

Absence of Developmental Change in Epenthetic Vowel Duration in Japanese Speakers' English

Takayuki Konishi, Kakeru Yazawa, Mariko Kondo

Abstract—This study examines developmental change in the production of epenthetic vowels by Japanese learners of English in relation to acquisition of L2 English speech rhythm. Seventy-two Japanese learners of English in the *J-AESOP* corpus were divided into lower- and higher-level learners according to their proficiency score and the frequency of vowel epenthesis. Three learners were excluded because no vowel epenthesis was observed in their utterances. The analysis of their read English speech data showed no statistical difference between lower- and higher-level learners, implying the absence of any developmental change in durations of epenthetic vowels. This result, together with the findings of previous studies, will be discussed in relation to the transfer of L1 phonology and manifestation of L2 English rhythm.

Keywords—Vowel epenthesis, Japanese learners of English, L2 speech corpus, speech rhythm.

I. INTRODUCTION

JAPANESE has a relatively simple syllable structure (mostly $J/(C)V/$) compared to other languages. This affects first language (L1) Japanese speakers' production and perception of second language (L2) speech. L1 Japanese speakers epenthesize vowels in consonant clusters and after syllable-final consonants, except for the moraic nasal, when they speak languages with more complex syllable structure such as English and French; for example, *street* /stri:t/ becomes /sutori:to/ (e.g. [4], [12], [13], [17]). Vowel epenthesis has been attributed to the perception of "illusory" vowels, which is caused by the transfer of L1 phonotactics; L1 Japanese speakers tend to hear /ebzo/ as /ebuzo/ while French speakers do not [4]. This view has been supported by acoustic and articulatory studies (e.g. [5], [16], [19]) showing that the quality of epenthetic vowels in Japanese learners' English is influenced by Japanese phonotactic constraints. The type of epenthesized vowel depends on the type of preceding consonant similar to the patterns observed in Japanese loanword phonology: the patterns are shown below.

- (1) $\emptyset \rightarrow i / tʃ, dʒ _$
- (2) $\emptyset \rightarrow o / t, d _$
- (3) $\emptyset \rightarrow u / \text{elsewhere}$

On the other hand, the transfer of L1 syllable structure is not the only cause of vowel epenthesis in non-native speech. For example, Davidson and her colleagues (e.g. [2], [3]) argue that the production of epenthetic vowels is better explained by

taking into account the "fine-grained phonetic details" of the consonant sequences. As discussed in [6], if a non-native consonant cluster is produced with insufficiently overlapping configuration ("mistiming") due to the articulatory difficulty, a transitional vowel-like structure appears within the cluster. Other studies have also reported that the frequency of vowel epenthesis by Japanese learners of English is subject to adjacent phonetic environments such as voicing, providing additional evidence for articulatory difficulties as a cause of vowel epenthesis (e.g. [12], [17]).

Despite the abundance of research on vowel epenthesis in L1 Japanese speakers' English, few studies have discussed its relationship with the learners' L2 proficiency. A previous study [19] demonstrated that, although the frequency of vowel epenthesis decreases in Japanese advanced learners' English, the quality of the epenthetic vowels is not affected by the learners' proficiency.

Also, the majority of previous studies discussed vowel epenthesis with a focus on articulatory difficulties or transfer of L1 phonotactics; not much attention has been paid to vowel epenthesis in relation to acquisition of L2 speech rhythm, although the epenthesis will certainly impede the proper manifestation of L2 English rhythm by changing the syllable structures.

Previous studies on L2 prosodic acquisition [8], [9], [11] found that advanced Japanese learners of English can accurately manifest durations of English stress contrast, i.e. unstressed vowels are produced with shorter duration, but the lack of spectral contrasts between stressed and unstressed vowels seems rather persistent.

Based on these, we can formulate a hypothesis that if there is an effect of L2 English speech rhythm on vowel epenthesis, the duration of epenthetic vowels will be shorter for Japanese advanced learners of English.

In an earlier study [10], we conducted an acoustic analysis to investigate whether duration of epenthetic vowels is shorter for Japanese advanced learners of English. However, the analysis did not take into account each learner's speech rate, which is also associated with the proficiency level. Therefore, the current study aims to apply more accurate statistical methods to test whether there is a developmental change in the duration of epenthetic vowels produced by Japanese learners of English.

II. METHOD

A. Data

The current study used 72 annotated English read speeches of *The North Wind and the Sun* by Japanese learners of English. The data has been extracted from the *J-AESOP* corpus that was

Takayuki Konishi and Kakeru Yazawa are PhD students in linguistics at Waseda University, 1-6-1, Nishi-Waseda, Shinjuku-ku, Tokyo, 169-8050, Japan (e-mail: tkonishi@aoni.waseda.jp; kyazawa@gmail.com).

