# The role of vowel duration cue in L1: Effects on L2 learners' identification of phonological vowel length in Japanese

Izumi Takiguchi

Bunkyo Gakuin University
itakiguchi@bgu.ac.jp

## ABSTRACT

This study investigated the effects of learners' L1 on their identification of Japanese phonological vowel length at three speaking rates. Native listeners of Japanese (NJ) and learners of Japanese with L1 backgrounds in Finnish (NFin), American English (NE), Russian (NR), and French (NFr) participated in the study. The results showed that the proportion of "long" responses increased as a function of vowel duration for all groups. Meanwhile, only NJ and NFin shifted the category boundary location according to speaking rate, which occurred at a shorter duration for a faster speaking rate. In addition, NFr's boundary width was significantly greater than NJ's and NFin's. These results suggest that L2 learners can access vowel duration as a cue regardless of their L1, but only those whose L1 uses vowel duration as a cue for phonological vowel length can shift the category boundary appropriately and identify L2 vowel length sharply as NJ do.

**Keywords**: L2 speech perception, phonological vowel length, Japanese, L1 influence, duration cue

## 1. INTRODUCTION

Languages differ in what phonological contrasts they have and how they use phonetic features as cues for the contrasts. Second language (L2) learners use phonetic features as cues differently than native speakers (NS) in the perception of L2 contrasts (e.g., [3, 5, 7, 10, 16]). Since phonetic features are utilized as cues for phonological contrasts differently across languages, learning an L2 might require learners to use phonetic features that are less important as cues for phonological contrasts in the L1 but are important in the L2.

Japanese has phonological vowel length. Word meanings can differ by a difference in vowel length alone, such as /kado/ (corner) with a phonemic short vowel versus /kado:/ (Japanese flower arrangement) with a phonemic long vowel. Vowel duration is the primary cue for the contrasting length categories [4].

Much research has pointed out that L2 learners often have difficulty identifying vowel length correctly. It has been shown that learners are less accurate in the identification compared to native listeners of Japanese (NJ) (e.g., [11, 12]) especially in word-final position (e.g., [9, 12]). NJ perceive vowel length categorically, but learners might not [4, 6, 13, 14, 15, 17]. The difference in speaking rates affects learners' identification accuracy, but it does not significantly affect NJ's [11, 12, 17]. NJ shift the category boundary as a function of speaking rate, but learners might not [6, 13, 17]. These differences in performance may relate to the role of vowel duration as a cue for phonological contrasts in L2 learners' languages. However, no previous studies on L2 learners' perception of Japanese vowel length seem to have examined this yet. Thus, the current study explores the role of vowel duration as a cue for L1 phonological contrasts and how this affects L2 learners' use of vowel duration in the identification of Japanese phonological vowel length.

A comprehensive approach for such an investigation includes collecting and comparing data from L2 learners of Japanese whose L1 differs in terms of vowel duration cues. Hence, based on the role of vowel duration in their language, native listeners of four languages, Finnish (NFin), American English (NE), Russian (NR), and French (NFr) were selected. Finnish has vowel length distinctions for which vowel duration is the primary cue. American English, which has no such contrast, may use vowel duration as a cue for tense versus lax vowel contrasts, for a postvocalic voicing contrast, and for lexical stress. Russian also does not have phonological vowel length, but exploits vowel duration as a cue for lexical stress. French does not have phonological vowel length, nor is vowel duration utilized to differentiate any words. English and Russian, but not French, use vowel duration as one of the possible cues to differentiate words even though they have no phonological vowel length.

The Desensitization Hypothesis [2] and the Feature Hypothesis [8] both consider the role of vowel duration as a cue in the L1 and its use in the L2. The former claims that vowel duration in the L2 is easy to access and that listeners use it to differentiate L2 vowel contrasts regardless of its role in their L1. The latter argues that prominence of L1 phonetic features determines how well listeners can use those features as cues and in turn, how successfully they can acquire L2 contrasts based on the features. The two hypotheses appear to predict

different results; however, their formulation is based on different types of studies and they might not conflict with each other in that they look at a different aspect of processing vowel duration as an L2 cue. Thus, we hypothesize that L2 learners can access L2 vowel duration irrespective of their L1, but their L1 experience with vowel duration affects how successfully they can use it in the L2.

