Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Production of Non-native Quantity Contrasts by Native Speakers of Cantonese, English, French, and Japanese

Albert Lee[1], *Yasuaki Shinohara*[2-4], *Faith Chiu*[5], *and Tsz Ching Mut*[1]

[1]The Education University of Hong Kong, [2]Waseda University, [3]City University of New York, [4]University of Delaware, [5]University of Glasgow

albertlee@eduhk.hk, y.shinohara@waseda.jp, faith.chiu@glasgow.ac.uk, tcmut@eduhk.hk

Abstract

In this study we compared the duration ratios of native speakers of Cantonese, English, French, and Japanese who produced non-native phonemic quantity contrasts in Japanese (two-way) and Estonian (three-way). These four L1 backgrounds differ in terms of the extent to which duration is used to mark quantity contrasts (e.g. short vs. long), ranging from non-phonemic (i.e. French) to systematic two-way (i.e. Japanese). A shadowing task was used to elicit participants' production. Estonian and Japanese stimuli (N = 360) were played in two separate blocks. The participants wore a pair of headphones in a quiet room and repeated the words they heard. The results showed that all participant groups were able to tell apart the quantity conditions, though for the Estonian target words Short vs. Long were better differentiated than Long vs. Over Long. The unexpectedly good performance of the French speakers in perception tasks [1] was partially replicated, as was the relatively poor performance of the Cantonese speakers. The theoretical implications of these findings are discussed.

Index Terms: phonemic quantity, shadowing task

1. Introduction

L2 phonemic quantity holds the key to answering some important questions in L2 phonological acquisition. Specifically, in this line of work researchers ask whether learning L2-specific phonetic contrasts is affected by L1-L2 differences at the level of discrete sound categories (i.e. discrete segments like /p/ vs. /b/) or continuous phonetic dimensions (or 'features' like VOT, *cf.* [2]) (see review in [3]). As phonemic quantity contrasts are primarily based on duration (continuous dimension), it offers an excellent opportunity for understanding the category vs. feature problem.

In [4], researchers compared native speakers of Estonian, English, and Spanish in terms of their mastery of Swedish two-way quantity distinctions (short vs. long). Spanish speakers do not use duration to mark phonemic quantity contrasts even as a secondary cue, unlike English which has short vs. long vowels (e.g. *bit* vs. *beat*) although duration is only one of the acoustic cues (alongside vowel quality) [5]. The results demonstrated that both English and Spanish speakers identified Swedish vowels less well than the Estonian speakers, who have three-way quantity contrasts in their L1 vowels and consonants (i.e. short, long, and overlong). However, the English speakers showed slightly better performance than the Spanish speakers. This suggests that an L1 background with more extensive use of duration as a quantity cue is beneficial for L2 quantity acquisition.

Meanwhile, although some quantity-sensitive learners can produce correct timing patterns in the target language (e.g. L1 Dutch-L2 Italian [6]), others do not (e.g. L1 Italian-L2 Japanese [7]). In the latter case, it could even be argued that L1 quantity contrasts hinder the acquisition of L2 timing control (*ibid.*). Furthermore, it is also possible that vowel quantity and consonant quantity are not equally easy to acquire for L2 learners [8], just as different pairs of quantity conditions (e.g. **CVCV** vs. **CVVCV**, **CVVCV** vs. **CVVVCV**) are not equally challenging to L2 speakers [9].

Despite the growing body of empirical studies in L2 quantity acquisition, several things have remained underexplored: (i) studies directly comparing more than two languages (except e.g. [4]), (ii) controlled proficiency levels of participants, and (iii) controlled influences of secondary cues. To this end, [1] compared native speakers of Cantonese, English, French, and Japanese who perceived non-native phonemic quantity contrasts in Japanese (two-way) and Estonian (three-way). Synthesised stimuli contrasting only in duration were used. These four L1 backgrounds differ in terms of the extent to which duration is used to mark quantity contrasts (e.g. short vs. long), ranging from non-phonemic (i.e. French) to systematic two-way (i.e. Japanese). The results showed that while Japanese listeners, who have systematic phonemic quantity contrasts in their L1, outperformed other listeners in discrimination and identification, their identification accuracy for *overlong* Estonian vowels and consonants was not as high as that for *long* Estonian vowels and consonants. Meanwhile, French listeners, who have no quantity contrasts in their L1 phonology, did not perform worse than the other groups (Cantonese and English, both having quantity contrasts to some degree) as predicted. These findings appear to be at odds with [4], and thus call for further verification with other methods.

