PHONETIC REALIZATION OF MULTIPLE STRESS LEVELS BY SPEAKERS OF A NON-STRESS LANGUAGE: A CASE OF JAPANESE-ACCENTED ENGLISH

Keiichi Tajima[1], Mafuyu Kitahara[2], and Kiyoko Yoneyama[3]

[1]Hosei University, [2]Sophia University, [3]Daito Bunka University
tajima@hosei.ac.jp, mafuyu@sophia.ac.jp, yoneyama@ic.daito.ac.jp

ABSTRACT

The present study investigated how well Japanese learners of English (JE) can differentiate stressed and unstressed syllables in English, by analyzing ten word pairs and five word triplets that exhibit stress shift, e.g., *compúter-còmputátion*, produced by high-proficiency (JEH) and mid-proficiency (JEM) learners and native American English speakers (AE). Results were as follows. First, JEH and JEM frequently produced devoiced vowels in unstressed syllables, e.g., *còmp[jʊ̥]tátion*, approximately in the same phonetic environments where vowels are devoiced in Japanese. Second, vowel duration and first-formant frequencies of primary-stressed, secondary-stressed, and unstressed syllables differed significantly across AE, JEH, and JEM. In particular, the acoustic differences between secondary-stressed and unstressed syllables were smaller for JE than for AE, but those between primary- and secondary-stressed syllables were larger for JE than for AE. Altogether, JE diverge from AE in their phonetic realization of English stress, showing native-language influences and difficulties with more than two stress levels.

Keywords: vowel reduction, vowel devoicing, vowel duration, vowel quality, stressed/unstressed syllables

1. INTRODUCTION

Speakers of a non-stress language such as Japanese often have difficulty producing distinctions in lexical stress in languages such as English. In particular, they may struggle with the phonetic realization of unstressed vowels, for instance, when producing the first vowel in the word *compúter*. Past studies that investigated the duration, pitch, intensity, and formant frequencies of weak syllables produced by second-language (L2) learners revealed that vowel reduction was highly correlated with English proficiency [1]. However, vowel quality adjustment was found to be difficult even for highly proficient Japanese learners of English, often reflecting a strong influence of orthography [2-4].

While theories vary in the treatment of English stress, English is commonly described as having two levels of stress: primary and secondary. Previous studies have focused on primary (1-stress) and unstressed (0-stress) vowels, but much less is known about how L2 learners produce secondary-stressed (2-stress) vowels.

Several preliminary studies have investigated how Japanese learners of English (JE) produce syllables with different degrees of stress by analyzing their production of English words that exhibit stress shift, e.g., *phótogràph-photógraphy* [5-7]. Results showed that JE often produced 0-stress vowels as devoiced vowels [5-7], approximately in the same phonetic environment in which short high vowels are devoiced in Japanese [8]. Furthermore, compared to 1-stress vowels, 0-stress vowels produced by JE were shorter in duration, but not lower in first-formant frequencies (F1), which would indicate a smaller jaw opening [6]. Finally, duration and F1 differed between 2- and 0-stress for vowels produced by native American English speakers (AE), but not by JE [6], suggesting that JE may not reliably distinguish between 2- and 0-stress vowels. Comparison between 1- and 2-stress vowels led to mixed results due to data limitations [7].

The purpose of the present study is to systematically compare the phonetic realization of three levels of stress, 1-, 2-, and 0-stress, produced by JE and AE, through acoustic analysis of 15 sets of English words that exhibit stress shift. Three hypotheses are examined. (1) JE do not differentiate 2- and 0-stress vowels, while AE do. (2) JE reduce 0-stress vowels more strongly in terms of duration than F1. (3) High-proficiency JE show characteristics more similar to AE than do mid-proficiency JE.

2. METHODS

2.1. Participants

Sixteen native Japanese speakers (JE) were recruited (5 males, 11 females). They were students at universities in the Tokyo area who had no experience of living in English speaking countries for more than 3 months. Their mean age was 21.5 years (SD: 1.55). JE were divided into two groups of eight participants each, high-proficiency (JEH) and mid-proficiency (JEM) learners, based on TOEIC scores. The average scores for JEH and JEM were 823 (SD: 64.6) and 531

(SD: 65.6), respectively. Eight native speakers of General American English (AE) living in the US were also recruited as a control group (4 males, 4 females). Their average age was 23.6 years (SD: 3.07). All participants had no history of hearing or speaking problems.

