantly different, suggesting that turbulence continued

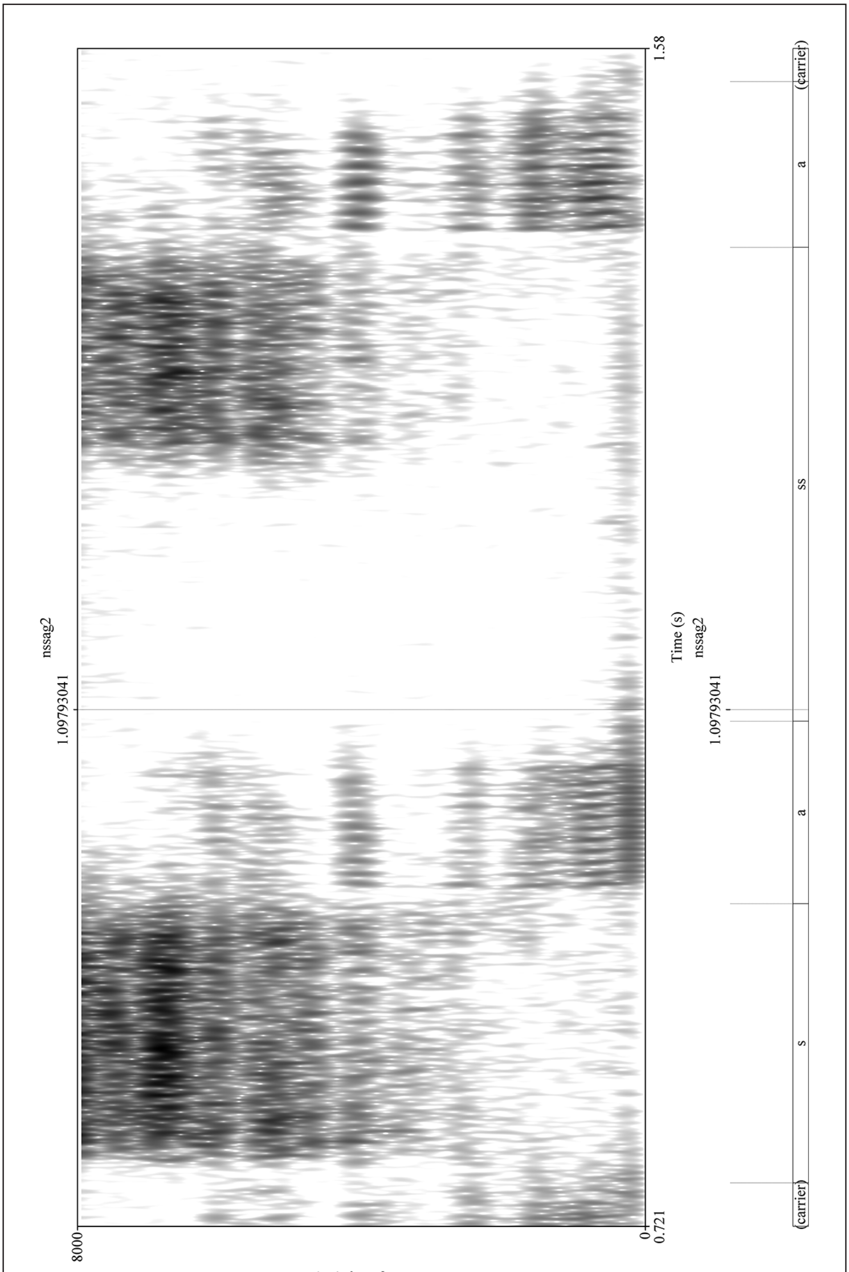

Figure 2. Spectrogram of *sassa* produced by a beginner learner at slow speaking rate.
Note. The segment 'ss' starts with a period of silence without acoustic energy.

