# Production of Syllable Structure in a Second Language: Factors Affecting Vowel Epenthesis in Japanese-accented English

Keiichi Tajima*, Donna Erickson†, and Kyoko Nagao††

*ATR Human Information Science Laboratories

† Gifu City Women's College

†† Indiana University

A salient characteristic of English spoken by native Japanese speakers is their tendency to insert epenthetic vowels between consonants or following word-final consonants (e.g., English stress as /su.to.re.su/). This study conducted speech analysis of a set of Japanese-accented English utterances to obtain quantitative estimates of how often vowel epenthesis actually occurs, and evaluated the effects of various factors that may influence vowel epenthesis. Results revealed significant effects of speaking rate and phonetic environment on the rate of occurrence of epenthetic vowels, but little effect of word familiarity and presence of Japanese loanwords for the English words.

## 1 Introduction

### 1.1 Vowel Epenthesis in Japanese-Accented English

In order to develop speech communication skills in a second language (L2), one needs to learn not just the individual speech sounds of L2 but also the global rhythmic characteristics of L2. One aspect of rhythmic structure involves the inventory of syllable shapes that are permitted in a language. This shows considerable cross-linguistic variation (Blevins 1995) as demonstrated, for example, by comparing Japanese and English. In Japanese, there are only four basic syllable shapes: (C)V, (C)VV, (C)VN (N = nasal), and (C)VQ (Q = first part of a geminate obstruent), whereas English syllables exhibit greater variation in structure and can be quite complex, e.g., strict (CCCVCC). Such differences in syllable structure are a potential source of difficulties for Japanese learners attempting to master the sound system of English. In fact, a salient characteristic of Japanese-accented English utterances is said to be the tendency to avoid consonant clusters and word-final consonants through vowel epenthesis, the production of extra, "non-lexical" vowels, e.g., sneezed /snizd/ pronounced as /sunizudo/. This problem is further exacerbated by the abundance of foreign words, many of which are English, that have been borrowed into Japanese as loanwords. In the loanwords, the foreign source words are phonologically assimilated to conform to the primarily CV-based structure of Japanese, as illustrated below:

| | |
|---|---|
| dream→/dori:mu/ | breakfast→/burek:ufa:suto/ |
| text→/tekisuto/ | McDonald's→/makudonarudo/ |

The effects of cross-linguistic differences in syllable structure have been studied in several domains. For example, sound changes that words undergo when borrowed into another language have been the focus of many phonological studies. Various proposals have been made to account for the complex pattern of sound correspondences that exist between Japanese loanwords and the foreign source words (Lovins 1973, Katayama 1998). Likewise, cross-linguistic differences in syllable structure have also been studied with respect to their effect on L2 speech perception. For example, Erickson et al. (1999) and Tajima et al. (2000) used a syllable-counting task to show that native Japanese speakers do not accurately perceive syllables in spoken English words. Dupoux et al. (1999) ran a series of perception tests to demonstrate that when native Japanese listeners are presented with consonant clusters in stimuli such as /ebzo/, they tend to perceive it as /ebuzo/ with an "illusory" epenthetic vowel intervening the two consonants, even though no vowel was intentionally produced within the cluster. Finally, Otake (2000) has utilized a tapping task to show that native Japanese listeners tend to perceive stimuli in Japanese as well as English in terms of moras, while native English listeners do so in terms of syllables, suggesting cross-linguistic differences in the rhythmic unit (moras vs. syllables) that is dominant during perception.

The effect of syllable structure differences across languages on L2 speech production seems obvious from anecdotal evidence, but there have been relatively few laboratory studies of this phenomenon. Some studies in the ESL literature have examined vowel epenthesis in utterances produced by Japanese learners of English, in the context of various phonological theories to predict where vowel epenthesis would occur (Ross 1994, Hancin-Bhatt and Bhatt 1997). But the method of determining the presence of epenthetic vowels was not clearly specified, or was based on auditory judgments, often from a single judge, without verification using acoustic analysis. As will be discussed later, however, epenthetic vowels are not categorically present or absent in speech, but instead tend to vary in the degree to which they can be reliably labeled as vowels. Thus, auditory inspection may not be a sufficient method for identifying epenthetic vowels. In one study that did employ acoustic analysis of Japanese-accented English (Saunders 1987), epenthetic vowels were not clearly identified in the Japanese speakers' productions, but the consonantal context examined was rather limited, to only voiceless /-ps/, /-ts/, and /-ks,/ clusters.