2.2. Material

Thirty-five derivationally related English words with alternating stress positions were chosen for the material. For example, *còmputátion-compúter* are a pair of words whose stress patterns are [2-0-1-0] and [0-1-0], respectively. This pair contains a 2-0 contrast in the first syllable and a 0-1 contrast in the second syllable. In this design, target vowels shared the same context; that is, consonants before and after the target vowel were consistent across pairs.

In addition to 10 such pairs of words, 5 triplets were included in the material. For example, *demócracy-démocràt-dèmocrátic* are a triplet of words whose stress patterns are [0-1-0-0], [1-0-2], and [2-0-1-0], respectively. Focusing on 1-0 contrast, there is one in the first syllable (*démocràt* vs. *demócracy*) as well as in the second syllable (*demócracy* vs. *démocràt-dèmocrátic*). One 2-0 contrast (*dèmocrátic* vs. *demócracy*) and two 1-2 contrasts (*démocràt* vs. *dèmocrátic* and *dèmocrátic* vs. *démocràt*) appear in this triplet as well.

2.3. Procedure

Participants were instructed to read target words in isolation presented on a computer screen. They read aloud the 35 words along with other filler words in one pseudo-random order and then in another order. The order of lists was counter-balanced across participants. Trained phoneticians checked their pronunciation on-site and asked for repetition when they made errors. For JE participants, phonetic symbols in IPA and stress positions were provided in addition to orthography.

For JE speakers, recordings were done in a sound-proof room. For AE speakers, recordings were done either in a quiet room or a sound-proof room. All the recordings were digitized at 44.1 kHz and 16 bits.

2.4. Analysis

Acoustic analyses were conducted using Praat [9]. Three acoustic properties were measured: (a) presence vs. absence of vowels, where a vowel was defined as present if it possessed two or more glottal cycles and had a clear vowel-like formant structure, (b) vowel duration (neighboring liquids which were hard to segment from the vowel were included in the vowel duration), and (c) first formant frequency (F1) as an indicator of jaw opening either at the center of the vowel, or at the 1/3 point when an inseparable liquid came after the vowel, or at the 2/3 point when it came before the vowel.

Statistical analyses were conducted on the *z*-score of either duration or F1 as a dependent variable and stress and speaker groups as independent variables. Linear mixed effects model implemented in the lme4 package [10] on R [11] was used. Speakers' intercept was the only random effect in the model.

3. RESULTS

3.1. Devoicing of unstressed vowels

As witnessed in previous studies, JE produced some words with devoiced vowels, i.e., vowels that did not meet the criterion in (a) above. Table 1 shows cases where "devoicing rate" was 25% or greater for one or more speaker groups. "Devoicing rate" is the percentage of productions among all tokens in which a particular vowel was devoiced.

Word	AE	JEH	JEM
còmputátion	0.0	62.5	50.0
cóntinent	12.5	12.5	37.5
còntinéntal	25.0	0.0	0.0
dìplomátic	0.0	0.0	25.0
hóspital	0.0	37.5	37.5
hòspitálity	0.0	25.0	37.5
imàginátion	0.0	25.0	42.9
nátional	0.0	50.0	25.0
nàtionálity	12.5	50.0	14.3
orìginálity	0.0	37.5	12.5
phòtográphic	0.0	25.0	25.0
photógraphy	12.5	25.0	12.5
polítical	0.0	37.5	37.5

Table 1: Vowels whose "devoicing rates" were 25% or greater for one or more speaker groups. Target vowels are underlined.

Table 1 shows that devoicing rate for JE was relatively high (37.5% or higher) for vowels in words such as *còmputátion*, *hóspital*, and *polítical*. Note that these vowels are flanked by voiceless consonants. Such phonetic environments are similar to those in which vowels undergo devoicing in Japanese. Table 1 further shows that devoiced vowels also appeared in contexts where the target vowel is not surrounded on both sides by voiceless consonants, e.g., *cóntinent*, *imàginátion*, and *nátional*. Moreover, devoiced vowels occurred occasionally in 2-stress vowels, e.g., *dìplomátic* and *phòtográphic*.