## 2. METHODOLOGY

### 2.1. Participants

There were 50 participants in total: 8 NJ, 8 NFin, 10 NR, 12 NE, and 12 NFr. No participants reported any hearing or speaking disorder. The 8 NJ were all born and grew up in the greater Tokyo area (Tokyo, Chiba, or Saitama). Their ages ranged from 22 to 48 with the mean being 31.8 (SD=10.6). The 8 NFin were all born and grew up in Finland and their dialects all have phonological vowel length contrasts. Their ages ranged from 22 to 38 with the mean being 28 (SD=5.8). The 12 NE were all born and grew up in the US and their mean age was 20.7 (range: 18-25, SD=1.7). The 10 NR were all born and grew up in Russia and their ages ranged from 21 to 28 with the mean being 24.3 (SD=2.2). The 12 NFr were all born and grew up in France, and spoke dialects which have no distinction in vowel length. Their mean age was 23.3 (range: 19-31, SD=3.6).

### 2.2. Stimuli

45 stimuli were created from a token of a nonsense word /nono:/ produced in isolation with an HHH accent pattern. A female native speaker of Tokyo Japanese in her twenties pronounced /nono:/ in an HHH accent pattern 20 times and in a LHH pattern 12 times. The words were recorded using a linear PCM recorder (SONY PCM-D50) at a 44.1 kHz sampling rate and 16-bit quantization with an electric condenser microphone (SONY ECM-MS957) in a sound-attenuated room.

**Table 1**: Segment durations and F0 values of the original token.

| segment | | n | o | n | o: |
|---|---|---|---|---|---|
| duration (ms) | | 26.4 | 96.3 | 70.6 | 339.6 |
| F0 (Hz) | onset | 235.6 | 226.4 | 226.4 | 227.5 |
| | offset | 226.4 | 226.4 | 227.5 | 209.5 |

The duration of each segment for the 20 tokens with an HHH pattern was measured using Praat [1]. One token, which had a duration similar to the mean, was selected as the original token. Table 1 summarizes acoustic characteristics of the token.

The accent pattern, V2 duration, and the duration of C1V1C2 of the original token were manipulated using the PSOLA function in Praat. First, noise before C1 and after V2 was cut off and 20 ms-silence was inserted. Next, the accent pattern, LHH was manipulated by removing all pitch points of the original and setting new pitch points as the F0 values shown in Table 2, which were the mean values of the 12 tokens produced in the LHH pattern. Then V2 duration was shortened or lengthened from 40 ms to 390 ms in 25 ms steps (15 steps) by setting duration points in relative forms from 0.12 (40 ms) to 1.15 (390 ms). Finally, the duration of C1V1C2, was shortened to 135 ms or lengthened to 270 ms by setting the relative duration (0.7 or 1.4). The three different durational patterns including the original 193 ms (1.0) simulated changes in speaking rate.

**Table 2**: Manipulation of the LHH pattern: Loci of pitch points in time (sec) and F0 values (Hz).

| | C1 onset | V1 onset | C2 onset | V2 onset | max | L% |
|---|---|---|---|---|---|---|
| loci | 0.019962 | 0.046315 | 0.142616 | 0.213236 | 0.316174 | 0.552823 |
| F0 | 206.8 | 182.1 | 180.8 | 210.9 | 220.5 | 209.7 |

### 2.3. Procedure

Each stimulus was presented to participants in isolation 6 times in a random order. Thus, the test section consisted of 270 trials (15 stimuli x 3 CVC durations x 6 times). They were divided into 15 blocks (18 stimuli per block). The participants did a practice session, which contained 12 trials (2 endpoints x 3 CVC durations x twice) prior to the test session. Participants were asked to identify whether the word was /nono/ or /nono:/ and to give a response by pressing a specified key, which was labeled “nono” or “nono:” in Japanese *katakana*, on a PC. It took approximately 10 - 17 minutes to complete.

## 3. RESULTS

Figure 1 shows the logistic functions for each L1 group for each CVC durational pattern. The lines and the shapes represent the fitted logistic curves and the observed proportion of “long” responses, respectively. The proportion of “long” responses increases as a function of V2 duration, giving an s-shaped curve for all groups. On the other hand, for NJ and NFin, the three curves are separately located, which indicates that the identification as long occurs earlier as the CVC duration is shorter.

**Figure 1**: Proportion of "long" responses as a function of vowel duration for the three durational patterns. Top to bottom: NJ, NFin, NE, NR, & NFr.

