# PERCEPTION OF VOWEL AND CONSONANT QUANTITY CONTRASTS BY CANTONESE, ENGLISH, FRENCH, AND JAPANESE SPEAKERS

Albert Lee[1], Yasuaki Shinohara[2-4], Faith Chiu[5], 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] Glasgow University Laboratory of Phonetics, University of Glasgow
albertlee@eduhk.hk, y.shinohara@waseda.jp, faith.chiu@glasgow.ac.uk, tcmut@eduhk.hk

## ABSTRACT

This paper reports on the findings of a study which investigated Cantonese, Japanese, English and French listeners' ability to perceive non-native quantity contrasts. In these languages, duration is used to mark phonemic quantity contrasts to different degrees. We had native listeners of these languages listen to resynthesized Estonian nonce word stimuli in AXB discrimination and identification tasks. The stimuli contrasted in consonant and vowel quantity. The results showed that Japanese listeners, who have short vs. long contrasts in both vowels and consonants, outperformed the other listeners in discrimination and identification. However, 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. The theoretical implications of these findings are discussed.

**Keywords**: feature hypothesis, L2 phonology, phonemic quantity contrasts

## 1. INTRODUCTION

### 1.1. Background

The issue of L2 phonemic quantity contrasts has received increasing attention in recent years as it holds the key to answering some theoretical questions in L2 phonological acquisition. Specifically, this line of research asks whether learners' difficulty in learning L2-specific phonetic contrasts is affected by L1-L2 differences at the level of discrete sound categories (i.e. discrete segments) or continuous phonetic dimensions (or 'features') (see review in [1], §2.5).

Existing experimental results have painted a mixed picture of the relationship between L1 phonology and L2 quantity acquisition. Although generally learners seem to be able to tell apart short vs. long in the target language at least to some extent (e.g. [2] for Cantonese), they also deviate from native speakers in numerous ways (e.g. ability to exaggerate short vs. long differences at slower speech rates [2]), making answering the category vs. feature question less than straightforward. A further complication is that these studies often compared only one language pair at a time (e.g. [3] who looked at Mandarin-speaking learners of Italian), with learners of varying proficiency levels, using different experimental methods. Moreover, the languages tested also differed in terms of orthographic depth, with some where quantity is clearly marked in the writing system (e.g. Japanese) whereas for others the same is not true (e.g. Cantonese). Thus, to examine the relationship between L1 phonology and L2 quantity acquisition, ideally one should compare participants from L1 backgrounds that make quantity distinctions to varying degrees. Other factors should also be held constant, such as proficiency (e.g. naïve) and orthographic depth (e.g. naïve participants without literacy of the target language). This study is an attempt to make such direct comparison.

### 1.2. Phonemic quantity in Estonian

Estonian has three-way quantity distinctions in both consonants and vowels. There are three levels of quantity, namely short, long, and overlong. The short vs. long distinction is faithfully represented in the writing system, e.g. *sada* 'hundred' and *saada* 'send', but the long vs. overlong difference is not [4]. Table 1 (in §2 below) shows relative segment durations of typical disyllabic Estonian words of different syllable structures.

### 1.3. Line-up of L1 backgrounds

Here we tested native listeners of Japanese, English, Cantonese, and French, who use duration in their L1 phonology to different degrees. Japanese has systematic short vs. long differences in both vowels and consonants [5]. Both obstruent consonants (e.g. /kita/ 'came' vs. /kitta/ 'cut') and vowels (e.g. /obasan/ 'aunt' vs. /oba:san/ 'grandmother') phonologically contrast in quantity, with duration being the primary acoustic cue [5]. English has short vs. long vowels (e.g. *bit* vs. *beat*) although duration is only one of the acoustic cues (alongside vowel quality) [6]. English also has false geminates at word boundaries such as *cat tail*. Cantonese has short vs.

long vowels but only limited to a small set of pairs (e.g. /ɐi/ vs. /aːi/). It also has false geminates at morpheme boundaries like in English. French has no phonemic quantity contrasts [7], and is said to be 'quantity insensitive'.

### 1.4. Experimental findings on L2 quantity acquisition

Experimental findings have pointed to a clear relationship between L1 phonology and mastery of L2 quantity distinctions. In [8], Estonian, English, and Spanish speakers were assessed on 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. The results demonstrated that although both the English and Spanish speakers identified Swedish vowels less well than the Estonian speakers, the English speakers showed slightly better performance than the Spanish speakers. This shows that an L1 background with more extensive use of duration as a quantity cue is beneficial for L2 quantity acquisition.