It appears, then, that even though vowel epenthesis in Japanese-accented English is well-known, a thorough empirical investigation of the phenomenon has not been carried out. Such a study would be well-motivated, given some evidence that deviations in syllable structure and temporal structure in foreign-accented utterances can reduce intelligibility by native listeners (Tajima et al. 1997).

The purpose of this study is to obtain quantitative estimates of how commonly vowel epenthesis occurs in Japanese-accented English based on a dataset of speech utterances collected from adult native Japanese speakers. In obtaining such quantitative estimates, it was expected that the likelihood of producing an epenthetic vowel would be influenced by a multitude of factors, such as the degree to which Japanese speakers are familiar with a given English word, the overall speaking rate, and the phonetic content of the stimuli. Thus, a number of variables were examined, as described in Section 1.2. Moreover, unlike most "lexical" vowels, production of an epenthetic vowel was not expected to be a clear-cut, all-or-none phenomenon. That is, many utterances were expected to contain borderline cases, in which judgment of the presence vs. absence of an epenthetic vowel could fluctuate depending on the criteria adopted. The current study attempted to address this problem by employing a fairly explicit set of judgment criteria, and obtaining independent judgments from multiple phonetically trained judges based on acoustic analysis and auditory inspection.

### 1.2 Factors that Potentially Influence Vowel Epenthesis

This section describes factors that may influence how often native Japanese speakers might produce epenthetic vowels in English words. The hypotheses below are not meant to represent an exhaustive list of factors influencing vowel epenthesis. Instead, they reflect the view that foreign-accented speech is potentially influenced by a multitude of factors, some of which are psychological factors (e.g., word familiarity), some of which are task-specific variables (e.g., speaking rate), and other of which involve phonetic parameters (e.g., consonantal properties). First, the likelihood that native Japanese speakers would produce epenthetic vowels in an English word may be influenced by their subjective familiarity with the word, or in a related way, the presence or absence of a Japanese loanword for the English word.

Native Japanese speakers take six years of English classes during secondary school, where grammar is the primary focus rather than conversational skills. While spoken English is not the primary focus at schools, spoken English is nowadays frequently heard in Japan in the mass media, e.g., imported music, radio shows, TV commercials, etc. As such, it is likely that some English words are very familiar to native Japanese speakers, while other English words are not familiar to them at all. It seems reasonable to hypothesize that speakers are better able to produce native-like renditions of words that they are familiar with than words that they have not encountered before. For clarity, this is re-stated in Hypothesis 1:

**Hypothesis 1:** Vowel epenthesis occurs less frequently in English words that are subjectively rated as more familiar than in words that are rated less familiar.

In addition to degree of familiarity, the presence vs. absence of Japanese loanwords for the English words, and the frequency of use of the loanwords, may also be a factor affecting vowel epenthesis. It is possible that the presence of a Japanese loanword for an English word would lead Japanese speakers to produce the word in a "loanword-like" fashion, increasing the likelihood of vowel epenthesis. Alternatively, it is also possible that the presence of loanwords is associated with greater familiarity with the English words. If this is so, and if Hypothesis 1 is correct, then English words with loanwords would be produced in a more native-like fashion than English words with no loanwords. These two considerations yield two conflicting hypotheses:

**Hypothesis 2a:** Vowel epenthesis occurs more frequently in English words that exist in Japanese as loanwords than in words that do not. Furthermore, the more frequently the loanwords appear in Japanese, the stronger the influence of loanword forms on English pronunciation, hence the more likely it is for vowel epenthesis to occur.

**Hypothesis 2b:** Vowel epenthesis occurs less frequently in English words that exist in Japanese as loanwords than in words that do not. Furthermore, the more frequently the loan words appear in Japanese, the more familiar the words are to Japanese speakers, hence the less likely it is for vowel epenthesis to occur.