In this paper, we revisited these four L1 groups using a shadowing production task of phonemic quantity contrasts in Japanese and Estonian. The production data were analysed using a series of duration ratios as in [10]. Larger ratios will indicate participants' better ability to contrast a given quantity pair.

Based on [1], we tested the following hypotheses: (H1) Japanese speakers' duration ratios are the greatest whereas (H2) those of Cantonese speakers the smallest; (H3) for Japanese speakers duration ratios for Estonian long vs. overlong contrasts will be smaller than those for short vs. long ones.

2. Methodology

2.1. Participants

Twenty native speakers of Cantonese (11 females, aged 19 to 50), 20 Japanese (11 females, aged 18 to 25), 20 English (10 females, aged 19 to 49) and 15 French (14 females, aged 18 to 24) were recruited. All participants also took part in [1]. They had no (history of) hearing or language impairments. All

10.21437/SpeechProsody.2024-29

Cantonese participants spoke English and Mandarin as L2 with varying proficiency levels. The Japanese participants had been learning English at school but did not speak it in their daily lives, nor did they have experience living outside Japan for over two months. The French participants were university students studying in the United Kingdom. No participant reported to speak any other language.

2.2. Stimuli

Synthetic pseudo-Estonian (Estonian henceforth) and pseudo-Japanese (Japanese henceforth) word stimuli were generated using VocalTractLab 2.2 [11]. There were 75 (nonce) Estonian words (15 **CVCV** base real words × 5 quantity conditions: **CVCV, CVVCV, CVVVCV, CVCCV, CVCCCV**), and 45 Japanese words (15 **CVCV** nonce words × 3 quantity conditions: **CVCV, CVVCV, CVCCV**). These were spoken in three synthetic voices, differing in fundamental frequency (male 110 Hz, male 150 Hz, female 200 Hz), vocal tract length, and voice quality. The actual duration of each segment is listed in Table 1 and Table 2. For all data collection sites, we used e-Prime to present stimuli and record responses.

Table 1: *Typical segment duration of Estonian quantity conditions (in ms) (based on* [12]*).*

	CVCV	CVVCV	CVVVCV	CVCCV	CVCCCV
C1	80				
V1	90	170	240	90	
C2	80			140	200
V2	140				

Table 2: *Typical segment duration of Japanese quantity conditions (in ms) (based on* [10]*).*

	CVCV	CVVCV	CVCCV
C1	80		
V1	70	130	70
C2	70	70	130
V2	140		

2.3. Procedure

In the shadowing task, Estonian and Japanese stimuli (N = 360) were played in two separate blocks. Participants heard one nonce word stimuli through a pair of headphones and repeated the word they heard.

2.4. Data analysis

In total, 26,510 utterances were analysed. They were labelled by the segment, of which duration were then used to calculate the following ratios: (Vocalic) V1 duration ratio (**V:VV**, e.g. *kato:kaato* [13]), Word duration ratio (**CVCV:CVVCV**, e.g. *kato:kaato* [13]), Vowel-to-word duration ratio (**V:CVCV**, e.g. *a:kato* [13]), Closure duration ratio (**C:CC**, e.g. *kato:katto* [14]), (Consonantal) V1 duration ratio (**V:V**, e.g. *kato:katto* [15]).

3. Results

3.1. Average segment duration

The mean segment duration of target words was first checked to assure that the length of segments differed according to the quantity conditions.

