10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Effects of Japanese Prosody on English Word Production: Interaction between Voicing and Gemination

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

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

Abstract

It is well attested that vowels are longer before voiced than voiceless consonants in English. Vowel duration in productions by American English speakers (AE) is affected especially by the voicing of postvocalic consonants and the quality of the target vowel [1]. Similar voicing effects in Japanese infants and adults, and also in English produced by Japanese learners (JE) were found as well [2]. However, previous studies on voicing effects in JE overlooked an important factor: gemination in loanwords. Data from a production experiment using monosyllabic English words with and without gemination in their borrowed counterparts were analyzed. Participants were 15 university students in Tokyo from intermediate to advanced proficiency levels in English, and two AE speakers for control. Results showed that vowel duration for JE was conditioned not by post-vocalic voicing, but instead by geminatability of the word, which was independently judged by 12 Japanese listeners in a different experiment. Post-vocalic stop closure duration was longer for voiceless than voiced stops, and longer for words that were judged to have a geminate stop than words that did not, for JE but not for AE speakers. Thus, voicing effects attested in AE speech is not straightforwardly realized in JE because of the interference from Japanese prosody.

Index Terms: loanword phonology, geminatability, English learners of Japanese, duration

1. Introduction

Vowels tend to be longer before voiced than voiceless consonants in many languages. For example, vowel duration in productions by American English speakers (AE) is affected especially by the voicing of postvocalic consonants and the quality of the vowel [1]. In Japanese, a similar voicing effect has been observed in Japanese produced by adults as well as infants. Furthermore, a voicing effect was also discovered in English produced by Japanese learners of English (JE), although the effects were substantially smaller than those reported for AE and were found only for advanced JE [2].

When English words are borrowed into Japanese as loanwords, the phonological structure of the original English words is altered to fit the phonological and mora-based prosodic structure of Japanese. Among the many modifications that take place during this process, such as epenthetic vowel insertion to simplify complex syllables, e.g. *stress* → *sutoresu*, post-vocalic stops undergo gemination when conditions are met. That is, certain words borrowed from English are realized with a geminate in Japanese (e.g. *kid* → kiddo, *rush* → rasshu). This can be seen as insertion of a moraic obstruent, e.g. ki.d.do (“.” indicates a mora boundary), reflecting the mora-based prosodic structure of Japanese. Gemination in Japanese loanwords is intricately conditioned by various factors such as tenseness and stress of the preceding vowel, and voicing of the obstruent in the source language [3], [4].

One question that arises in this context is the extent to which phonological and prosodic properties of the native language (L1) surfaces in the phonetic realization of words produced in a second language (L2). There is a line of research on the effect of loanwords written in katakana syllabary in English education [5] (and references therein). A wide range of issues in learning English under the effect of often misleading usage and pronunciation of loanwords are discussed. However, the effect of loanwords on the production of English words by learners is often a relatively uncultivated area in L2 phonetics. Generally, it is suspected that the Japanese loanword lexicon may interfere with the production of English words by JE speakers. If so, then gemination in Japanese loanwords might influence the durational properties of segments in English words produced by Japanese learners.

The purpose of this study is to investigate the degree to which post-vocalic voicing effects interact with gemination in loanwords in the phonetic realization of English words by JE speakers. Two hypotheses were addressed. The first hypothesis is that voicing effects are observed for both AE and JE productions because of general phonetic properties of post-vocalic voicing. The second hypothesis is that JE productions are further affected by the putative geminatability of English loanwords in Japanese. Here, “geminatability” means native Japanese speakers’ intuitions about the presence or absence of a geminate in Japanese loanwords. In order to test these hypotheses, a production experiment was conducted. A post-hoc analysis of the effect of vowel quality will also be discussed.

2. Methods

2.1. Corpus

A subset of the English words spoken by Japanese speakers of English (JE) in a previous study [2] were used in the present study. The JE group consisted of 15 participants, who were undergraduate students, graduate students and staff members at Daito Bunka University and Waseda University in Tokyo, Japan (male participants: 8, female participants: 7). They were native speakers of Japanese, most of whom had no experience living in English speaking countries for more than three

10.21437/SpeechProsody.2020-184

months. Their TOEIC scores were in the range between 450 and 750. Among them, three participants were in the 220-470 range, seven participants were in the 470-730 range, and five participants were in the 730-750 range. In addition to the JE speakers, two male native speakers of American English (AE) participated as a control group. All participants had no hearing difficulties.