An additional set of factors that might affect the rate of occurrence of vowel epenthesis pertains to the experimental conditions under which productions are elicited from native Japanese speakers. It is possible, for example, that instructions to speak at a slower rate leads to more instances of vowel epenthesis than speaking at a normal rate, since a slow rate of speech might induce a clear and careful articulation of individual consonants, accompanied by an audible vowel-like release following each consonant. This leads to Hypothesis 3:

**Hypothesis 3:** Vowel epenthesis occurs more frequently in a slower speaking rate than in a normal one.

It is also conceivable that the likelihood of vowel epenthesis is affected by whether a target word is produced in isolation or within a sentence. On one hand, it is possible that producing a single isolated English word might be a more favorable condition for speakers to attempt a "best" (i.e., native-like) pronunciation of the word than producing a whole sentence. On the other hand, however, presentation of a single word might lead speakers to pay close attention to each sound, leading to more careful articulation of individual consonants (potentially accompanied by vowel epenthesis) than when presented within a sentence. These considerations lead to two alternative hypotheses:

**Hypothesis 4a:** Vowel epenthesis occurs less frequently in a word when produced in isolation than when produced in a sentence.

**Hypothesis 4b:** Vowel epenthesis occurs more frequently in a word when produced in isolation than when produced in a sentence.

As a final set of factors, it is possible that epenthetic vowels are more likely produced in certain consonantal contexts than others. For example, it is possible that epenthetic vowels are more likely when surrounded by voiced consonants than by voiceless consonants, due to assimilation of voicing with the surrounding consonants. Moreover, Japanese speakers tend to devoice vowels between two voiceless obstruents in their native language, and may exhibit this tendency in production of English as well. (Note that epenthetic vowels are assumed to be voiced. The possibility of a voiceless epenthetic vowel is not pursued because its presence is difficult to determine instrumentally.) This leads to Hypothesis 5.

**Hypothesis 5:** Vowel epenthesis occurs more frequently in voiced environments than in voiceless environments.

It is also possible that stop consonants are more likely followed by an epenthetic vowel than affricates or fricatives. When a stop is released, the volume of air that has accumulated in the vocal tract during closure is pushed out. Affricates and fricatives, however, are not followed by a strong outward airflow because these sounds involve sustained production of frication noise, during which the volume of air inside the vocal tract gradually escapes through a narrow constriction. The relatively stronger airflow following stops than following affricates or fricatives may make it more likely for the former to be released and followed by a noticeable vowel-like segment than the latter, leading to Hypothesis 6.

**Hypothesis 6:** Vowel epenthesis occurs more frequently following stop consonants than following affricates or fricatives.

Finally, it is possible that labial stops are more likely followed by a vowel-like segment than alveolar stops, and alveolars more likely followed by a vowel than velars, since an anterior place of articulation allows a larger volume of air to accumulate in the vocal tract than a posterior one. Greater air volume may lead to a more forceful airstream following the constriction release, which may provide a more favorable condition for production of a vowel-like sound after release. Interestingly, an opposite trend has been reported when the rate at which sentence-final stops are released was compared across different places of articulation (Byrd 1993a). This study reported that native English speakers recorded in the TIMIT database (Zue 1990a) produced released stops as opposed to unreleased stops more frequently following velar stops than alveolar stops, and more frequently following alveolar stops than following labial stops. Note that Byrd's study counted released stops produced by native English speakers, which are not the same as epenthetic vowels produced by nonnative speakers. However, if the same trend also holds for epenthetic vowels, then two conflicting hypotheses are possible:

**Hypothesis 7a:** Vowel epenthesis occurs more frequently following word-final stops produced in an anterior place of articulation than following posterior ones.

**Hypothesis 7b:** Vowel epenthesis occurs less frequently following word-final stops produced in an anterior place of articulation than following posterior ones.

## 2 Methods

### 2.1 Dataset

A list of 90 English words were compiled. The words were selected such that the word set as a whole contained representative consonant clusters of English. Consonant clusters were classified according to word position, number of consonants in the cluster, as well as voicing, manner, and place of each consonant. The list of words are shown in Figure 1. Some of the words existed in Japanese as loan-words, written using katakana script, e.g., *basketball, platform*. Others were expected to be English words that Japanese students would encounter in school, e.g., *children, explain*. Other words were expected to be relatively unfamiliar to Japanese speakers, e.g., *glimpsed, thrive*.