Devoicing rate for AE was not zero (12.5–25.0%) for some words. Rather than being "devoiced" per se, these vowels were likely deleted or reduced, as often seen in casual speech [12]. Such instances occurred

	0-1 contrast		0-2 contrast		1-2 contrast	
	Est.	*t*-value	Est.	*t*-value	Est.	*t*-value
--- duration ---						
Intercept	-0.570	-4.324***	-0.657	-4.308***	0.462	2.513*
JEH	0.054	0.291	0.305	1.419	0.283	1.089
JEM	0.186	0.992	0.493	2.289*	0.134	0.513
stress	1.393	17.049***	0.951	7.301***	-0.651	-5.217***
JEH:stress	-0.024	0.204	-0.674	-3.660***	-0.590	-3.326***
JEM:stress	-0.370	-3.149***	-0.924	-4.974***	-0.462	-2.543*
--- F1 ---						
Intercept	-0.458	-2.378*	-0.312	-1.455	-0.334	1.435
JEH	0.101	0.369	-0.105	-.347	-0.036	-0.110
JEM	0.352	1.287	0.007	0.024	0.005	0.002
stress	0.786	9.325***	0.876	7.974***	0.188	-1.700
JEH:stress	-0.045	-0.377	-0.611	-3.935***	-0.405	-2.593**
JEM:stress	-0.150	-1.237	-0.754	-4.793***	-0.408	-2.537*

$^{*}p < .05$ $^{**}p < .01$ $^{***}p < .001$

Table 2: Linear mixed effects model results for duration (top) and F1 (bottom) for 0-1 contrasts, 0-2 contrasts, and 1-2 contrasts. "Est" = parameter estimate. For speaker group, AE was the baseline condition.

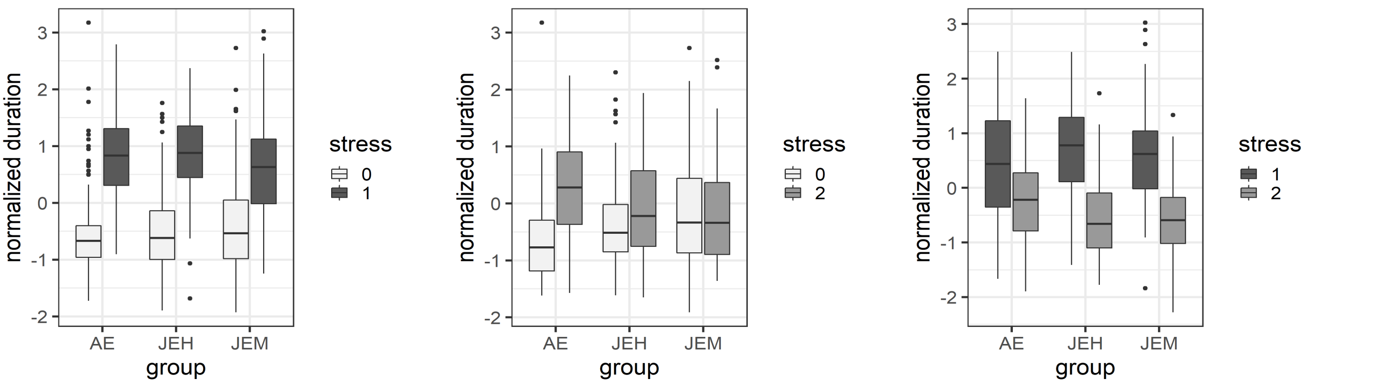

Figure 1: Boxplots of normalized vowel duration as a function of speaker group and stress level. Left: comparison between unstressed and primary-stressed vowels. Middle: comparison between unstressed and secondary-stressed vowels. Right: comparison between primary- and secondary-stressed vowels.

in fewer words and were less frequent in AE than JE. In fact, most words that exhibited vowel devoicing by JE were never produced with devoiced vowels by AE.

3.2. Vowel duration

To examine the effect of stress on vowel duration, vowel duration was compared for vowel pairs with different contrasts. Vowels in word-final syllables were excluded from duration comparisons because their duration may be adversely affected by final lengthening. A total of 17 vowel pairs were compared for 0-1 contrasts. Since the vowels vary in vowel identity and phonetic environment across different word pairs, vowel durations were normalized (*z*-transformed) for each pair and speaker group, and then aggregated for analysis. In addition to 0-1 contrasts, 10 vowel pairs each were compared for 0-2 and 1-2 contrasts.

Figure 1 shows boxplots of normalized vowel duration as a function of speaker group and stress level. The left panel illustrates results for 0-1 contrasts. Statistical analysis in Table 2 indicate that 0-stress vowels were significantly shorter than 1-stress vowels for all speaker groups, but that the magnitude of the distinction was smaller for JEM than for JEH and AE.