Mariko Kondo is a professor in phonetics and speech communication at Waseda University, Japan (e-mail: mkondo@waseda.jp).

constructed by members of the *Asian English Speech cOrpus Project (AESOP)* in Japan [14], [18].

Each subject in the dataset was assigned a proficiency score, which was the mean score evaluated by 8 English teachers (4 native English speakers and 4 native Japanese speakers) on a 9-point scale with 0.5 increments from 1 (very poor) to 5 (very good, nativelike).

B. Annotation

The data was first automatically annotated using the Hidden Markov Model Toolkit (HTK) [7] with a pronunciation dictionary based on the TIMIT corpus [1]. After the automatic alignment, epenthetic vowels were manually annotated together with modification of segmental boundaries. Four criteria [17] were used to identify epenthetic vowels (TABLE I); if the four conditions were met, then we identified it as an epenthetic vowel.

Table II shows examples of words that were identified to have epenthetic vowels.

TABLE I
CRITERIA FOR JUDGING VOWEL EPENTHESIS [17]

| Criteria |
|--|
| A. Clear vowel-like formant structure (especially F1) on the spectrogram |
| B. Periodic waveform |
| C. At least two pitch periods |
| D. Evidence as vowels (if the epenthetic vowel is adjacent to a consonant with formant structure). |

TABLE II
EXAMPLES OF WORDS THAT WERE IDENTIFIED TO HAVE EPENTHETIC VOWELS ([V] REPRESENTS AN EPENTHETIC VOWEL)

| Word | Pronunciation | Environment |
|-----------------|----------------|-------------------|
| <i>along</i> | /aɒŋ[V]/ | syllable-final |
| <i>of</i> | /əv[V]/ | syllable-final |
| <i>blew</i> | /b[V]lu/ | consonant cluster |
| <i>stronger</i> | /s[V]t[r]ɒŋgə/ | consonant cluster |

C. Analysis

In the statistical analysis the subjects were divided into lower-level learners (n = 37) and higher-level learners (n = 32) using the median proficiency score of 2.8125 (Fig. 1). Three subjects (whose scores were 3.50, 3.69 and 4.88) were excluded because no epenthesis was observed for them.

The mean durations of epenthetic vowels in the two groups were statistically compared. Since higher-level learners are assumed to speak faster, the durations of epenthetic vowels in both groups were normalized using (1), where d_{ep} is the duration of each epenthetic vowel, d_{wd} is the duration of the word containing the epenthetic vowel, n_{sg} is the number of segments in the word, and d_{nor} is the normalized duration.

$$d_{nor} = \frac{d_{ep}}{d_{wd}/n_{sg}} \quad (1)$$

The distributions of the normalized durations of the epenthetic vowels for each learner group are shown in Fig. 2. However, the normalized durations of all the lower- and

higher-level learners' epenthetic vowels were not normally distributed (Sapiro-Wilk normality test [$ps < .01$] [15]). Therefore, the data were normalized by calculating the mean duration for each subject (Fig. 3), neither lower- nor higher-learner distributions were statistically different from normal distribution (Sapiro-Wilk normality test [$ps > .2$] [15]). Then, performing a Welch's Two Sample t-test on this normalized data did not show any significant difference between the two learner groups ($t(58.911) = -1.4793, p > .1$ [15]). The result is summarized in Fig. 4.

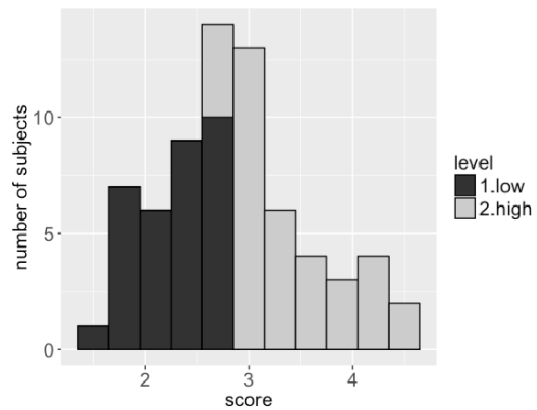


Fig. 1 Distribution of the proficiency scores (black: lower-level learners / grey: higher-level learners)

