Developmental Change in English Stress Manifestation by Japanese Speakers

Takayuki Konishi[1] & Mariko Kondo[1,2,3]

GSICCS[1], SILS[2], LASS[3], Waseda Univeristy
tkonisi@aoni.waseda.jp, mkondo@waseda.jp

ABSTRACT

This study investigated the developmental change in the phonetic realization of English stress, using corpus data of Japanese English learners with varying proficiency. Previous studies demonstrated that highly proficient Japanese learners of English can produce native-like English lexical stress in terms of intensity, F0, and vowel duration, but not vowel quality. The results of our study showed that the contrasts in intensity, F0 and duration were manifested by all proficiency groups while spectral contrast was consistently absent. In addition, significant differences in vowel duration were found between native speakers and Japanese speakers of low-to-medium English proficiency level. The results imply that Japanese mora-timed rhythm is an obstacle in manifesting native-like lexical stress. These two findings suggest that it is more difficult to overcome L1 interference of segmental phonology than of suprasegmental phonology.

Keywords: SLA, English stress, L1 interference, Japanese phonology

1. INTRODUCTION

There has been a growing interest in non-native varieties of English in recent years. Studies have found that it is more important to focus on Second Language (L2) speech phenomena that are hard for interlocutors to understand [7]. Acquisition of proper prosody is very important in this regard because wrong manifestation of prosody leads to lexical-semantic, syntactic, and pragmatic misunderstandings. In addition, the latest theories of phonetics and phonology support the view that the perception of prosody precedes that of segments [10]. Accordingly, accurate manifestation of prosody is fundamental in L2 English communication.

Despite its importance, there are much fewer suprasegmental studies on L2 English compared to segmental ones. For example, acquisition of English phonemes by Japanese speakers has been extensively studied (e.g. [9]), but there are few studies on suprasegmental aspects of English spoken by Japanese speakers. Also, recently there have been an increasing number of corpus-based studies on Second Language Acquisition (SLA), but most of these studies use written corpora rather than spoken corpora.

2. ENGLISH STRESS BY JAPANESE SPEAKERS

English is a stress accent language. Existing studies show that the acoustic correlates of English lexical stress are intensity, F0, vowel duration and vowel quality, i.e. spectra. In contrast, Japanese is a pitch accent language; the accent is realized by change of F0, and it does not affect vowel duration, quality or intensity. Therefore, first language (L1) interference is expected to happen in Japanese speakers' acquisition of English lexical stress. The fundamental rhythmic structures of the two languages also differ; Japanese is a mora-timed language and English a stress-timed language.

It has been reported that Japanese learners of English with high proficiency can manifest native-like stress in terms of intensity, F0 and vowel duration [4], [5]. However, in these two studies spectral contrast between stressed and unstressed vowels was consistently absent. A significant difference was found in duration of unstressed vowels between native English speakers and Japanese English learners in one study [4], but not in the other [5]. These contradictory results may have been due to the different proficiency of the subjects; in the study showing a difference all the subjects were L2 English learners in Japan [4], whereas in the other study they were Japanese-English bilinguals in the US [5]. Therefore, in this study, we investigated the developmental change in the phonetic realization of English stress, using corpus data of Japanese English learners with varying proficiency.

3. ANALYSIS

3.1. Data

In this study, we extracted data from *J-AESOP* comprising of the data of 16 native English speakers and 71 native Japanese speakers. *J-AESOP* is part of the *AESOP* Corpus (e.g. [6], [11]). The data of one speaker was not available for the task used in the

current study due to the technical problem at the time of the recording. Therefore 70 Japanese speakers' data were analyzed for this study. The speakers' proficiency levels were evaluated by 8 English teachers on 9-point scale with 0.5 point increments, from 1 (very poor), 3 (medium) up to 5 (very good, native-like), and the scores of the 8 evaluators were averaged.

The test words in a reading task were analyzed. Some examples of the test words (underlined) in the carrier sentence are listed (a)-(c) below.