The middle panel of Figure 1 shows results for 0-2 contrasts. In addition to Table 2, individual comparisons within each group indicated that 2-stress vowels were significantly longer than 0-stress vowels for AE ($p < .001$) and JEH ($p < .05$), but not for JEM. These results indicate that AE produced systematically longer 2-stress vowels than 0-stress vowels, whereas JEM did not make such distinctions. JEH showed intermediate results.

Finally, the right panel of Figure 1 shows results for 1-2 contrasts. Individual comparisons per group indicated that 1-stress vowels were significantly longer than 2-stress vowels for all groups ($p < .001$).

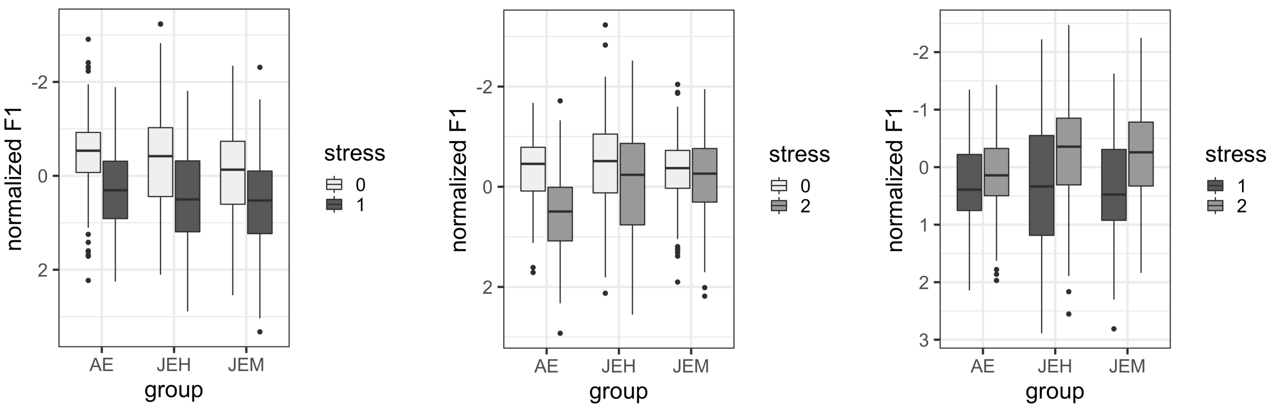

Figure 2: Boxplots of normalized F1 as a function of speaker group and stress level. The vertical axis is reversed so that downward direction indicates greater jaw opening. The left, middle, and right panels show the same pairwise comparisons of stress levels as Figure 1.

Together with Table 2, these results indicate that the duration difference between 1- and 2-stress syllables is greater for JEH and JEM than for AE.

3.3. Vowel formant frequencies

To examine the effect of stress on vowel quality, F1 was analyzed including vowels in word-final syllables, for a total of 22 pairs for 0-1 contrasts, 11 pairs for 0-2 contrasts, and 13 pairs for 1-2 contrasts.

Figure 2 shows boxplots of *z*-transformed F1 as a function of speaker group and stress level. The left panel shows results for 0-1 contrasts. Statistical analyses in Table 2 shows that 0-stress vowels were produced with significantly less jaw opening than 1-stress vowels for all speaker groups.

The middle panel of Figure 2 illustrates results for 0-2 contrasts. Results of statistical analysis matched those of duration. Individual comparisons for each group also indicated that 2-stress vowels were significantly more open than 0-stress vowels for AE ($p < .001$) and JEH ($p < .05$), but not for JEM. These results demonstrate that AE produced more open 2-stress vowels than 0-stress vowels, whereas JEM did not make such distinctions. JEH showed intermediate results.

The right panel of Figure 2 illustrates results for 1-2 contrasts. Individual comparisons per group showed some differences from those of duration: AE showed no significant difference in F1 between 1- and 2-stress vowels, while both JE groups produced significantly more open vowels for 1- than 2-stress vowels ($p < .001$). These results suggest that AE do not differentiate 1- and 2-stress vowels in terms of F1, but JE realize a systematic difference in F1.