Figure 1 shows that for Japanese targets, **V1** is longer in **CVVCV** and **C2** is longer in **CVCCV**. Figure 2 shows that for Estonian targets, **V1** is longer in **CVVCV** than in **CVCV**, and **C2** is longer in **CVCCV** than in **CVCV**. Similarly, **V1** is longer in **CVVVCV** than in **CVVCV** and **C2** is longer in **CVCCCV** than in **CVCCV**.

A two-way ANOVA was performed for Estonian and Japanese mean segment duration data, with the fixed factors of Target Language (Estonian, Japanese), Participants' L1 (Cantonese, English, French, Japanese) and Stimulus Type (Japanese: **Short**, Long consonants ('**LC**' henceforth), Long vowels ('**LV**' henceforth); Estonian: **Short, LC, LV**, Over Long consonants ('**OverLC**' henceforth), Over Long vowels ('**OverLV**' henceforth)), and their interactions. Stimulus Type ($F(4,5728) = 71.221$, $p < .001$), Participants' L1 ($F(3,5728) = 3.982$, $p < .01$) and Target Language ($F(1,5728) = 26.046$, $p < .001$) have significant main effects. Post-hoc Bonferroni tests confirmed that for all L1 groups, all quantity conditions were significantly different from one another in mean segment duration (all $p < .001$).

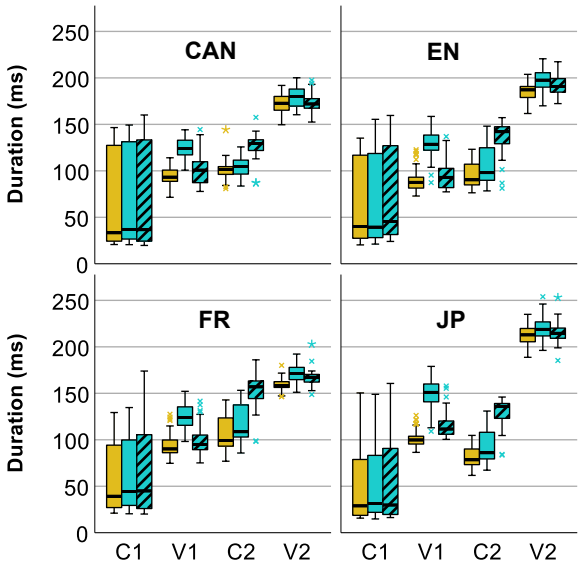

Figure 1: *Duration (ms) (Japanese) by Qty and L1. Gold = S, turquoise = L, solid = V, dashed = C.*

3.2. Short vs. long and long vs. over long vowels

A ratio greater than 1:1 means that the longer vowels are longer than the shorter vowels. As shown in Figure 3, for all speaker groups, **V1** Duration Ratio exceeded the 1:1 threshold. The same holds true for Word Duration Ratio -- ratios of all speaker groups exceeded 1:1 (see Figure 4), meaning both long and over long words are longer than the short words across groups. Specifically, **LV** and **LC** were longer than **Short** words. Likewise, **OverLV** and **OverLC** words were longer than **LV** and **LC** words.

We performed two-way ANOVA to the Estonian stimuli with the fixed factors of Participants' L1 and Stimulus Type (**CVCV:CVVCV, CVVCV:CVVVCV**) and their interactions, and one-way ANOVA to the Japanese stimuli (see Table 3 for test results for Japanese data).