(a) I said apartment five times.
(b) I said elevator ten times.
(c) I say tomorrow ten times.

This experimental setting enabled us to study whether the subjects differentiated stressed and unstressed vowels.

Using the proficiency score as the criterion, Japanese speakers were divided into three proficiency groups, low-, medium-, and high-proficiency learners. This was done using the 24th score (low-med break) and 47th score (med-high break) from the lowest as reference points. The learners with scores lower than the low-med break (2.625) were grouped as low-proficiency learners (low-group), those with scores higher than the med-high break (3.125) as high-proficiency learners (high-group), and those in between as medium-proficiency learners (med-group). Since some of the learners had the same score, the number of learners was not the same in each group. The low-group consisted of 23 learners, the med-group had 26 learners and the high-group had 21 learners (Figure 1).

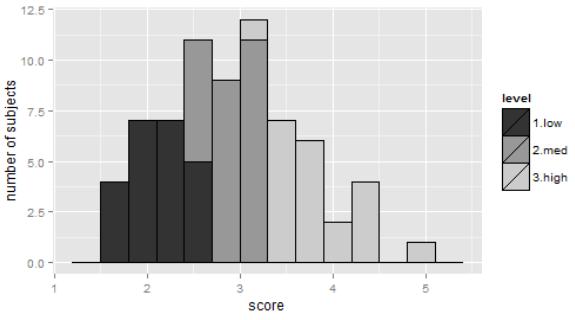

Figure 1. Distribution of the subjects' scores

The whole data set consisted of 320 utterances of native English speakers (20 test words ×16 subjects) and 1400 utterances of Japanese English learners (20 test words ×70 subjects). The recorded sounds were manually annotated so that the phonetic information of target words and their segments could be obtained. Primary and secondary stress placements were additionally annotated.

Acoustic parameters of each vowel were extracted to conduct statistical analysis. Non-canonical vowels, i.e. those different from the canonical ones in English and which are inexplicable by interlanguage phonology, were excluded from the data. Some of these substitutions seemed to be caused by the orthography of English loanwords in Japanese, which is irrelevant to the current analysis. It was sometimes the case that even for native speakers, the stress placements differed from the ones prescribed in dictionaries. For the current analysis, if the vowels were extremely different from the normal allophonic variance, they were excluded from the data. It was also sometimes the case that the low-group placed stress on vowels which are canonically realized as schwa /ə/. Those vowels were also excluded from the data as were vowels with secondary stress.

3.2. STATISTICAL DESIGN

The current study aims to test the results of the previous studies, [4], [5], using corpus data of Japanese learners of English with varying proficiency. Replicating the methods adopted in the two previous studies, separate ANOVA was conducted for each of the four parameters (intensity, F0, vowel duration and vowel spectra). Stress (+ or – stressed) was set as a within-subject variable and proficiency (low-/ medium-/ high-group/ native) as a between subject variable.

Intensity (in dB), F0 (in semitone, re = 1 Hertz), duration (in ms) and first and second formants (F1 and F2; in Hertz) of each vowel were measured. In order to test the reduction of vowels with single ANOVA, F1 and F2 were converted to Euclidean distance from the center of each speaker's vowel space using formula (1). F1 and F2 denote first and second formant values of each vowel. μ_{F1} and μ_{F2} are means of first and second formant values of the center of each vowel space. The center of each vowel space for each speaker was defined as the mean F1 value and mean F2 value of all the vowel tokens.

$$(1) \qquad \mathrm{d(F1,F2)} = \sqrt{(F1 - \mu_{F1})^2 + (F2 - \mu_{F2})^2}$$

After the conversion, all four parameters were normalized by calculating the z-score for each speaker. The equation for the conversion is shown in (2) where x represents each value, μ is the average value and σ the standard deviation.

$$(2) \qquad \mathrm{z} = \frac{x - \mu}{\sigma}$$

The conversion was especially important for the analysis of F0 and formants in order to eliminate the

effect of physiological difference between males and females.