The target items used in the current production study were 22 minimal pairs containing a postvocalic voiced/voiceless consonant, for a total of 44 words. Table 1 shows the items, along with the geminatability rates collected in a judgement experiment reported in section 2.3. The minimal pairs were monosyllabic CVC pairs or CCVC pairs containing one of seven different vowels ([aɪ][i:][ɪ][ɛ][æ][ɑ][ʌ]) and ending with a voiced or voiceless consonant. The stimuli were taken from an English phonetics textbook for college students [6].

2.2. Procedure

The production experiment was carried out individually in a quiet room. The task was administrated using Praat [7]. The participants were instructed to pronounce each target word embedded in a carrier sentence, "Say _____ now." Each target word in the carrier sentence was presented two times. The order of the stimuli was randomized, and each participant received a different unique randomization. If the participants made mistakes or hesitations within the sentence, they were instructed to repeat the trial sentences. Their productions were recorded with a high-quality microphone and were digitized at a 44.1-kHz sampling rate with 16-bit quantization. The production experiment was self-paced, and thus the total duration of the experiment varied from approximately 15 minutes to 25 minutes

2.3. Geminatability judgment

English words are sometimes pronounced with a geminate consonant by Japanese speakers. Kawagoe [3] reported that this geminate insertion in English words are triggered by vowel qualities. In order to check whether or not the target items used in the present study are treated by Japanese college students as containing a geminate consonant, a geminatability judgement experiment was conducted, separately from the production experiment with a separate group of participants.

Twelve native speakers of Japanese who were different from the speakers recorded in [2] participated in the geminatability judgment experiment. They were instructed to transcribe each target English word in katakana characters. The geminate rate, i.e. percentage of participants who transcribed the word with the geminate character "ッ", was calculated for each target word. Table 1 shows the geminate rate (%) for each item.

A target word was categorized as "geminate" if 9 (75%) or more responses contained a geminate (e.g. *back* バック, *bag* バッグ). A target word was categorized as "singleton" if 3 (25%) or fewer responses had a geminate (e.g. *write* ライト, *ride* ライド). As Table 1 shows, most target words were categorized more or less unambiguously as "geminate" or "singleton". Five words, i.e. *bead*, *Sid*, *said*, *cad*, and *dug*, had intermediate geminate rates (25%–75%) and were therefore categorized as "mixed."

Incidentally, while not the focus of the present paper, Table 1 also suggests other tendencies that have been pointed out previously, e.g. [4]. For example, words containing the vowel [i:] tend to be judged as singleton (see section 3.3), while words ending in [b] also tend to be judged as singleton.

2.4. Analysis

The five target words that were categorized as "mixed" in the geminatability judgment experiment were excluded in the following analyses. Also excluded from the analyses were mispronounced words, such as "bead" pronounced as [bɛd] and "cub" as [kæb], as well as tokens which did not have a visible release of the final stop or affricate. As a result, 1125 tokens were analyzed (17 speakers × 39 words × 2 repetitions – 201 excluded). Segmentation was performed using Praat [7]. The start and end points of each target vowel were first segmented automatically, and manually corrected by the authors, who all had been trained as phoneticians. The starting point of the vowel generally corresponded to the initial sharp rise in the power of the first formant, and the end of the vowel, which also marked the beginning of the following consonant closure, corresponded to the abrupt decrease in power due to the consonantal closure. The end point of the closure was also marked based on the existence of a burst of the consonant.