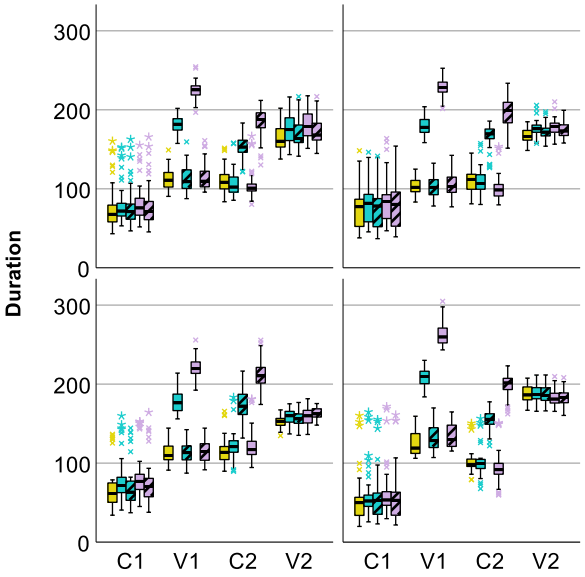

Figure 2: *Duration (ms) (Estonian) by Qty and L1. Lilac = OverL*

Table 3: *ANOVA test results.*

Duration ratios of (Japanese data)				
	Vocalic	Word	Consonantal	
	V1	-	C2	V1
Stimulus Type				
df	-	1	-	-
MS		< .001		
F		0.043		
Participants' L1				
df	3	3	3	3
MS	.147	.036	.389	.035
F	2.086	6.867 ***	6.522 ***	2.503

Regarding **V1** Duration Ratio (see Figure 3), the two-way ANOVA revealed that only Stimulus Type had a significant main effect ($F(1,142) = 173.282$, $p < .001$). Post-hoc Tukey test confirmed that that all speaker groups differentiated **LV** from **Short** vowels better than **OverLV** from **LV** (all $p < .001$).

As for Word Duration Ratio, the two-way ANOVA showed that only Participants' L1 had a main effect on the duration of Japanese words ($F(3,142) = 6.867$, $p < .001$), whereas only Stimulus Type has a main effect on the duration of Estonian words ($F(3,284) = 85.646$, $p < .001$). Post-hoc Tukey tests confirmed that for the Estonian stimuli, all speaker groups differentiated **LV** from **Short** vowels better than **OverLV** from **LV** (all $p < .001$); for the Japanese stimuli, Cantonese speakers differentiated **Short** from **LV** worst ($p < .01$), whereas the other groups did not differ from one another significantly.

3.3. Singleton vs. geminate consonants

The production of singleton vs. geminate consonants was analysed in terms of **C2** duration ratio (i.e. Closure Duration Ratio) as well as that of neighbouring **V1**. As shown in Figure 5, all speaker groups exceeded the 1:1 threshold in all consonantal pairs in Estonian and Japanese stimuli, meaning that phonemically longer consonants do have longer duration than shorter ones.

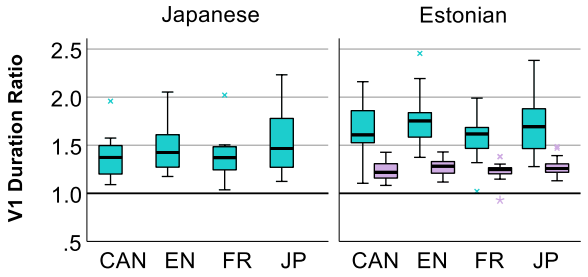

Figure 3: *V1 duration ratio (CV:CVV & CVV:CVVV). Solid = vowels. Turqoise = Short:L, Lilac = L:OverL.*