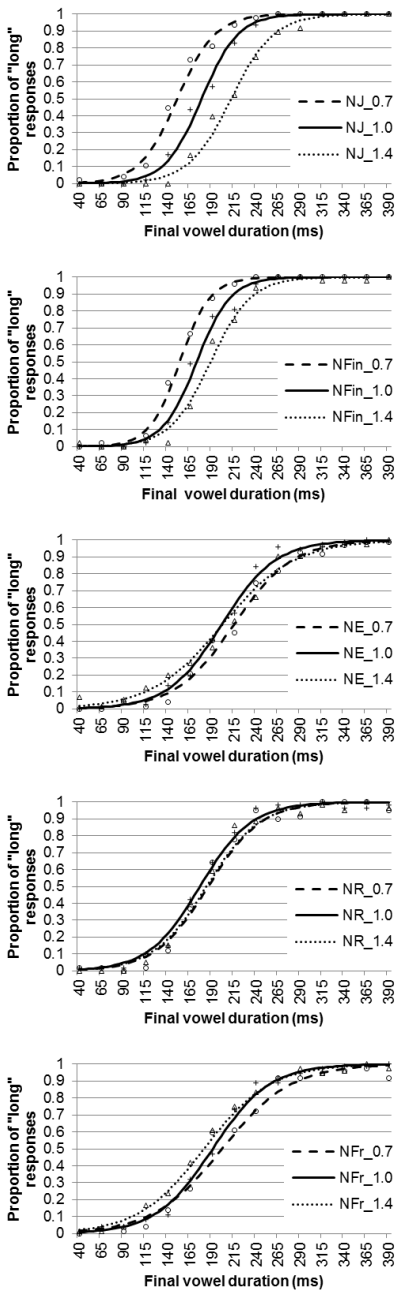



Figure 2 shows mean estimates of the category boundary. A two-way repeated measures ANOVA with L1 as a between-subjects factor and CVC durational pattern as a within-subjects factor showed no significant main effects of L1 [$F(4, 45)=2.2$, $p=0.087$, *n.s.*] and CVC durational pattern [$F(1.58, 71.2)=3.4$, $p=0.050$, *n.s.*]. However, an interaction between L1 and CVC durational pattern was significant [$F(6.33, 71.2)=5.8$, $p<.001$]. Thus, a one-way repeated measures ANOVA was run for each L1 group. ANOVAs for NE, NR, and NFr showed no main effect of CVC durational pattern [NE: $F(2, 22)=1.2$, $p=0.320$, *n.s.*; NR: $F(2,18)=0.9$, $p=0.418$, *n.s.*; NFr: $F(1.31, 14.37)=1.7$, $p=0.221$, *n.s.*], but ANOVAs for NJ and NFin showed a main effect [NJ: $F(2, 14)=28.7$, $p<.001$; NFin: $F(2, 14)=18.1$, $p<.001$]. A post hoc test for NJ revealed that the category boundary for 1.4 occurred at a significantly longer duration than those for 0.7 ($p<.01$) and 1.0 ($p<.05$). Also, the boundary for 1.0 occurred at a significantly longer duration than that for 0.7 ($p<.01$). That is, the boundary value was 0.7 (151.1 ms) < 1.0 (178.2 ms) < 1.4 (211.5 ms) for NJ. Quantitatively, NFin's category boundary was also 0.7 (153.9 ms) < 1.0 (172.1 ms) < 1.4 (188.2 ms). A post hoc test for NFin showed significant differences between 0.7 and 1.0 ($p<.05$), and 0.7 and 1.4 ($p<.01$), and a marginal difference between 1.0 and 1.4 ($p<0.1$).

**Figure 2**: Mean estimates of category boundary.

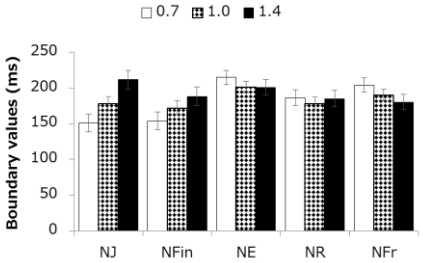
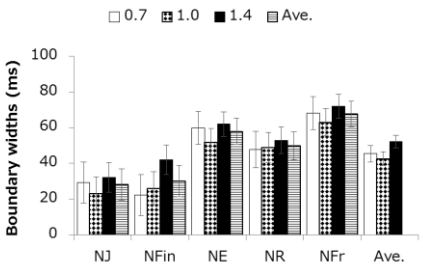



