Discrimination and Identification of Japanese Quantity Contrasts by Native Cantonese, English, French, and Japanese Listeners

Albert Lee[1], Yasuaki Shinohara[2], Faith Chiu[3] & Tsz Ching Mut[1]

[1]The Education University of Hong Kong (Hong Kong), [2]Waseda University (Japan), [3]University of Glasgow (UK)
albertlee@eduhk.hk, y.shinohara@waseda.jp, faith.chiu@glasgow.ac.uk, stcmut@gmail.com

One line of research in L2 phonological acquisition asks whether the relationship between L1 phonology and acquiring L2 sounds is based on discrete sound categories or on phonetic features. The latter implies that, for example, having a /t/-/d/ contrast in L1 would help learners acquire a new /p/-/b/ contrast in L2, as the primary acoustic cue to both contrasts is voice onset time [1]. However, in answering this question, most previous studies have not achieved to (i) compare more than two languages at a time, (ii) compare languages with similar orthographic depths, or (iii) test participants of proficiency levels comparable to those in other studies. Therefore, in this paper, we compared listeners from **four L1** backgrounds and compared their ability to perceive Japanese quantity contrasts. This paper is part of a larger project in which we also investigated their ability to perceive quantity contrasts in another language.

We selected to consider Japanese, English, Cantonese, and French in this study, as these languages use duration as a quantity cue to different degrees. Japanese has systematic short vs. long differences in both vowels and consonants [2]. Both obstruent consonants (e.g. /kita/ 'came' vs. /kitta/ 'cut') and vowels (e.g. /obasan/ 'aunt' vs. /obaːsan/ 'grandmother') phonologically contrast in quantity, with duration being the primary acoustic cue [2]. English has short vs. long vowels (e.g. *bit* vs. *beat*) although duration is only one of the acoustic cues (alongside vowel quality) [3]. Cantonese has short vs. long vowels but only limited to a small set of pairs (e.g. /ɐi/ vs. /aːi/). French has no phonemic quantity contrasts [4], and is said to be 'quantity insensitive'.

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. In addition, we are also interested in whether Cantonese or English speakers would perform better, as both languages partially use duration to cue quantity contrasts, but in different ways.

Twenty native listeners of Cantonese, 20 Japanese, 20 English and 15 French were recruited. They had no (history of) hearing or language impairments. They completed an AXB discrimination task and an identification task. Stimuli were 45 Japanese (nonce) words (15 CVCV base real words × 3 quantity conditions: CVCV, CVVCV, CVCCV) generated using VocalTractLab 2.2 [5]. These were produced 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 based on [6].

Figure 1 (left panel) displays the discrimination accuracy of Japanese quantity contrasts (short vs. long) by Cantonese, English, French and Japanese listeners. A logistic mixed effects model was fitted to the correct or incorrect responses. The fixed factor was participant's L1 and the random effects were participant, token, and presentation order (e.g., long-long-short, short-long-long). Orthogonal contrasts were set for the L1 factor. Results show that Japanese speakers' discrimination accuracy was significantly higher than the other three language groups, $\beta = 0.14$, $SE = 0.07$, $z = 2.03$, $p = .042$. English speakers yielded significantly higher accuracy than Cantonese speakers, $\beta = -0.47$, $SE = 0.16$, $z = -2.94$, $p < .01$.

The right panel displays identification accuracy of Japanese short and long stimuli by the same participants. We fitted another logistic mixed effects model to the correct or incorrect responses. The fixed factors were participants' L1, stimulus quantity (short, long), and their interaction. Orthogonal contrasts were set for those categorical variables. The random effects for participants and stimuli were also included.

Japanese listeners' identification accuracy was significantly higher than the other three groups overall, $\beta = 0.16$, $SE = 0.05$, $z = 3.52$, $p < .001$. There was a significant effect of stimulus quantity (short vs. long), $\beta = -0.48$, $SE = 0.05$, $z = -10.06$, $p < .001$, and the quantity effect was significantly smaller for Japanese speakers than the other group, $\beta = -0.03$, $SE = 0.01$, $z = -2.25$, $p = .024$. In addition, the quantity effect was significantly larger for English group than for Cantonese group, $\beta = -0.09$, $SE = 0.03$, $z = -3.30$, $p < .001$.

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. Cantonese listeners performed worse than English listeners in discrimination but not in identification.

That Japanese listeners did not do as well on identifying long sounds may be attributed to the fact that in this study we used synthetic stimuli, in which non-durational cues to quantity were held controlled. In addition to duration, native Japanese speakers also rely on pitch movement as a quantity cue [7] when listening to natural speech.

What is puzzling is why the French listeners performed much better than expected, despite the fact that their L1 is often deemed 'quantity-insensitive' [4]. It is unclear what their good performance in the present study can be attributed to. Further investigation is needed.

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 quantity contrasts (i.e. above chance accuracy). However, it is unclear why they performed less well than English listeners, to whom duration is only one of multiple acoustic cues to vowel quantity. Their performance in identification, however, was not significantly different.

Although we have selected to consider multiple L1 backgrounds and compared listeners' perception accuracy in non-native word stimuli, we found that only Japanese listeners unambiguously outperformed the others. 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.
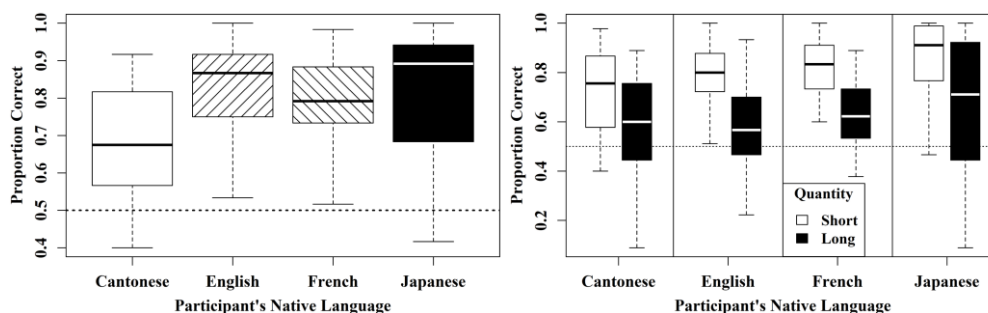

Figure 1. Discrimination (left) and identification (right) accuracy of Japanese vowel and consonant quantity contrasts by different L1 groups. The horizontal dashed lines represent chance levels.

References
[1] J. E. Flege and R. F. Port, "Cross-language phonetic interference: Arabic to English," *Lang. Speech*, vol. 24, no. 2, pp. 125–146, 1981.
[2] Y. Hirata, "Effects of speaking rate on the vowel length distinction in Japanese," *J. Phon.*, vol. 32, no. 4, pp. 565–589, 2004.
[3] A. S. House, "On vowel duration in English," *J. Acoust. Soc. Am.*, vol. 33, no. 9, pp. 1174–1178, 1961.
[4] 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.
[5] P. Birkholz, "Modeling consonant-vowel coarticulation for articulatory speech synthesis," *PLoS One*, vol. 8, no. e60603, pp. 1–17, 2013.
[6] 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.
[7] I. Takiguchi, H. Takeyasu, and M. Giriko, "Effects of a dynamic F0 on the perceived vowel duration in Japanese," 2010.
[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.