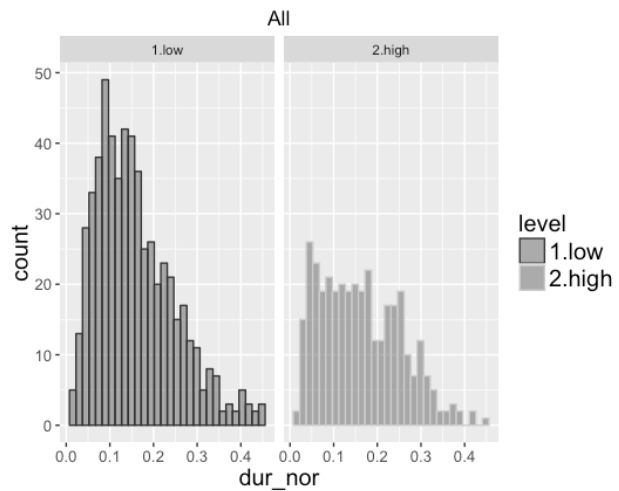


Fig. 2 Normalized durations of epenthetic vowels left: lower-level learners / right: higher-level learners

Finally, the statistical analysis was conducted again, this time of two learner groups categorized according to the frequency of vowel epenthesis. Once again, a Welch's Two sample t-test did not show any statistical difference between the duration of epenthetic vowels of the lower-level group (n = 35, mean frequency of epenthesis = 21.4) and higher-level group (n = 34, mean frequency of epenthesis = 5.2) ($t(49.153) = -0.59147, p > .5$ [15]).

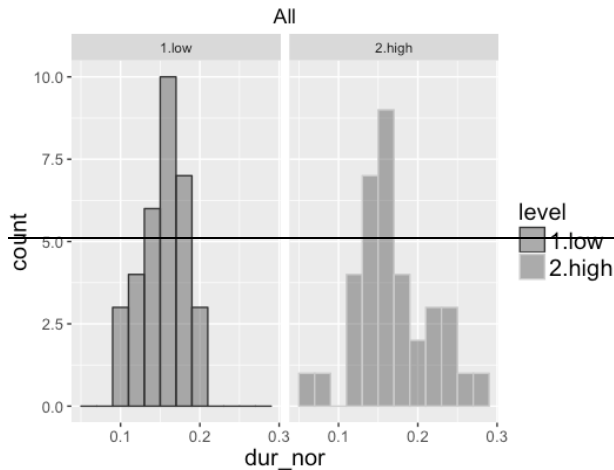


Fig. 3 Means of normalized durations of epenthetic vowels, calculated per subject (left: lower-level learners / right: higher-level learners)

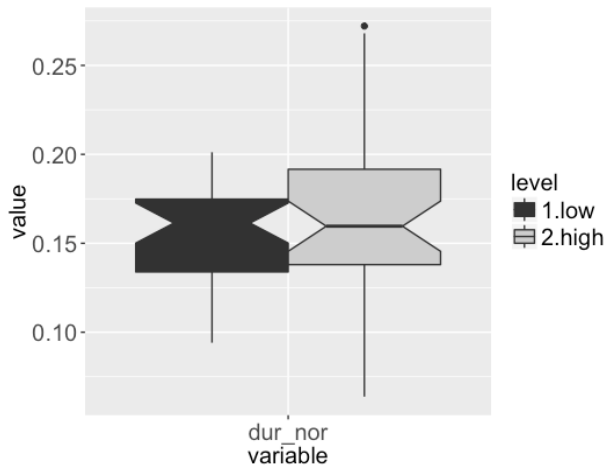


Fig. 4 Mean of normalized durations for each learner group (left: lower-level learners / right: higher-level learners)

III. DISCUSSION

The absence of any change in the durations of epenthetic vowels in production is in contrast to the effect of phonetic details on the perception of epenthetic vowels by L2 English speakers [2], [3], [16]. Although L2 English speakers are sensitive to the fine phonetic details of speech sound, our results show that the manipulation of such phonetic details in production is difficult even for advanced learners.

The results of our current analysis are summarized together with the results by [17], [6], [7], [9] in Table III. Based on these results, we propose that Japanese learners' ability/inability to control each L2 phonetic feature can be explained by the transfer of L1 phonology or articulatory difficulty of the L2 sound.

The absence of any developmental changes in the duration of epenthetic vowels suggests that, at least in Japanese learners' English, the occurrence of vowel epenthesis is a result of L1 transfer and therefore it is not incorporated into L2 rhythm. In

other words, the acquisition of L2 English rhythmic structure does not seem to affect the production of epenthetic vowels, whereas it does appear to be the case for the duration of L2 English unstressed vowels. Therefore, our latest finding, together with earlier results [19], indicate that although Japanese learners of English epenthesis fewer vowels as they overcome the articulatory difficulties of producing L2 sound sequences, even advanced learners epenthesize *full* vowels in terms of both quality and duration, suggesting a persistent transfer of L1 phonotactics. In other words, regardless of the learner's proficiency, the epenthesis is either fully present or completely absent.