4. DISCUSSION AND CONCLUSION

The present paper investigated acoustic correlates of different stress levels in English as produced by JE and AE. Results indicated that JE differentiated 1- and 0-stress vowels as well as AE did for the most part, in general agreement with past studies [1-4]. As for 2-stress vowels, AE clearly distinguished them from 0-stress vowels, and did not distinguish them from 1-stress vowels in F1. In contrast, JE showed a much smaller distinction, if at all, between 2- and 0-stress vowels in both duration and F1. Vowel devoicing was also observed in both 0-stress and 2-stress vowels. Together, these results support hypothesis (1), and suggest that JE may not be fully aware of the contrast between 2- and 0-stress vowels.

Analysis of vowel duration and F1 in sections 3.2 and 3.3 indicated many similarities and few differences between these two acoustic parameters. JE showed systematic differences in both duration and F1 for 0-1 and 1-2 contrasts, but not for 0-2 contrasts. These results do not provide strong support for hypothesis (2), nor do they agree with studies that reported greater difficulty with vowel quality than duration for L2 learners of English [1-7]. Since only normalized F1 values were analyzed in the present study, it remains to be seen whether inclusion of other parameters, e.g. F2, would lead to alternative results.

Finally, some differences were found between the two JE groups, such as a larger 0-1 contrast in duration for JEH than JEM, and a somewhat larger 2-0 contrast in duration and F1 for JEH than JEM. These results support hypothesis (3) to some extent, and suggest that increased English proficiency, as measured by TOEIC scores, may be correlated with improved production of English stress by JE.

In conclusion, the present study suggests that JE can generally realize the distinction between 1- and 0-stress vowels, but they are not fully aware of 2-stress vowels, and often produce them as if they are 0-stress vowels. Insofar as such tendencies lead to intelligibility and communication problems, production of stress levels should be an integral element of L2 training and education.

5. ACKNOWLEDGMENT

This research was supported by JSPS Kakenhi Grant Numbers: 18K00549, 18K00662, 19K00585, 22K00625, 23K00489. We are grateful to Maria Arvelo, Prof. Miwako Hisagi, and Prof. Benjamin Munson for their help coordinating experiments in the US.

6. REFERENCES

[1] Lee, B., Guion, S. S., Harada, T. 2006. Analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28, 487–513.

[2] Kondo, Y. 2000. Production of schwa by Japanese speakers of English: An acoustic study of shifts in coarticulatory strategies from L1 to L2. In: Broe, M. B., Pierrehumbert, J. B. (eds), *Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge University Press, 29–39.

[3] Kondo, M. 2009. Is acquisition of L2 phonemes difficult? Production of English stress by Japanese speakers. *Proc. 10th GASLA 2009*, 105–112.

[4] Yazawa, K., Ozaki, Y., Short, G., Kondo, M., Sagisaka, Y. 2015. A study of the production of unstressed vowels by Japanese speakers of English using the J-AESOP corpus. *The 18th Oriental COCOSDA /CASLRE*. DOI: 10.1109/ICSDA.2015. 7357872.

[5] Kitahara, M., Tajima, K., Yoneyama, K. 2022. Phonetic Realization of Vowel Reduction by Japanese Learners of English: A Preliminary Corpus Analysis. In Kubozono, H., Morimoto, M. (eds), *Prosody Kenkyu no Shintenkai (New Developments in Prosody Research)*, Kaitakusha, 258–278 (in Japanese).

[6] Tajima, K., Kitahara, M., Yoneyama, K. 2022. Vowel reduction by speakers of a non-stress language: A preliminary corpus analysis of Japanese-accented English. Oral presentation at *New Sounds 2022*, held at Universitat de Barcelona, Apr. 20–22, 2022.

[7] Yoneyama, K., Kitahara, M., Tajima, K. 2022. Acoustic analysis of English vowel reduction by Japanese college students. In *Proc. 2022 General Meeting of the Phonetic Society of Japan*, 109–114 (in Japanese).

[8] Kondo, M. 1997. Mechanisms of vowel devoicing in Japanese. Doctoral Thesis, University of Edinburgh.

[9] Boersma, P., Weenink, D. 2022. Praat: doing phonetics by computer [Computer program]. Version 6.2.17, retrieved from http://www.praat.org/.

[10] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, 1–48. https://doi.org/10.18637/jss.v067.i01.

[11] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-roject. org/.

[12] Tajima, K. and Shattuck-Hufnagel, S. 2012. Phonetic characteristics of syllable reduction and enhancement in American English. *Journal of the Acoustical Society of America.* 132, 2005.