**abduct**, accept, acting, actress, **adds**, **admit**, advance, asked, asks, **basketball**, blasts, bleached, **bragged**, buildings, caged, **chapter**, children, **clog**, clues, cringed, cubes, described, **Disney**, districts, drink, drums, entry, example, **exchange**, **explain**, extreme, films, flash, **formed**, **forms**, **friends**, front, **gently**, glimpsed, grapes, helps, **husband**, insult, legs, lens, looked, **magnets**, **monster**, **necklace**, oblige, obtain, **pacemaker**, photograph, **placed**, platform, postman, printer, proves, queens, **raindrop**, respond, rugby, sandstorm, scold, **scream**, shred, sickness, silver, six, skid, **sleeve**, slender, smart, smear, sneak, **sneezed**, spies, spills, **splash**, spray, **statement**, stretch, stunts, tank, tasty, **text**, **texts**, thrive, tray, **tribe**

Figure 1. List of 90 English words used in this study. Of these, 30 words (in boldface) were embedded in the carrier sentence, *I say___ in the list.*

Participants were 26 monolingual native Japanese college students from Doshisha University, Kyoto, Japan (15 females and 11 males, between 18 and 21 years of age). Most subjects spoke the Kansai dialect of Japanese. The subjects received six years of English education in secondary school, but had no experience living in an English-speaking community. In addition, a control group of 6 native American/Canadian English speakers also participated (3 females, 3 males, between 22 and 42 years of age). They were visiting researchers at ATR Laboratories, Kyoto, Japan. Their length of residency in Japan ranged from one to five years.

The recordings took place in an anechoic chamber at ATR Laboratories. During each subject's recording, (s)he was given a randomized list of the words to be produced. For each item, subjects heard the item produced by a male native American English speaker; this ensured that the subjects knew the pronunciation of each word. They were told not to imitate the native-speaker production. After they heard the word, subjects produced the word in isolation. Subjects read through the list twice, at two speaking rates: normal and slow. In addition, 30 of the 90 words (indicated in boldface in Figure 1) were embedded in the carrier sentence, *I say___ in the list*, and were read at a normal rate. The order of the three speaking conditions was counterbalanced across subjects. The utterances were recorded onto a DAT, and later saved as audio files at 22050-Hz sampling rate with 16-bit resolution. Each audio file was then amplitude-normalized so that peak amplitude was the same across all files.

### 2.2 Speech Analysis

Judgment concerning the presence vs. absence of epenthetic vowels in the utterances was obtained from three trained phoneticians (the authors of this paper). One of them is a native English speaker, one is a native Japanese speaker, and one is a bilingual in both languages. All received a graduate degree in Linguistics in the United States.

The three judges made independent judgments of the utterances collected. For each utterance, a waveform, spectrogram, fundamental frequency contour, amplitude envelope, and probability of voicing profile were displayed using ESPS/waves+. The judges listened to each token as often as necessary. The general criterion for marking an epenthetic vowel was that it had to be both audible and visible on the waveform or spectrogram. Care was taken to ensure that epenthetic vowels were clearly visible, by enforcing the following criteria: (1) Clear vowel-like formant structure must be visible on the spectrogram, particularly the first formant. (2) The waveform must show periodicity from phonation. (3) Very short epenthetic vowels must contain at least two pitch periods. (4) Epenthetic vowels that are adjacent to another consonant with formant structure (e.g., a nasal or liquid) must show evidence that identifies them as vowels (e.g., presence of a noticeable change in amplitude or a rapid formant transition).

To help the judges make consistent judgments, they were given a manual prepared by the first author that specified the above criteria and also pointed out specific situations in which judgment must be made carefully, e.g., detecting the presence of a very short epenthetic vowel following release of a voiced stop, and deciding between a sustained (syllabic) nasal or liquid vs. a nasal or liquid plus an epenthetic vowel (potentially nasalized if adjacent to a nasal segment). The judges were instructed to put a label on any '"extra", non-lexical vowel that met the criteria above, including vowels appearing within consonant clusters and vowels following word-final consonants. Inspection of the entire dataset took two to four months depending on the judge.