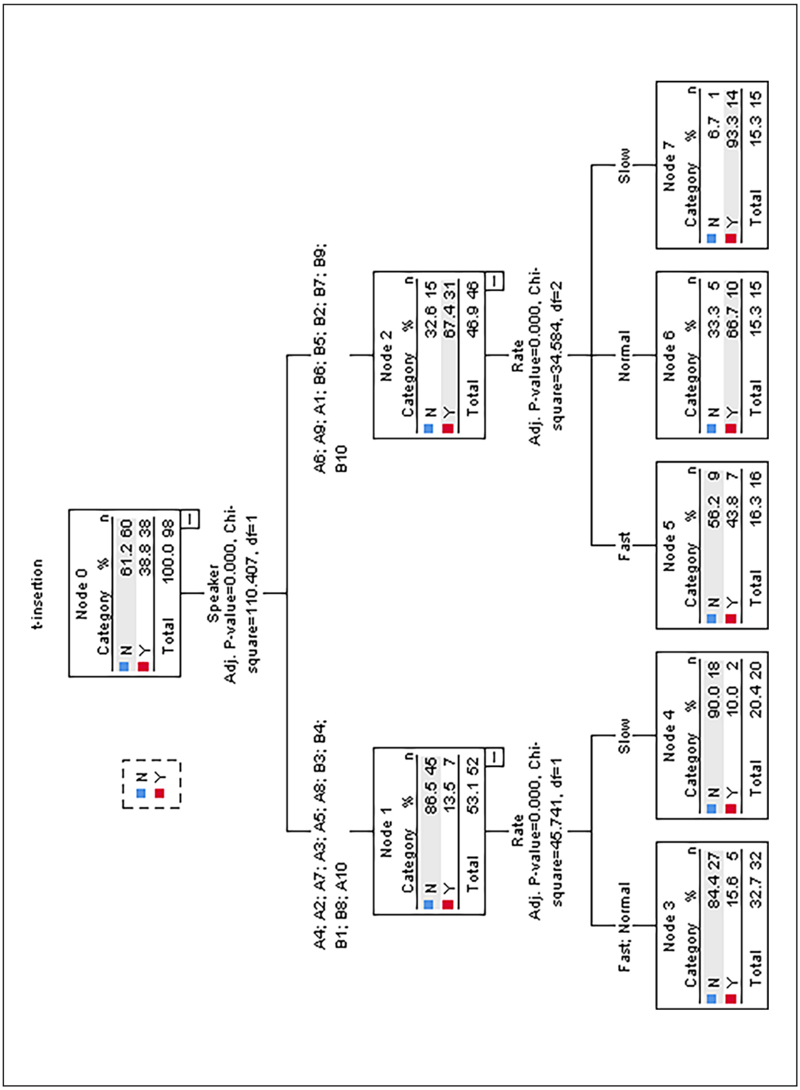

Figure 3. Decision tree predicting cases of /t/-insertion by the learner groups.
Note. 'Y' represents cases of /t/-insertion, whereas 'N' represents otherwise (colour online).