Figure 3 shows mean estimates of the boundary width. A two-way repeated measures ANOVA on boundary width with L1 as a between-subjects factor and CVC durational pattern as a within-subjects factor showed significant main effects of L1 [$F(4, 45)=4.6$, $p<.01$] and CVC durational pattern [$F(1.7, 75.7)=5.7$, $p<.01$]. An interaction between L1 and CVC durational pattern was not significant [$F(6.7, 75.7)=0.69$, $p=0.670$, *n.s.*]. A Tukey (HSD) post hoc test for the main effect of L1 revealed that NFr's boundary width (67.8 ms) was significantly greater than NJ's (28.1 ms) ($p<.01$) and NFin's (30.1 ms) ($p<.05$). NE's boundary width (58.0 ms) was also marginally greater than NJ's (28.1 ms) ($p<0.1$). There were no other significant differences between any pair. In addition, the post hoc test for the main effect of CVC durational pattern showed significant difference between 1.0 (42.6 ms) and 1.4 (52.2 ms) (1.0 < 1.4, $p<.01$).

**Figure 3**: Mean estimates of boundary width.



## 4. DISCUSSION

The proportion of "long" responses increased as a function of vowel duration, giving an s-shaped curve. This indicates that all groups identified vowels as short and long when vowel duration was sufficiently short and long, respectively. That is, all L2 learners are sensitive to vowel duration as a cue. The role of vowel duration in the L1 does not relate to the ability to access vowel duration for length identification in the L2, as suggested by the Desensitization Hypothesis [2].

Only NFin among the L2 learners shifted their category boundary as a function of speaking rate in the same fashion as NJ. The current results suggest that NJ and L2 learners with experience using vowel duration for phonological vowel length in the L1 refer to surrounding segment duration, detect durational differences, and shift their boundary location. NE and NR, whose L1 uses vowel duration for segmental contrasts and/or lexical stress, and NFr, whose L1 does not utilize vowel duration to differentiate any words, behaved similarly. Thus, we can conclude that only L1 experience with vowel duration for phonological vowel length provides L2 learners with the ability to shift the category boundary appropriately to adapt to speaking rate, as suggested by the Feature Hypothesis [8].

The inability to shift the category boundary shown by the three groups is in line with the report by [13] and [17]. On the other hand, [6] states that American English listeners can learn to shift the boundary location according to the temporal context with a 5-day perceptual training program. This implies that L2 learners without L1 experience using vowel duration for length distinctions need some specific training to be able to shift the boundary location according to speaking rate. By contrast, L2 learners whose L1 has a vowel length contrast have already developed such an ability in their L1, and can use this ability in the L2.

NFr's boundary width was significantly larger than those of NJ and NFin. This result suggests that the role of vowel duration in the L1 relates to the sharpness of length identification in the L2, as suggested by the Feature Hypothesis [8]. First, L2 learners with L1 experience using vowel duration for vowel length distinction identify vowel length sharply, at the level of NJ. Next, L2 learners without L1 experience using vowel duration to differentiate words identify vowel length less sharply compared to NJ and L2 learners whose L1 has phonological vowel length. Finally, L2 learners whose L1 uses vowel duration to differentiate words but not for phonological vowel length do something in-between. On the other hand, NE's boundary width was marginally greater than that of NJ. That is, NE's length identification is less similar to NJ's than NR's in terms of the sharpness. This suggests that factors other than vowel duration in the L1 also affect the sharpness of length identification in the L2.

The boundary width was affected by speaking rate and that of 1.4 was significantly larger than that of 1.0. This indicates that listeners were less sharp at identifying vowel length when the stimuli were spoken at a slow speaking rate than at a normal speaking rate. It is possible that stimuli for which the CVC duration was lengthened, due to the manipulation, might have sounded unnatural compared to stimuli with original duration. However, the identification sharpness did not differ between the stimuli with the original duration and those with the shortened duration. Consequently, unnaturalness due to the manipulation does not fully explain the result. On the other hand, [17] also reported less sharp length identification at a slower speaking rate. Thus, there might be some mechanism for relating length identification and speaking rate.