### 2.3 Word Familiarity and Frequency

To obtain estimates of native Japanese speakers' familiarity with the English words tested, familiarity ratings of the 90 words were obtained from six native Japanese speakers, 4 females, 2 males, between 24 and 31 years of age, all of whom were different from the 26 speakers in the dataset. Each rater was asked to rate the degree to which (s)he was familiar with each English word on a 7-point scale ranging from 1 (very unfamiliar) to 7 (very familiar). Each word was shown visually on the computer screen, and a male native speaker's production of the word was also presented through headphones. The raters responded by entering a number between 1 and 7 using the keyboard. The 90 words were presented once each in a random order. The raters were told to rate the overall familiarity with the words, and were not given any instructions to rate their familiarity specifically with the words' orthography, pronunciation, or their loanword counterparts in Japanese (if they exist).

In addition to familiarity ratings, we determined whether a Japanese loanword existed for each English word; if it did, we also obtained estimates of its frequency of use. This was done using the NTT Lexical Properties of Japanese database (Amano and Kondo 2000) a text corpus containing 14 years of Japan's Asahi newspaper articles. Of the 90 test words, we took 57 words which contained no inflectional suffixes (e.g., past tense or plural morphemes). We then converted each word into *katakana* script, used to transcribe foreign words in Japanese, then computed the word frequency of each *katakana* word in the NTT database. Test words with inflectional suffixes were not included here because it was unclear whether to search the database for the root + suffix form (which is not typically found among Japanese loanwords) or the root-only form (whose frequency may not be a good predictor of vowel epenthesis in the root + suffix form).

## 3 Results

### 3.1 Inter-judge Reliability

First, estimates of the level of agreement across the three judges were obtained. The dataset contained 17,978 possible positions in which an epenthetic vowel could potentially occur, i.e., between every pair of adjacent consonants and following word-final consonants, in every utterance collected. Of these, the three judges agreed on 90.2% of the judgments. That is, for 90.2% of the time, all judges agreed that an epenthetic vowel was present in a given position or that it was absent. For the remaining 9.8%, the judges gave mixed responses. Therefore, the three judges were generally in good agreement with one another. In addition, the percentage of times each of the three judges reported the presence of epenthetic vowels was 12.3%, 13.9%, and 11.6%. Thus, the judges did not differ much in their bias toward labeling whether an epenthetic vowel was present or absent.

### 3.2 Effect of Word Familiarity and Loanword Frequency

The first factor we examined is the effect of word familiarity on the likelihood of producing epenthetic vowels. Figure 2 plots for each test word the mean familiarity rating across the six raters on the x-axis, and the percentage of productions that contained epenthetic vowels on the y-axis. The latter was calculated by computing the percentage of tokens of the word that each judge labeled as containing at least one epenthetic vowel, then taking the mean across the three judges. Productions in the isolated normal-rate and isolated slow-rate conditions were analyzed; productions in the embedded condition were not considered because only 30 of the 90 words were produced in this condition. The figure reveals a reasonably wide variation across words in familiarity as well as rate of epenthesis. Familiarity ranged from 1.0 (*abduct*) to 7.0 (*drink, friends, front, six*), and rate of occurrence of epenthesis ranged from 0.0% (*monster, pacemaker, tasty*) to 71.2% (*formed*). There seems to be a somewhat denser population of data points near the lower right-hand corner of the scatterplot than other areas, suggesting that there were a number of high-familiarity words in which vowel epenthesis was relatively uncommon. There was a slight but significant negative correlation between familiarity and rate of epenthesis [Pearson's $r$ = -0.25; $t(88)$ = -2.65; $p$ - .006]. Although the figure suggests that the relationship is weak, the correlation is at least in the same direction as predicted by Hypothesis 1.