For those from a 'quantity-sensitive' L1 background (e.g. Japanese), acquiring more complex L2 quantity contrasts can be challenging. In [9] (replicated in [10]) Japanese learners of Estonian were able to distinguish between Estonian long vs. overlong consonants, but not long vs. overlong vowels, though both are absent in Japanese.

Interestingly, recent studies ([3], [11]) showed that arguably "quantity-insensitive" Mandarin learners were able to tell apart Italian short vs. long consonants. Similarly, Cantonese learners were found to be able to distinguish between Japanese short vs. long vowels and consonants in their production [2]. All in all, it seems that insofar as their ability to tell apart short vs. long is concerned, with enough exposure learners can acquire quantity distinction without major problems. What is interesting, however, is how participants from different L1 backgrounds *relatively* perform in a direct comparison, where known confounds are controlled.

### 1.5. Hypotheses

Our overarching goal is to understand the relationship between how far duration is used in L1 as a quantity cue and how successfully one can acquire L2 quantity contrasts. To this end, it is necessary to line up L1 backgrounds that make quantity distinctions to varying degrees. The stimuli and L2 proficiency of participants should also be controlled.

Here we tested the following hypotheses: (H1) Japanese listeners' perception accuracy in both discrimination and identification is the highest; (H2) French listeners' perception accuracy is the lowest; (H3) Japanese listeners would not be able to discriminate or identify the long vs. overlong contrast as well as they do the short vs. long contrast in Estonian.

## 2. METHODOLOGY

### 2.1. Participants

Twenty native listeners of Cantonese (11 female, aged 19 to 50), 20 Japanese (11 female, aged 18 to 25), 20 English (10 female, aged 19 to 49) and 15 French (14 females, aged 18 to 24) were recruited. They had no (history of) hearing or language impairments. All 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 Estonian nonce word stimuli were generated using VocalTractLab 2.2 [12]. We chose to use synthetic stimuli to control for non-durational secondary cues that could influence perception, e.g. pitch [13]. There were 75 (nonce) words (15 CVCV base real words × 5 quantity conditions: CVCV, CVVCV, CVVVCV, CVCCV, CVCCCV). 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 (based on [2]). For all data collection sites, we used e-Prime to present stimuli and record responses.

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

**Table 1.** Segment duration of Estonian stimuli (in ms)

### 2.3. Procedures

#### *2.3.1 AXB discrimination task*

In the AXB discrimination task, participants heard a sequence of three speech stimuli through a pair of headphones, and judged whether the middle one was the same as the first or the last one. They were asked to respond as quickly as possible.

*2.3.2 Identification task*

In each trial of the identification task, participants heard one speech stimulus, and identified the quantity of vowels and consonants. In the vowel block, participants chose from CVCV, CVVCV, and CVVVCV (e.g. pada, pa ː da, pa ː ː da); in the consonant block, they chose from CVCV, CVCCV, and CVCCCV (e.g. pada, pad:a, pad::a). The order of the two blocks was counterbalanced across participants.

## 3. RESULTS

*3.1 AXB discrimination task*

Figure 1 displays the AXB discrimination accuracy of the Estonian nonce word stimuli by Cantonese, English, French and Japanese listeners. We fitted logistic mixed effects models to the correct/incorrect binomial responses using the R package lme4 [14]. The model included the fixed factors of participants' L1 (Cantonese, English, French, Japanese), stimulus pair (short-long, long-overlong) and their interaction. Orthogonal contrasts were set for these categorical variables. Random intercept for participant was also included.

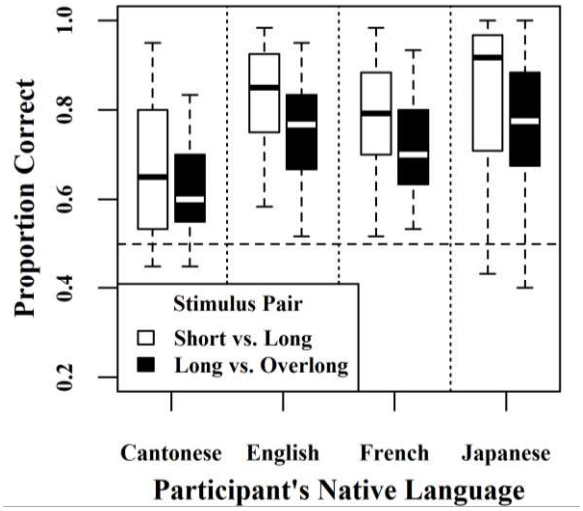


**Figure 1.** Discrimination accuracy of Estonian vowel and consonant quantity contrasts by Cantonese, English, French, and Japanese speakers. The horizontal dashed line represents 50% chance level.