Here, it is interesting that not only NJ and NFin, who showed a boundary shift as a function of speaking rate, but also the others, who did not show such a shift, demonstrated the effect of speaking rate on the boundary width. This indicates that the learners hear surrounding segment durations and are sensitive to durational changes. Thus, it seems that what is challenging for L2 learners who lack L1 experience using vowel duration for phonological vowel length is being able to shift the boundary in relation to speaking rate. Their sensitivity to durational changes, on the other hand, further supports the idea that L2 learners can access vowel duration as a cue regardless of its role in their L1.

## 5. CONCLUSION

This study investigated how vowel duration in learners' L1 affects its use in the identification of vowel length contrasts in the L2. The major findings are as follows. First, L2 learners, regardless of their L1, can access vowel duration as an L2 cue. Second, experience using vowel duration as a cue for phonological vowel length in the L1 allows L2 learners to shift the category boundary appropriately as a function of speaking rate and to identify vowel length sharply in the L2. To conclude, the role of vowel duration in the L1 does not affect the ability to access L2 vowel duration, as suggested by the Desensitization Hypothesis. Instead, it relates to the ability to shift the category boundary in relation to speaking rate and to identification sharpness, i.e., the ability to use vowel duration successfully, as suggested by the Feature Hypothesis.

## 6. REFERENCES


[1] Boersma, P., Weenink, D. 1992-2012. Praat: doing phonetics by computer http://www.praat.org/

[2] Bohn, O. S. 1995. Cross-language perception in adults: First language transfer doesn't tell it all. In: Strange, W. (ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* Timonium, Maryland: York Press, 379-410.

[3] Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34, 372-387.

[4] Fujisaki, H., Nakamura, K., Imoto, T. 1975. Auditory perception of duration of speech and non-speech stimuli. In: Fant, G. and Tatham, M. A. A. (eds.), *Auditory Analysis and Perception Speech*. London: Academic, 197-219.

[5] Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.

[6] Kato, H., Tajima, K., Rothwell, A., Akahane-Yamada, R., Munhall, K. 2004. Perception of phonemic length contrasts in Japanese with or without a carrier sentence by native and non-native listeners. *Proc. of the 18th International Congress on Acoustics*, Kyoto, Japan, I-609-I-612.

[7] Lippus, P., Pajusalu, K. Allik, J. 2009. The tonal component of Estonian quantity in native and non-native perception. *Journal of Phonetics* 37, 388-396.

[8] McAllister, R., Flege, J. E. Piske, T. 2002. The influence of L2 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30, 229-258.

[9] Minagawa-Kawai, Y., Maekawa, K., Kiritani, S. 2002. Nihongo gakushuusha no chou/tanboin no doutei ni okeru picchigata to onsetsuichi no kouka [Effects of pitch accent and syllable position in identifying Japanese long and short vowels: Comparison of English and Korean speakers]. *Journal of the Phonetic Society of Japan* 6, 88-97.

[10] Munro, M. J. Bohn, O. S. 2007. The study of second language speech: A brief overview. In: Bohn, O. S., Munro, M. J. (eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company, 3-11.

[11] Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R., Sagisaka, Y. 2013. Non-native perception and learning of the phonemic length contrast in spoken Japanese: training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics* 22, 373-398.

[12] Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., Munhall, K. G. 2008. Training English listeners to perceive phonemic length contrasts in Japanese. *J. Acoust. Soc. Am.* 123, 397-413.

[13] Toda, T. 2003. *Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese*. Lanham, Maryland: University Press of America.

[14] Uchida, T. 1993. Chuugokujin nihongo gakushuusha ni okeru chouon to sokuon no choukakuteki ninchi no tokuchou [Characteristics of auditory cognition of long vowels and double consonants for Chinese students in learning Japanese language]. *Japanese Journal of Educational Psychology* 41, 414-423.

[15] Uchida, T. 1998. Nihongo tokushuhaku no shinritekina ninchikatei kara toraeta onsetsu to haku: Teijouteki onsei kukan no jizokujikan ni taisuru kategorii teki chikaku [Categorical perception of relatively steady-static speech sound duration in Japanese moraic phonemes]. *Journal of the Phonetic Society of Japan* 2, 71-86.

[16] Wang, 2008. L2 stress perception: The reliance on different acoustic cues. *Proc. of the 4th International Conference on Speech Prosody*, Campinas, Brazil.

[17] Wilson, A., Kato, H., Tajima, K., Akahane-Yamada, R. 2005. Native and non-native perception of phonemic length contrasts in Japanese: Effects of speaking rate and presentation context. *Proc. of 2005 Spring Meeting the Acoustical Society of Japan*, 2-2-5, 331-332.