Table 1: *Target words and their geminate rates (%).*

	voiceless		voiced	
vowel	word	gem rate	word	gem rate
[aɪ]	write	0.0	ride	0.0
[i:]	beat	0.0	bead	33.3
	feet	8.3	feed	0.0
	leak	0.0	league	0.0
[ɪ]	kit	100.0	kid	100.0
	lit	100.0	lid	83.3
	mitt	100.0	mid	83.3
	rich	100.0	ridge	91.7
	sit	100.0	Sid	58.3
[ɛ]	bet	100.0	bed	100.0
	peck	83.3	peg	25.0
	set	100.0	said	41.7
[æ]	back	100.0	bag	100.0
	batch	100.0	badge	83.3
	cap	100.0	cab	0.0
	cat	100.0	cad	50.0
[ɑ]	got	100.0	god	100.0
	knots	100.0	nods	100.0
	mop	100.0	mob	0.0
[ʌ]	but	91.7	bud	75.0
	cup	91.7	cub	0.0
	duck	100.0	dug	33.3

Duration of the vowel and the following stop or affricate closure were then analyzed using linear mixed effects models implemented on R version 3.4.0. The fixed factors of the models were speaker group (AE, JE), voicing (voiceless, voiced) and geminatability (geminate, singleton) of the word-final post-vocalic stop. The random factors were the by-speaker and by-word random intercepts, and by-speaker random slopes for voicing and geminatability, depending on the model. Appropriate random factors were selected by starting from the simplest model, incrementally adding one

random factor at a time, and testing if the latter model has greater explanatory power than the former. Models which had the least number of random factors while having the greatest explanatory power were selected. The dependent variables were the vowel duration and following stop closure duration.

3. Results

3.1. Vowel duration

Figures 1 shows boxplots of the vowel duration as a function of speaker group (AE, JE), voicing (voiceless, voiced) and geminatability (singleton, geminate). Results for the AE speakers in the left half of Figure 1 show a clear effect of voicing for native speakers, and only a small effect of geminatability, if any. In contrast, results for the JE speakers in the right half of Figure 1 show a noticeable effect of geminatability, but only a small effect of voicing.

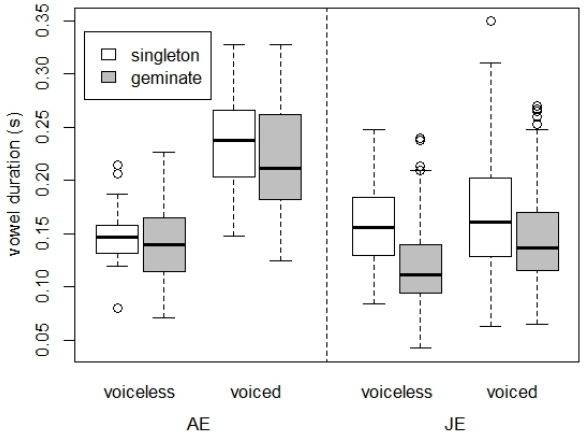

Figure 1: *Vowel duration (s) across conditions. Vowels were longer on average before voiced than voiceless consonants; this trend was stronger for AE than JE speakers. Vowels were also shorter before geminate than singleton consonants for JE speakers.*

A mixed effects model on the data in Figure 1 showed a significant main effect of voicing (t = 3.569; p < .01), a significant group-by-voicing interaction (t = –3.660; p < .01), and a significant group-by-geminatability interaction (t = –2.175; p < .05). No other main effects or interactions were significant. Further analysis of the group-by-voicing interaction indicated that vowels were significantly longer before voiced than voiceless stops for both AE speakers (t = 5.904; p < .01) and JE speakers (t = 3.074; p < .01), although the magnitude of the difference was larger for AE than JE speakers. Likewise, further analysis of the group-by-geminatability interaction indicated that vowels were significantly shorter before geminate than singleton consonants for JE speakers (t = –4.226; p < .001), but not for AE speakers (t = –1.809; *n.s.*).

3.2. Stop closure duration

Figure 2 shows boxplots of the post-vocalic stop closure duration as a function of speaker group, voicing, and geminatability. Figure 2 reveals a slight tendency for stop closures to be longer for voiceless than voiced stops for both AE and JE speakers. For JE speakers, closure duration further appears to be somewhat longer for geminate than singleton stops.