The results demonstrated that Japanese listeners were significantly better at discriminating Estonian quantity contrasts than other groups of listeners, $\beta$ = 0.14, SE = 0.05, $z$ = 3.07, $p$ < 0.01. English listeners performed significantly better than Cantonese listeners, $\beta$ = -0.40, SE = 0.11, $z$ = −3.66, $p$ < 0.001, whereas French speakers did not perform significantly worse than Cantonese or English speakers, $\beta$ = 0.04, SE = 0.07, $z$ = 0.55, $p$ > 0.05. The stimulus pair effect (short-long vs. long-overlong discrimination) was significant, $\beta$ = -0.21, SE = 0.02, $z$ = -11.28, $p$ < 0.001, suggesting that the short vs. long contrast was easier to discriminate than the long vs. overlong contrast across L1 groups.

The two-way interaction of participant's L1 and stimulus pair demonstrated that the effect of stimulus pair was larger for Japanese speakers than other language groups, albeit marginally significant, $\beta$ = -0.02, SE = 0.01, $z$ = -1.90, $p$ = 0.058. This means that, compared to other language speakers, the short vs. long contrast was easier for Japanese speakers to discriminate than the long vs. overlong contrast. In addition, the stimulus pair effect was significantly larger for English speakers than Cantonese speakers, $\beta$ = 0.08, SE = 0.02, $z$ = 3.44, $p$ < 0.001.

Further analysis was conducted for each language group. The results demonstrated that the long vs. overlong discrimination was significantly more difficult than the short vs. long discrimination for all language groups, $p$ < 0.01.

*3.2 Identification task*

Figure 2 shows the identification accuracy of Estonian vowels and consonants quantity contrasts by the four language groups. The logistic mixed effects model included the fixed factors of participants' L1, quantity condition (short vs. long, long vs. overlong), and their interaction. The random effects were participant and talker voice.

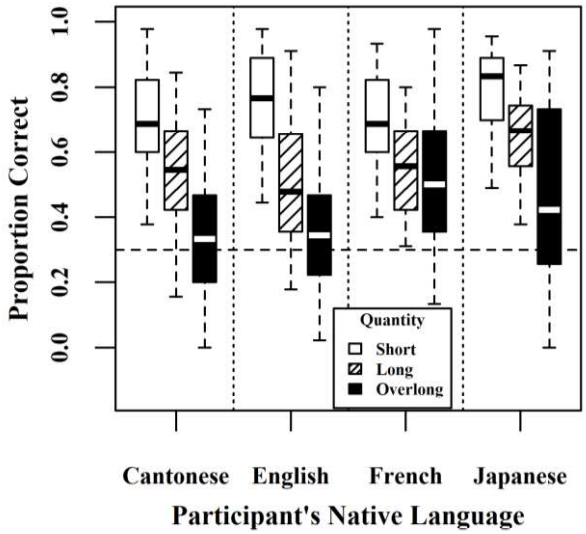


**Figure 2.** Identification accuracy of Estonian vowels and consonants in the quantity contrasts by Cantonese, English, French, and Japanese speakers. The horizontal dashed line indicates 33.3% chance level.

The logistic mixed effects model demonstrated that Japanese listeners were better at identifying the Estonian quantity contrasts than all other groups of listeners, $\beta$ = 0.07, SE = 0.03, $z$ = 2.64, $p$ < 0.01. The identification accuracy of short stimuli was higher

than that of the long stimuli, $\beta = 0.44$, SE = 0.07, $z = 6.38$, $p < 0.001$, and the overlong stimuli were less accurately identified than the other two quantity conditions (short & long), $\beta = 0.36$, SE = 0.04, $z = 10.05$, $p < 0.001$.

The two-way interactions of participant's L1 and quantity condition demonstrated that, although the quantity effect (overlong vs. short & long) was not significantly larger for Japanese speakers than other language groups, $\beta = 0.003$, SE = 0.01, $z = 0.55$, $p > 0.05$, that was larger for Cantonese and English groups than for French speakers, $\beta = -0.08$, SE = 0.01, $z = -8.53$, $p < 0.001$. Similarly, the quantity effect of the short vs. long contrast was larger for Cantonese and English groups than for French, $\beta = -0.04$, SE = 0.02, $z = 2.64$, $p < 0.01$, and that was larger for English than Cantonese group, $\beta = 0.11$, SE = 0.03, $z = 4.23$, $p < 0.001$.