TABLE III
DEVELOPMENTAL CHANGES IN PHONETIC FEATURES OF ENGLISH REDUCED VOWELS AND EPENTHETIC VOWELS PRODUCED BY JAPANESE LEARNERS OF ENGLISH

| | Phonetic feature | Developmental change |
|-------------------|------------------|----------------------|
| Unstressed vowels | Quality | No |
| | Duration | Yes |
| Epenthetic vowels | Quality | No |
| | Duration | No |

In [9], we argued that L1 interference on L2 prosody is easier to overcome than L1 interference on L2 segments. Now, the result of the current study further suggests that persistent L1 interference exists both on L2 segments and L2 syllable structure.

IV. CONCLUSION

The result of the above analysis suggests that vowel epenthesis is not a phenomenon of L2 speech rhythm. The epenthesis of Japanese learners of English was either fully present or completely absent; the developmental change is on epenthetic frequency rather than the realization of epenthetic vowels.

In addition, taken together with previous studies, this latest study shows that, in addition to L2 segments, L2 syllable structure is also more difficult to acquire than L2 prosody.

ACKNOWLEDGMENT

This research was supported by Waseda University Grant for Special Research Projects (#2016S-144).

REFERENCES

- [1] The DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus (TIMIT). http://www.ldc.upenn/readme_files/timit.readme.html
- [2] Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 37(1), 270.
- [3] Davidson, L., Martin, S., & Wilson, C. (2015). Stabilizing the production of nonnative consonant clusters with acoustic variability. *The Journal of the Acoustical Society of America*, 137(2), 856-872.
- [4] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of experimental psychology: human perception and performance*, 25(6), 1568.
- [5] Funatsu, S., Imaizumi, S., Fujimoto, M., Hashizume, A., & Kurisu, K. (2008). Do Japanese speakers perceive nonexistent vowels in non-native

- consonant clusters? *Journal of the Acoustical Society of America*, 123(5), 3072. doi: 10.1121/1.2932847
- [6] Hall, N. (2003). *Gestures and segments: Vowel intrusion as overlap* (Unpublished doctoral dissertation). University of Massachusetts.
- [7] Hidden Markov Model Toolkit (HTK). htk.eng.cam.ac.uk
- [8] Kondo, M. (2009). Is Acquisition of L2 Phonemes Difficult? Production of English Stress by Japanese Speakers. In Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009), 105-112. Somerville, MA.
- [9] Konishi, T., Kondo, M. (2015) Developmental Change in English Stress Manifestation by Japanese Speakers *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, Scotland
- [10] Konishi, T., Yazawa, K. & Kondo, M. (2016) Japanese Advanced Learners of English Produce Shorter Epenthetic Vowels. *Proceedings of The 30th General Meeting of the Phonetic Society of Japan*. pp. 70-75.
- [11] Lee, B., Guion, S. G., 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(03), 487-513.
- [12] Masuda, H. & Arai, T. (2010). Processing of consonant clusters by Japanese native speakers: Influence of English learning backgrounds. *Acoustical Science and Technology* 31(5). 320- 327.
- [13] Mazuka, R., Cao, Y., Dupoux, E. & Christophe, A. (2011). The development of phonological illusion: A cross-linguistic study with Japanese and French infants. *Developmental Science* 14(4), pp693-699.
- [14] Meng, H., Tseng, C. Y., Kondo, M., Harrison, A. M., & Visceglia, T. (2009). Studying L2 Suprasegmental Features in Asian Englishes: A Position Paper. In *Interspeech*. pp. 1715-1718.
- [15] R Version 3.3.0. <https://cran.r-project.org>
- [16] Shibuya, Y., & Erickson, D. (2010). Consonant cluster production in Japanese learners of English. In *Proceedings of the 10th annual conference of the International Speech Communication Association*. Tokyo: Waseda University.
- [17] Tajima, K., Erickson, D., & Nagao, K. (2000). Factors affecting native Japanese speakers' production of intrusive (epenthetic) vowels in English words. In Proceedings of the 6th International Conference on Spoken Language Processing (pp. 558–561). Beijing.
- [18] Visceglia, T., Tseng, C. Y., Kondo, M., Meng, H., & Sagisaka, Y. (2009). Phonetic aspects of content design in AESOP (Asian English Speech corpus Project). In *Speech Database and Assessments, 2009 Oriental COCOSDA International Conference on*. pp. 60-65. IEEE.
- [19] Yazawa, K., Konishi, T., Hanzawa, K., Short, G. & Kondo, M. (2015) Vowel Epenthesis in Japanese Speakers' L2 English *Proceedings of The 18th International Congress of Phonetic Sciences (ICPhS XVIII)*, Glasgow, Scotland