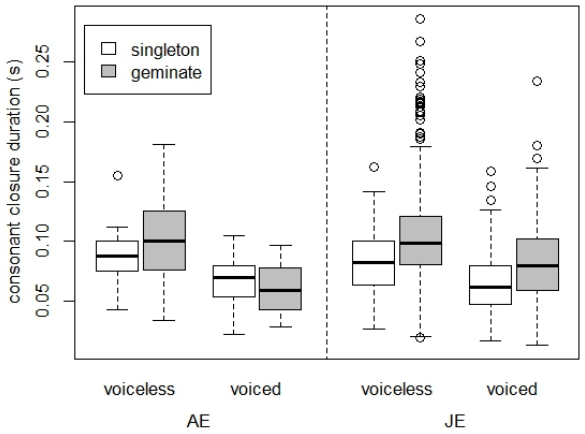

Figure 2: *Post-vocalic consonant closure duration (s) across conditions. Closure durations were on average longer for voiceless than voiced consonants.*

A linear mixed effects model fitted to the data in Figure 2 revealed a significant main effect of voicing (t = –2.099, p < .05). That is, closure duration was significantly longer for voiceless than voiced consonants. No other main effects or interactions were significant.

3.3. Results by vowel quality

Although the present study was not originally designed to investigate the effect of vowel quality, post-hoc analyses revealed that the results varied widely depending on the vowel. Thus, the data were separately examined for words containing different vowels. Table 2 shows the mean vowel durations and stop closure durations separately for target words with different vowels. Some of the cells in Table 2 are empty, indicating that some target words were consistently judged in the geminatability judgment experiment as either "singleton" or "geminate". For example, for target words with voiceless word-final stops, words containing [aɪ] or [i:] were consistently judged as singleton, e.g. *write* ライト, *beat* ビート, whereas words containing [ɪ], [ɛ], [æ], [ɑ], or [ʌ] were judged as geminate, e.g. *kit* キット, *bet* ベット, *back* バック, *got* ゴット, *but* バット. Such differences in vowel quality and geminatability are not associated with large differences in vowel or closure duration for AE speakers. In fact, Table 2 shows that target words with [aɪ][i:] have a mean vowel duration of 0.149s and a mean closure duration of 0.088s, while words with [ɪ][ɛ][æ][ɑ][ʌ] have a mean vowel duration of 0.140s and a mean closure duration of 0.101s. In contrast, JE speakers show a much larger difference in vowel duration between words with [aɪ][i:] (mean = 0.157s) and words with [ɪ][ɛ][æ][ɑ][ʌ] (mean = 0.117s). The difference in closure duration is also somewhat larger for JE speakers (0.083s vs. 0.106s) than for AE speakers.

For target words with voiced word-final stops, Table 2 shows that words containing [ɛ][æ][ɑ][ʌ] were judged as both "singleton", e.g. *peg* ペグ, *cab* キャブ, *mob* モブ, *cub* カブ, and "geminate", e.g. *bed* ベッド, *bag* バッグ, *god* ゴッド,

Table 2: *Vowel duration and following stop closure duration (s) for target words with different vowels as a function of speaker group, voicing, and geminatability. sin. = singleton, gem. = geminate.*

		AE				JP			
		voiceless		voiced		voiceless		voiced	
vowel	segment	sin.	gem.	sin.	gem.	sin.	gem.	sin.	gem.
ɑɪ	vowel	0.190		0.295		0.177		0.228	
	clo.	0.080		0.051		0.094		0.044	
i:	vowel	0.135		0.250		0.153		0.193	
	clo.	0.090		0.048		0.081		0.058	
ɪ	vowel		0.117		0.174		0.097		0.123
	clo.		0.086		0.066		0.104		0.082
ɛ	vowel		0.128	0.214	0.205		0.106	0.141	0.138
	clo.		0.098	0.071	0.047		0.108	0.080	0.086
æ	vowel		0.167	0.229	0.276		0.128	0.151	0.172
	clo.		0.109	0.074	0.063		0.104	0.064	0.077
ɑ	vowel		0.166	0.254	0.272		0.139	0.176	0.167
	clo.		0.104	0.088	0.055		0.107	0.073	0.087
ʌ	vowel		0.121	0.171	0.191		0.119	0.136	0.144
	clo.		0.109	0.085	0.056		0.108	0.072	0.095
mean	vowel	0.149	0.140	0.237	0.217	0.157	0.117	0.170	0.147
	clo.	0.088	0.101	0.066	0.060	0.083	0.106	0.066	0.084

bud バッド, depending on the word. For AE speakers, such differences in geminatability once again are not associated with large differences in vowel duration (mean = 0.237s for singleton and mean = 0.217s for geminate) and closure duration (mean = 0.066s for singleton and mean = 0.060s for geminate). On the other hand, for JE speakers, such differences in geminatability are associated with somewhat larger differences in vowel duration (mean = 0.170s for singleton and mean = 0.147s for geminate) and closure duration (mean = 0.066s for singleton and mean = 0.084s for geminate).