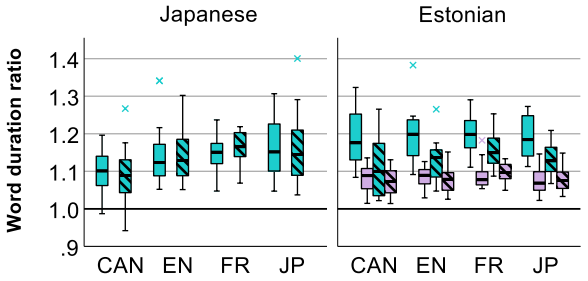

Figure 4: *Duration of target words. Dashed = Consonants.*

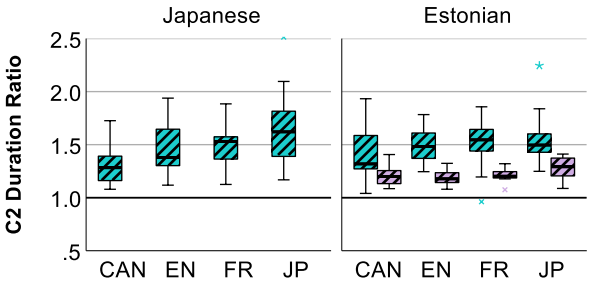

Figure 5. *C2 duration ratio.*

The two-way ANOVA including the fixed factors of Participants'1 L1, Stimulus Type and their interactions indicated that both Stimulus Type ($F(1,142) = 105.821$, $p < .001$) and Participants' L1 ($F(3,142) = 3.369$, $p < .05$) had main effects on the duration of Estonian **C2**. Post-hoc Tukey test revealed that for all L1 groups, speakers differentiated **LC** from **Short** consonants significantly better than **OverLC** from **LC** (for Cantonese, $p = .001$; for other groups, $p < .001$). Post-hoc Tukey test of one-way ANOVA for Japanese stimuli suggested that Japanese speakers differentiated **LC** from **Short** consonants significantly better than the Cantonese group ($p < .001$).

The effect of consonant quantity on the duration of **V1** was also examined (see Table 4). Although **V1** is 11% longer before a geminate in *natural* Japanese [15] (but not in the present stimuli), we found that 32% of our speakers did consistently lengthen **V1** across target languages and stimulus types. Post-hoc Tukey test on the two-way ANOVA (Participants' L1, Stimulus Type as fixed factors, and their interactions) suggested that Estonian **V1** duration ratios (**Short** vs. **LC**) of Japanese and

English speakers were significantly different (p = .032).

Table 4: *Duration ratio of (consonantal) V1.*

	Estonian				Japanese	
	Short vs. LC		LC vs. OverLC		Short vs. LC	
	Ratio	SD	Ratio	SD	Ratio	SD
CAN	1.02	0.10	1.03	0.06	1.08	0.14
EN	0.98	0.07	1.04	0.07	1.06	0.09
FR	1.01	0.10	1.01	0.07	1.04	0.08
JP	1.06	0.08	1.02	0.05	1.14	0.14

4. Discussion

We set out to test the following hypotheses: (H1) Japanese speakers' duration ratios will be the greatest whereas (H2) those of Cantonese speakers the smallest; (H3) for Japanese speakers' duration ratios for Estonian long vs. overlong contrasts will be smaller than those for short vs. long ones.

For Japanese targets, we found that Japanese speakers differentiated **LC** from **Short** consonants significantly better than the Cantonese group in terms of both Word Duration Ratio and C2 duration ratio; otherwise, they were not found to outperform any other L1 group. H1 is thus refuted. Likewise, Cantonese speakers were not otherwise found to perform less well than the other groups, thus refuting H2. Nevertheless, for both vowels and consonants, Japanese speakers as well as all other groups differentiated short from long better than long from over long, thus supporting H3. All in all, the relative performance of the four L1 groups in [1] was partially replicated, only that the difference among groups was much less obvious in the present production experiment (i.e. not observed in most duration ratios).

Three observations can be made from the present shadowing task. Firstly, largely regardless of L1 background, all naive speaker groups were able to tell apart different quantity conditions in their production. This is despite their relative performance in the discrimination and identification tasks in [1]. Secondly, what is interesting is that while the segments of interest were longer in relevant quantities, the overall duration of each word was similar, meaning that in this task speakers were maintaining a relatively stable duration at the word level. This is in contrast to [10], where word duration was 24% to 34% longer for **CVVCV** words than **CVCV**. Thirdly, 32% of the speakers lengthened **V1** before a geminate, despite the absence of this secondary cue in the synthesised stimuli. Typologically, it is more common to shorten **V1** before a geminate [16], hence it is unclear why these speakers spontaneously lengthened it in Table 4.

Although the relative performance of the four L1 groups largely followed the order of Japanese > English = French > Cantonese in the perception tasks of [1], this pattern became less obvious in the production data here. We seem to find that for perception [1], only the Japanese group unambiguously outperformed the others, whereas the rest did not appear to be very different (excluding the Cantonese data); in production (i.e. shadowing synthesised stimuli), no one group consistently stood out across duration ratios. In the next step, we will examine if there is any relationship between production and perception for individual speakers.