4. RESULTS

Figure 2 shows the difference in stressed and unstressed conditions of each parameter for each speaker group.

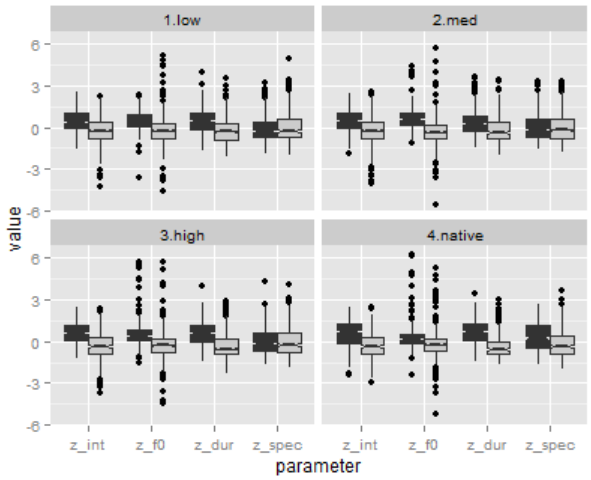

Figure 2. Each parameter in stressed (in black) and unstressed (in grey) conditions
(int: intensity f0: F0 dur: duration spec: spectra "z_" denotes the value is z-score normalized.)

For intensity, the main effect of stress and its interaction with proficiency level were both significant, $F(1, 82) = 279.744$, $p < .001$ and $F(3, 82) = 2.922$, $p < .05$, respectively. Post-hoc Tukey's HSD showed that stressed vowels had significantly greater intensity than unstressed vowels for all groups, all *p*s < .001. There was no significant difference in intensity between any two groups, all *p*s > .1.

As for F0, the main effect and interaction were again both significant, $F(1, 82) = 332.308$, $p < .001$ and $F(3, 82) = 4.181$, $p < .001$, respectively. Stressed vowels had higher F0 than unstressed vowels for all groups, *p*s < .001. A significant difference was found between native speakers and the med-group for stressed vowels, $p < .001$, with a higher F0 for the med-group (mean = 92.3 Hz) than for native speakers (mean = 88.1 Hz). We interpret this to be a hypercorrection due to L1 Japanese phonology; cf. Japanese uses F0 only to manifest its accent.

There was also a significant main effect of stress, $F(1, 82) = 274.147$, $p < .001$ on duration, together with its significant interaction with level $F(3, 82) = 7.103$, $p < .001$. Stressed vowels showed greater duration than unstressed vowels for all groups, *p*s < .001. There were also significant differences in stressed vowels between native speakers and both the low-group and the med-group, $p < .001$ and $p < .01$, respectively.

Finally, for vowel spectra, both the main effect of stress and interaction with level were again significant, $F(1, 82) = 6.375$, $p < .01$ and $F(3, 82) = 6.462$, $p < .001$. Native speakers' stressed vowels had significantly greater vowel spectra than their unstressed vowels $p < .001$, but there were no differences in vowel spectra between stressed and unstressed vowels for any of the learner groups, all *p*s >.9. These data support the conclusions of the previous studies that Japanese speakers cannot manifest native-like spectral contrast.

To sum up, these results revealed differences in each proficiency group between stressed and unstressed vowels for intensity, F0 and vowel duration. However, differences in vowel spectra were only observed for native speakers. The data indicate that intensity and F0 were native-like in all learner groups. The only parameter which showed any difference between learner groups was vowel duration, where the durational contrast between stressed and unstressed vowels for low- and med-groups was less than for the high-group which was similar to that of the native speakers.

5. DISCUSSION

For native English speakers, all four parameters showed significant differences between stressed and unstressed vowels, replicating the results in the previous studies [4], [5].

The results of the high-group were also consistent with the analyses on proficient Japanese speakers in these two studies [4], [5].

In addition, the current study also showed that the contrasts in intensity, F0 and duration were similar for all three learner groups.