Further analysis for each language group was conducted. For all L1 groups the identification accuracy of short stimuli was significantly higher than the long stimuli, $p < 0.001$, and that of overlong stimuli was significantly lower than the other two quantity conditions (short and long), $p < 0.001$.

## 5. DISCUSSION

This study set out to test the following hypotheses: (H1) Japanese listeners' perception accuracy is the highest; (H2) French listeners' perception accuracy is the lowest; (H3) Japanese listeners would not be able to discriminate or identify the overlong vowels and consonants as well as they do the long vs. short contrast. We found that Japanese listeners outperformed Cantonese, English and French listeners in both discrimination and identification, supporting H1. On the other hand, French listeners were not found to perform worse than other groups in any of the tasks, thus refuting H2. Discriminating and identifying overlong stimuli were more difficult than the short and long stimuli for Japanese speakers, and that was common for all language speakers, supporting H3.

Having systematic two-way quantity contrasts in their L1 may have allowed the Japanese listeners to discriminate similar two-way quantity contrasts of a non-native language. This is reminiscent of a previous study [15], in which a non-native vowel quantity contrast was perceived just as well as the native consonantal quantity contrast.

For the Cantonese listeners, their partial use of duration to mark vowel quantity contrasts (in only a small subset of vowels) in their L1 may have helped them discriminate non-native three-way quantity contrasts but not to the same extent as for the Japanese listeners. However, they performed less well in discrimination than English listeners, to whom duration is only one of multiple acoustic cues to vowel quantity. Note also that the English listeners were the only monolingual group not reporting to speak any L2. Our findings may appear to suggest that having many vowel pairs contrasting in quantity (i.e. English), even if duration is but one of multiple cues, is more beneficial to acquiring non-native quantity contrasts than having few quantity-contrasting vowel pairs in L1 (i.e. Cantonese). Needless to say, this speculation needs to be verified with more empirical evidence.

What is puzzling is why the French listeners performed much better than expected, despite the fact that their L1 is often deemed 'quantity-insensitive' [7]. As all the participants were naïve listeners of the target language, it is unclear what their good performance in the present study can be attributed to. The only conceivable factor that may set them apart from the other L1 groups is that they were immersed in an L2 at the time of testing, though how this might have contributed to the current findings is equally unclear. Also interesting is the significant interactions between L1 and quantity contrast, where the differences in identification accuracy of numerous quantity contrasts were significantly smaller for French than for other L1 groups. Further investigation is needed.

Although we have lined up multiple L1 backgrounds and compared listeners' perception accuracy in non-native word stimuli, we found that only Japanese listeners unambiguously outperformed the others. Meanwhile, the relative performance of Cantonese and English (partial quantity distinctions) as well as French ('quantity-insensitive') in different tasks does not seem to be easily attributable to their respective use of duration as a quantity cue (*contra* [8]). Although recent experimental findings (looking at two languages at a time) have improved our understanding of L2 quantity acquisition, the present direct comparison of *four* language backgrounds has shown that the picture is far from clear. A production study with these four listener groups is currently underway to shed further light on this.

## 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). We thank Mr. Mingyu Weng for assistance with creating the stimuli.

## 7. REFERENCES


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

[2] 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.

[3] Q. Feng and M. G. Busà, "Mandarin Chinese-speaking learners' acquisition of Italian consonant length contrast," *System*, vol. 111, no. 102938, pp. 1–13, 2022.

[4] I. Lehiste, "The function of quantity in Finnish and Estonian," 1965.

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

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

[7] P. A. Hallé, R. Ridouane, and C. T. Best, "Differential difficulties in perception of Tashlhiyt Berber consonant quantity contrasts by native Tashlhiyt listeners vs. Berber-naïve French listeners," *Front. Psychol.*, vol. 7, no. 209, pp. 1–16, 2016.

[8] 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.

[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, Y. Shinohara, and T. C. Mut, "Non-native length contrast perception by Japanese and Cantonese speakers," *Proc. Meet. Acoust.*, vol. 45, no. 060003, pp. 1–9, 2022.

[11] Q. Feng and M. G. Busà, "Acquiring Italian stop consonants: A challenge for Mandarin Chinese-speaking learners," *Second Lang. Res.*, 2022.

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

[13] Y. Minagawa, K. Maekawa, and S. Kiritani, "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]," *J. Phonetic Soc. Japan*, vol. 6, no. 2, pp. 88–97, 2002.

[14] D. M. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting linear mixed-effects models using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, 2015.

[15] 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.