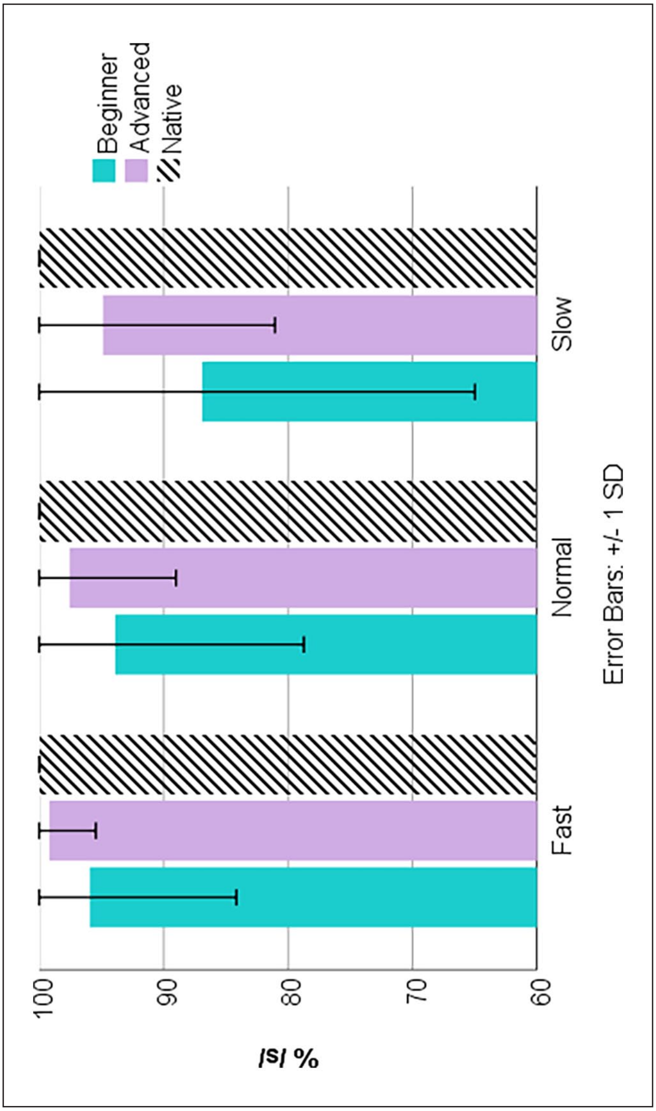

Figure 4. Barplot comparing percentage of /s/ duration in C2 of different speaker groups and speaking rates (N = 2,025).

throughout the entire C2, i.e. genuine gemination. Therefore, hypothesis 1 was only partially supported as the advanced learners did not behave like their beginner counterparts in their mean C2 intensity. We further observed an effect of speech rate on /t/-insertion, with more insertion cases in slower speech, thus supporting hypothesis 2.

2 /t/-insertion as an effective transitional strategy

It is interesting why the beginners improvised [t˺]–/s/ as a transitional strategy, rather than using /s.s/, especially given /s/ in the coda position had already been acquired in their L2 English. In fact /s/ is not described to be a challenging sound in the coda position for Hong Kong English (Peng and Setter, 2000; Setter et al., 2010). Even with extensive L2 English experience, they still resorted to an L1 false geminate to achieve the long target constriction duration of the geminate condition. Two points concerning this behavior are noteworthy here. First, their resorting to L1 as an improvised strategy to produce this new contrast type may be taken to illustrate just how challenging it is for them to produce genuine long consonants, which are absent in both Cantonese and English. As beginners three months into learning Japanese, they needed the easiest short-cut to producing long constriction duration from their first language, thus [t˺]–/s/ cluster from L1 Cantonese rather than /s.s/ as permitted in L2 English. In fact, that some advanced learners were still resorting to false geminates shows exactly how robust this strategy is. Second, while false geminates may not be genuine, the beginners were right on aiming for long constriction duration: the salient perceptual cue to long consonants (Han, 1992).

Our interpretation may also shed light on learners from other L1 backgrounds, e.g. English. For example, Hayes-Harb and Masuda (2008) reported a listening and production experiment in which English speakers were taught Japanese non-words contrasting in C2 quantity. They found that while native Japanese speakers outperformed the experienced learners, who in turn outperformed the inexperienced learners, both learner groups were able to lexically distinguish singleton and geminate words to some extent in the listening task. This is interesting because English does not have consonant quantity contrasts except at the phonotactic level (see Cantonese), while duration is only one of multiple acoustic cues to tense vs. lax vowel distinctions (House, 1961). Based on our discussion above, it is thus possible that the learners in Hayes-Harb and Masuda (2008) were in fact encoding real geminates as false ones (i.e. CVC.CV), a phonological structure they have in their L1.

3 A fresh look at Lee and Mok, 2018

Our findings have shed new light on a puzzle left behind in Lee and Mok (2018): Does partial use of duration in L1 vowel quantity distinctions benefit the acquisition of L2 consonant quantity contrasts? Do L1 false geminates benefit the acquisition of L2 underlying geminates? Although the durational ratios in Lee and Mok (2018) seemed to suggest a positive answer to both questions, the analysis of intensity trajectories revealed that the beginners were in fact improvising a transitional strategy, i.e. false geminates. The present study shows that in the case of Cantonese learners of Japanese, their having vowel quantity contrasts in L1 does not mean that they can readily use constriction duration to contrast singleton and geminate consonants as beginner learners. Meanwhile, the fact that their L1 permits false geminates echoes their transitional strategy of producing a coda [t˺]–/s/ cluster in place of a geminated /s/. Therefore, as far as beginner Cantonese learners are concerned, their production of L2 Japanese geminates may be better attributed to L1 false geminates (in turn, the fact that stops are permitted as codas) instead of vowel quantity contrasts. It would be interesting to verify our findings with speakers whose L1 does not permit codas ('simple syllable structure'; see Maddieson, 2005).