Furthermore, the results of the current study can also help explain the previous results that showed a significant difference in vowel duration between native English speakers and non-bilingual learners in one study [4], but no difference between native English speakers and Japanese-English bilinguals in the other study [5]. Our current study indicates that there are developmental stages in manifesting native-like vowel durations.

Language acquisition models proposed hitherto, especially Best's Perceptual Assimilation Model (e.g. [1]) and Flege's Speech Learning Model (e.g. [3]), are mostly consistent in their view that once the parameter is specified for a certain phonological feature, it is difficult to reset the parameter when acquiring a new language. As discussed in [2], if the parameter is not specified the transfer of L1 blank slate (i.e. absence of the distinction) and subsequent

Full Access (i.e. L1-like acquisition of the contrast) will occur.

Our results suggest that intensity is easier to be manipulated in L2 English lexical stress because it is not specified in L1 Japanese phonology. In other words, although L1 transfer of 'non-specification' occurs, the interference is relatively easy to overcome.

Although F0 is used phonologically to manifest Japanese lexical accent, the use of the parameter is same in Japanese and English; namely, accented syllables (or morae) are higher-pitched compared to unaccented ones. The potential positive transfer might explain the seeming hyper-correction by med-group in use of F0; they used even higher pitch than native speakers for stressed vowels.

Vowel duration is already specified in Japanese segmental phonology for short and long vowel contrast and is therefore more susceptible to Japanese mora-timed rhythm (e.g. [8]).

Since Japanese does not use vowel quality to manifest prosody, the transfer of L1 blank slate and subsequent full access are expected. In comparison to the results for intensity, spectral contrast was consistently absent in all learner groups. The difference in the results for intensity and spectral contrast could be due to the fact that intensity is basically a suprasegmental parameter, whereas spectral contrast is both segmental as well as suprasegmental.

Therefore, it appears that L1 interference on L2 segmental phonology seems to be more difficult to overcome than L1 interference on L2 prosody. This would explain why in the current analysis spectral contrast was far more difficult for Japanese speakers to manifest than the contrast in vowel duration. Other research has also shown that native-like spectral contrast is absent even in bilingual speakers [5].

6. CONCLUSION

The results of the current study suggest that L1 transfer happens differently depending on whether the feature is segmentally or prosodically contrastive in both first and second languages. In the case of Japanese learners of English, L1 interference on L2 English segments seems to be more difficult to overcome than its effect on English prosody. In particular, manifestation of vowel reduction, i.e. spectral contrast, seems to be the most difficult to achieve.

7. ACKNOWLEDGEMENTS

This research was supported by the Waseda University Grant for Special Research Projects No. 2014B-439 and JSPS KAKENHI Grant No.15H02729

8. REFERENCES

[1] Best, C. T. 1994. The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In H. C. Nusbaum (Ed.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge, MA: MIT Press. 167-224.

[2] Escudero, P., Boersma, P. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(04), 551-585.

[3] Flege, J. E. 1995. Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, 233-277.

[4] 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.

[5] 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.

[6] Meng, H., Tseng, C. Y., Kondo, M., Harrison, A., Viscelgia, T. 2009. Studying L2 Suprasegmental Features in Asian Englishes: a Position Paper. In *INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association*, 1715-1718.

[7] Munro, M. J. 2008. Foreign Accent and Speech Intelligibility. *Phonology and Second Language Acquisition*, 193-218.

[8] Port, R. F., Dalby, J., O'Dell, M. 1987. Evidence for Mora Timing in Japanese. *The Journal of the Acoustical Society of America, 81*(5), 1574-1585.

[9] Sheldon, A., Strange, W. 1982. The acquisition of /r/ and /l/ by Japanese Learners of English: Evidence that Speech Production Can Precede Speech Perception. *Applied Psycholinguistics, 3*(03), 243-261.

[10] Tatham, M., Morton, K. 2011. A Guide to Speech Production and Perception. Edinburgh University Press.

[11] 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 *2009 Oriental COCOSDA International Conference on Speech Database and Assessments*. 60-65. IEEE.