4. Discussion

In summary, the purpose of the present study was to investigate the voicing effects in production of English words by AE and JE where the words are subcategorized by the geminatability in Japanese loanwords.

Results demonstrated that AE speakers showed a systematic effect of voicing on vowel duration and following closure duration, as expected and as has been reported in the literature [1]. That is, vowels were longer and the following consonant closures were shorter for voiced than voiceless post-vocalic consonants.

For JE speakers, the present study found an effect of voicing on vowel duration. Although the magnitude of this effect was smaller than that found in AE speakers, this result is consistent with Hypothesis 1, and agrees with findings in our previous study [2].

In addition, the JE speakers showed a significant effect of voicing on postvocalic stop closure duration as well. However, the JE speakers showed a systematic effect of geminatability on vowel duration but not stop closure duration, which is in partial support of Hypothesis 2. This may in turn suggest that the phonological and prosodic structure of English loanwords in Japanese, in particular the phonological factors that predict the presence or absence of a moraic obstruent in loanwords, exerts an influence on the production of English words by Japanese speakers. From a practical standpoint, JE speakers need to be made aware of the tendency for Japanese prosodic structure to influence their English pronunciation, and appropriate instruction should be provided to mitigate such effects.

A secondary finding of this study is that JE speakers produced a much larger durational distinction between so-called "long" vowels, i.e. [aɪ][iː], and "short" vowels, i.e. [ɪ][ɛ][æ][ɑ][ʌ], than do native AE speakers. Such exaggerated distinctions may override the effect of voicing-related vowel lengthening, which again suggests that prosodic organization, in this case the moraic status of vowels in loanwords, constrains the English pronunciation of the target words by JE.

5. Conclusion

The results indicated that Hypothesis 1 was supported. The voicing effect was observed in both AE and JE speakers. Hypothesis 2 was also supported, suggesting that JE speakers' vowel duration was further modulated by the existence of geminates in the loanword counterpart in the Japanese lexicon. More extensive research is needed to understand the complex relationship among loanword production, English production, and L1 phonology/prosody.

6. Acknowledgements

The research was supported by JSPS Kakenhi 18K00549, 18K00662, 19K00585, and NINJAL collaborative research project 'Cross-linguistic Studies of Japanese Prosody and Grammar.'

7. References

[1] House, A. S. **(1961)**. On Vowel Duration in English, *JASA*, **33**, 1174-1178.

[2] Yoneyama, K., and Kitahara, M. **(2014)**. Voicing Effect on Vowel Duration: Corpus Analyses of Japanese Infants and Adults, and Production Data of English Learner, *Journal of Phonetics Society of Japan*, **18-1**, 30-39.

[3] Kawagoe, I. **(1995)**. Shakuyoogo ni miru sokuonka to rizumu shototsu (Consonant gemination in loanword phonology and rhythm clash) [in Japanese], ***Gengo Kenkyu***, **108**, 46-73.

[4] Ono, K. **(1991)**. Gairaigo to shiteno eigo no sokuonka ni tsuite. (/Q-insertion rule in English loan words in Japanese) [in Japanese], ***Gengo Kenkyu***, **100**, 67-88.

[5] Harada, Y., Morishita, M., and Hiramatsu, Y., (**2019**) Adverse Effects of Katakana Words in Japanese for Japanese Learners of English [in Japanese], *Proceedings of the 36th Cognitive Science Society of Japan 2019*, 503-511.

[6] Takebayashi, S., and Saito, H. **(2008)**. *Eigo Onseigaku Nyuumon* (An introduction to English Phonetics) [in Japanese], Taksyukan: Tokyo.

[7] Boersma, P., and Weenink, D. **(2019)**. Praat: doing phonetics by computer [Computer program]. Version 6.1.01, retrieved from http://www.praat.org/.