As for the advanced learners, the present findings suggest that they were indeed producing real geminates. This means that, with enough exposure, it is possible for Cantonese learners to use constriction duration to mark consonant quantity contrasts. Note that even so their geminates are not fully native-like: Lee and Mok (2018) have shown that (1) their durational ratios are different from the native speakers, and that (2) they were unable to exaggerate durational differences between CVCV and CVCCV in slower speech like their native speaker counterparts.

Taken together, the present findings and those in Lee and Mok (2018) may suggest the following with regard to the feature vs. category issue: (1) length contrasts confined to only a small subset of vowel pairs in L1 sufficed to facilitate beginner learners to distinguish between long and short vowels in L2 Japanese (see Lee and Mok, 2018: Table 7), and (2) the same vowel length contrasts did not help beginners to readily form an accurate representation of geminate consonants. The first point (1) is somewhat reminiscent of the English speakers in McAllister et al. (2002) who outperformed Spanish speakers due to their partial use of duration in vowel length contrasts. For Cantonese, having length contrasts in only a few L1 vowel pairs allowed the beginners to tell apart long vs. short in other L2 vowels, consistent with a featural prediction. For (2), the present findings lend further support to the view that this length feature is also constituent specific: having L1 vowel length contrasts does not readily help beginner learners to form an accurate representation for geminate consonants, somewhat comparable to German listeners not being as good at discriminating L2 Italian consonant length contrasts (Altmann et al., 2012).

4 Methodological implications

Our findings also show that learners' production can deviate from that of native speakers in unexpected ways. For Cantonese learners of Japanese, their non-native-like production of geminates could not be spotted in Lee and Mok (2018) using well-established durational ratios, because their transitional strategy for this L2 distinction was highly successful in terms of achieving target duration patterns. This had thus led Lee and Mok (2018) to believe that false geminates in Cantonese led to genuine positive transfer to these learners' Japanese speech. The present study has revealed the 'true colours' of these learners' geminates: false geminates like in their L1. To correctly understand L2 speech production data, it is necessary to use a wider range of measurements than for analysing L1 data. Failing to do so can result in a distorted view of the data and in turn misled interpretation of the findings.

5 Implications for teaching

Considering that the learners will eventually shift to authentic geminates anyway when they become more advanced, it may be advisable for teachers to refrain from correcting /t/-insertion when working with Cantonese-speaking beginners. Allowing them to map geminates onto an existing L1 category helps them focus on other more challenging aspects of their new language. The same may be said for other 'foreign accent features' that do not impede understanding; after all, unless communication is affected, whether to sound native-like should be the learner's choice (Levis, 2005) .

In this study we have shown acoustic evidence that Cantonese-speaking beginners may insert /t/ when producing a geminated /s/. Future researchers could verify our findings with articulatory data such as ultrasound. Furthermore, it would be interesting to investigate the effectiveness of /t/-insertion by recruiting native listeners for perception experiments.

VI Conclusions

In this article, we revisited Lee and Mok's (2018) data and examined whether Cantonese learners' production of geminates was genuine (i.e. lengthened constriction) or was the result of an unreleased coda + homorganic initial consonant sequence. Analysis of the mean intensity of geminated /s/ revealed that the latter was true for the beginners, but not so much for the advanced learners. In other words, while both learner groups appeared to achieve duration ratios to mark quantity contrasts, the beginners did so using phonotactic means; but they could eventually acquire the native production strategy of lengthening constriction duration of C2 by the time they became advanced learners. Furthermore, for both learner groups more instances of /t/ insertion were observed in slower speech. We conclude this study by calling for the use of a wider range of measurements in analysing L2 production data for a more comprehensive and accurate understanding of learners' production strategies.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work described in this article was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. ECS 28605120) awarded to the first author. We thank Zhiqiang Zhu (CUHK) for helping with a part of the data analysis.

ORCID iDs

Albert Lee https://orcid.org/0000-0002-3224-5788

Peggy Mok https://orcid.org/0000-0002-9284-6083

Notes

1. See Meister et al. (2015), who reported an asymmetry in the opposite direction.
2. There are alternative phonemic analyses with more short vs. long vowel pairs. For example, in an 11-vowel system (e.g. Bauer and Benedict, 1997; Kao, 1971), there can be four pairs of vowels contrasting in quantity. However, only the /aː/–/ɐ/ pair is unambiguously a short vs. long contrast which appears in all environments.
3. Mandarin does not have vowel or consonant length quantity contrasts except to the extent that the neutral tone syllables are considered unstressed (Shen, 1992). As all the learners had been exposed to English stress anyway, we did not consider this an additional confound.