Taking the present findings and [1] together, we conclude that understanding the relationship between L1 background and the acquisition of non-native quantity contrasts continues to be less than straightforward, despite the classic findings in [4]. Whereas it is possible to control for proficiency levels, like in the present study and in [1] (i.e. by testing naïve speakers), trying to control for confounding factors such as secondary cues is much harder. In [1], holding pitch cues constant appeared to have negatively impacted on the Cantonese listeners much more than other groups. Future studies should investigate more L1 backgrounds, and should also consider factors such as whether different groups exploit a given secondary cue similarly.

5. Conclusions

In this paper, we presented our findings of a shadowing task to explore the relationship between L1 backgrounds and the ability to differentiate quantity conditions in production. We found that contrary to their relative performance in perception tasks, our participants from numerous L1 backgrounds did not seem to differ from one another to the same extent in production. We call for more experimental studies comparing a wider range of language backgrounds to help us fully understand how L2 quantity contrasts are acquired.

6. Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China awarded to AL (Project No. ECS 28605120) and by the Waseda University Grant for Special Research Project (2019C-143) awarded to YS. We thank Mr. Mingyu Weng for preparing the stimuli.

7. References

[1] A. Lee, Y. Shinohara, F. Chiu, and T. C. Mut, "Perception of vowel and consonant quantity contrasts by Cantonese, English, French, and Japanese speakers," in *Proceedings of the 20th International Congress of Phonetic Sciences (ICPhS 2023)*, 2023, pp. 2477–2481.

[2] J. E. Flege and R. F. Port, "Cross-language phonetic interference: Arabic to English," *Lang. Speech*, vol. 24, no. 2, pp. 125–146, 1981.

[3] N. Jiang, *Second language processing: An introduction*. New York, NY: Routledge, 2018.

[4] R. McAllister, J. E. Flege, and T. Piske, "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian," *J. Phon.*, vol. 30, no. 2, pp. 229–258, 2002.

[5] A. S. House, "On vowel duration in English," *J. Acoust. Soc. Am.*, vol. 33, no. 9, pp. 1174–1178, 1961.

[6] B. de Clercq, S. Ellen, and C. Crocco, "Rosa versus rossa: The acquisition of Italian geminates by native speakers of Dutch," *Phrasis. Stud. Lang. Lit.*, vol. 50, pp. 3–29, 2014.

[7] C. Guillemot, "The role of L1 durational correlates in L2 acquisition: A production study of Japanese geminates by Italian, French and English L2 learners," 2018.

[8] H. Altmann, I. Berger, and B. Braun, "Asymmetries in the perception of non-native consonantal and vocalic length contrasts," *Second Lang. Res.*, vol. 28, no. 4, pp. 387–413, 2012.

[9] E. Meister, R. Nemoto, and L. Meister, "Production of Estonian quantity contrasts by Japanese speakers," *Eesti ja Soome-Ugri Keeleteaduse Ajak.*, vol. 6, no. 3, pp. 79–96, 2015.

[10] A. Lee and P. K. P. Mok, "Acquisition of Japanese quantity contrasts by L1 Cantonese speakers," *Second Lang. Res.*, vol. 34, no. 4, pp. 419–448, 2018.

[11] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PLoS One*, vol. 8(4), no. e60603, pp. 1–17, 2013.

[12] L. Meister and E. Meister, "Perception of the short vs. long phonological category in Estonian by native and non-native listeners," *J. Phon.*, vol. 39, no. 2, pp. 212–224, 2011.

[13] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *J. Phon.*, vol. 32, no. 4, pp. 565–589, 2004.

[14] M. S. Han, "The timing control of geminate consonants in Japanese: A challenge for nonnative speakers," *Phonetica*, vol. 49, pp. 102–127, 1992.

[15] M. S. Han, "Acoustic manifestations of mora timing in Japanese," *J. Acoust. Soc. Am.*, vol. 96, no. 1, pp. 73–82, 1994.

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