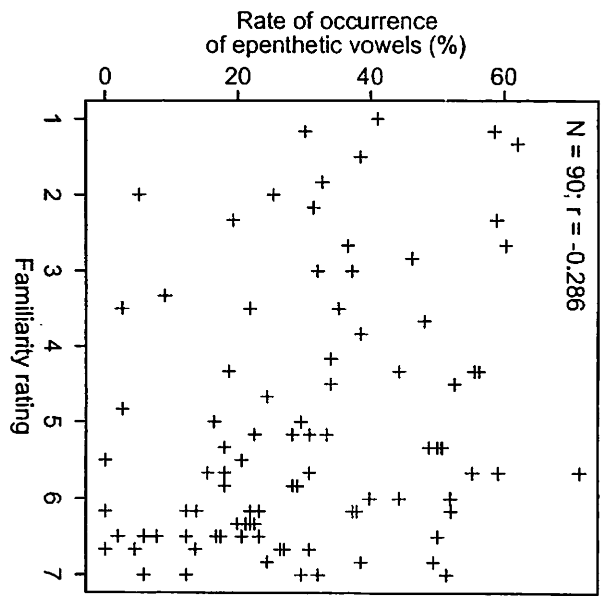


Figure 2. Rate of occurrence of epenthetic vowels in each word plotted against native Japanese speakers' subjective familiarity with the word.

In addition to degree of familiarity, we examined whether the presence vs. absence of Japanese loanwords affected the likelihood of vowel epenthesis. Of the 57 monomorphemic test words (see Section 2.3), Japanese loanwords were found in the NTT database (Amano and Kondo 2000) for 34 words (e.g., *Disney, drink, slender*), while loanwords were not found for the remaining 23 words (e.g., *admit, raindrop, sneak*). Figure 3 shows the mean rate of occurrence of epenthetic vowels for the two groups of words. The mean rate of epenthesis for words that had loanword counterparts was 22.5%, SD = 14.1, and the mean for words that did not have loanword counterparts was 30.7%, SD = 14.8, apparently in support of Hypothesis 2b. However, the same direction of difference was also found in the English data when the rate of epenthesis was computed for the same two groups of words: mean = 5.0%, SD = 7.0 (words with loanwords) and mean = 11.0%, SD = 11.3 (words with no loanwords). The differences were significant for both the Japanese data [$t(55) = 2.11$; $p = .039$] and English data [$t(55) = 2.47$; $p = .017$]. Since an effect of Japanese loanwords would not be expected for native English speakers' productions, it is suspected that the differences are the result of other overriding factors.

As an additional test, we examined whether rate of vowel epenthesis in English words that had loanword counterparts in Japanese was affected by word frequency of the loanwords. Among the 34 words that existed as loanwords, word frequency ranged from 1 for *chapter* and *photograph*, to 2,358 for *tank* and 2,840 for *rugby*. The correlation between the logarithm of the word frequency and rate of vowel epenthesis was found to be non-significant [Pearson's $r = -0.20$; $t(35) = -1.22$; n.s.].

Altogether, the results suggest a slight tendency for high-familiarity words to exhibit fewer instances of epenthetic vowels than low-familiarity words, in agreement with Hypothesis 1. However, neither Hypothesis 2a nor 2b concerning Japanese loanwords was supported; the presence or absence of loanwords and their word frequency did not turn out to be reliable predictors of how often Japanese speakers produce epenthetic vowels.

### 3.3 Effect of Speaking Condition

Next, we examine how vowel epenthesis in Japanese-accented English is influenced by experimental factors such as speaking rate and whether the speakers produced words or sentences. This analysis used 30 of the 90 test words that were produced in all three speaking conditions (indicated in boldface in Figure 1). For each word, we took all productions of the word by the speakers and calculated the percentage of productions that each judge labeled as containing an epenthetic vowel. Then a mean percentage across the judges was obtained separately for each speaking condition. Epenthetic vowels produced word-finally were not included in this analysis, since they were difficult to detect in the embedded condition, when they occur adjacent to the word in the carrier, *I say___ in the list.* Figure 4 shows the result, separately for the native Japanese and native English speakers.

The Japanese speakers' data reveal a higher rate of occurrence of epenthetic vowels at a slow speaking rate (28.2%) than at a normal rate (21.5%), a difference of about 7 percentage points. Rate of epenthesis in the embedded normal-rate condition (22.5%) is about the same as that in the isolated normal-rate condition (21.5%). The native English speakers' result indicates that epenthetic vowels are witnessed in a small percentage (3.0--5.4%) of their productions. Similar to the Japanese data, epenthetic vowels are slightly more common in the slow-rate condition (5.4%) than in the normal-rate condition (3.0%).