References

Altmann H, Berger I, and Braun B (2012) Asymmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research* 28: 387–413.

Asu EL, Teras P, and Khatiwada R (2009) Illustrations of the IPA: Estonian. *Journal of the International Phonetic Association* 39: 367–372.

Bates DM, Mächler M, Bolker BM, and Walker SC (2015) Fitting linear mixed-effects models using {lme4}. *Journal Of Statistical Software* 67: 1–48.

Bauer RS and Benedict PK (1997) *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.

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

Chang CB (2019) The phonetics of second language learning and bilingualism. In: Katz WF and Assmann PF (eds) *The Routledge handbook of phonetics*. Abingdon: Routledge, pp. 427–447.

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

Chiu F, Fromont L, Lee A, and 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)*, 1008. Glasgow.

Chomsky N and Halle M (1968) *The sound pattern of English*. Cambridge, MA: MIT Press.

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

Flege JE and Port RF (1981) Cross-language phonetic interference: Arabic to English. *Language and Speech* 24: 125–146.

Fry DB (1955) Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America* 27: 765–768.

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

Han MS (1994) Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96: 73–82.

Hancin-Bhatt B (1994) Segment transfer: A consequence of a dynamic system. *Second Language Research* 10: 241–269.

Hayes-Harb R and Masuda K (2008) Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research* 24(1): 5–33.

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

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

House AS (1961) On vowel duration in English. *Journal of the Acoustical Society of America* 33: 1174–1178.

Jiang N (2018) *Second language processing: An introduction*. Abingdon: Routledge.

Jongman A (1989) Duration of frication noise required for identification of English fricatives. *Journal of the Acoustical Society of America* 85: 1718–1725.

Kao DL (1971) *Structure of the syllable in Cantonese*. Berlin: Mouton.

Kawahara S (2015) The phonetics of *sokuon*, or geminate obstruents. In: Kubozono H (ed.) *Handbook of Japanese phonetics and phonology*. Berlin: Mouton de Gruyter, pp. 43–77.

Kuhl PK (1992) Psychoacoustics and speech and prototypes. In Werner LA and Rubel EW (eds) *Developmental psychoacoustics*. Washington, DC: American Psychological Association, pp. 293–332.

Kuhl PK (1994) Learning and representation in speech and language. *Current Opinion in Neurobiology* 4: 812–822.

Lee A and Mok PKP (2018) Acquisition of Japanese quantity contrasts by L1 Cantonese speakers. *Second Language Research* 34: 419–448.

Lee B, Guion SG, and Harada T (2006) Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28: 487–513.

Levis JM (2005) Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39: 369–377.

Maddieson I (2005) Syllable structure. In: Haspelmath M, Dryer MS, Gil D, and Comrie B (eds) *The world atlas of language structures*. Oxford: Oxford University Press, pp. 54–57.

McAllister R, Flege JE, and 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: 229–258.

Meister E, Nemoto R, and Meister L (2015) Production of Estonian quantity contrasts by Japanese speakers. *Eesti Ja Soome-Ugri Keeleteaduse Ajakiri* 6: 79–96.

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 preceding/following vowels]. *Journal of the Phonetic Society of Japan* 7: 70–76.

Oh GE and Redford MA (2012) The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40: 82–91.

Peng L and Setter J (2000) The emergence of systematicity in the English pronunciations of two Cantonese-speaking adults in Hong Kong. *English World-Wide: A Journal of Varieties of English* 21: 81–108.

Setter J, Wong SPC, and Chan HS (2010) *Hong Kong English*. Edinburgh: Edinburgh University Press.

Shen XS (1992) Mandarin neutral tone revisited. *Acta Linguistica Hafniensia* 24: 131–152.

Takiguchi I, Takeyasu H, and 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.

Tsukada K, Hirata Y, and 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.

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

Turk AE, Nakai S, and Sugahara M (2006) Acoustic segment durations in prosodic research: A practical guide. In: Sudhoff S, Lenertová D, Meyer R, et al. (eds) *Methods in empirical prosody research*. Berlin: Walter de Gruyter, pp. 1–28.