An analysis of variance (ANOVA) with Rate (normal, slow) as a within-subjects factor and Word (30 words) as a random factor revealed a highly significant effect of Rate [$F(1,29) = 18.28$; $p < .001$] for the Japanese speakers' data. The same analysis with the native English speakers' data revealed a marginally significant effect of Rate [$(1,29) = 4.64$; $p = .040$]. A similar ANOVA with Carrier Sentence (with carrier, without carrier) as a fixed factor showed no significant effect of Carrier, either for the Japanese data [$F(1,29) = 0.22$; n.s.] or for the English data [$F(1,29) = 1.00$ ; n.s.].

Thus, on average, roughly one out of every four to five words produced by the native Japanese speakers in this dataset exhibited epenthetic vowels. The native English speakers also did show instances of epenthetic, non-lexical vowels that met the judgment criteria above. It appears that speaking slowly leads to more instances of vowel epenthesis, confirming Hypothesis 3. In contrast, neither Hypothesis 4a nor 4b was supported by the data, since vowel epenthesis did not seem to be affected by whether a word is produced in isolation or embedded in a sentence.

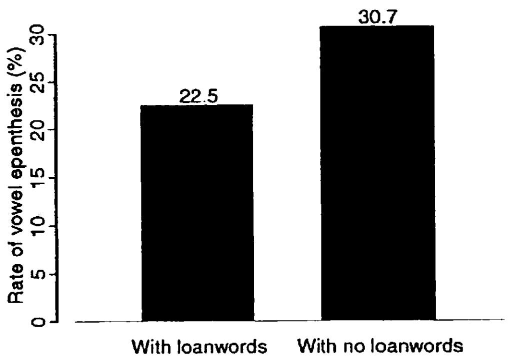


Figure 3. Percentage of words produced with epenthetic vowels as a function of the presence or absence of Japanese loanwords for the words. See main text for discussion of the result.

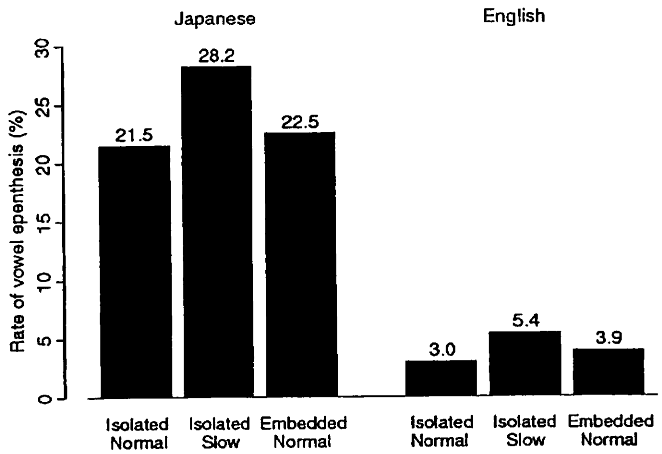


Figure 4. Percentage of words produced with epenthetic vowels as a function of speaking condition.

### 3.4 Effect of Phonetic Environment

We now turn to the effect of phonetic environment on the rate of vowel epenthesis in Japanese-accented English. In this analysis, we focus our attention on clusters which appear in word-final position, since it was relatively easy to devise a list of words containing word-final clusters with systematic combinations of consonantal properties such as voicing, place, and manner, e.g., by means of adding past tense /-d/ and plural /-z/ morphemes to words ending with various coda consonants. Clusters in word-initial and word-medial positions were also included in the dataset, but we will not examine them here.

To examine how voicing, place, and manner affect the rate at which epenthetic vowels occur between adjacent consonants, we examined data from ten words in the dataset that contained word-final 2-consonant clusters: *grapes* /-ps/, *accept* /-pt/, *magnets* /-ts/, *six* /-ks/, *looked* /-kt/, *cubes* /-bz/, *described* /-bd/, *adds* /-dz/, *legs* /-gz/, and *bragged* /-gd/. Figures 5, 6 and 7 examine the effect of voicing, place, and manner, respectively, for Japanese and English speakers. Each figure's caption indicates the words (and the clusters therein) based on which the rate of occurrence of vowel epenthesis was calculated. For example, in Figure 5, rate of epenthesis for voiced clusters was obtained by looking at productions of the words *cubes, described, adds, legs*, and *bragged*, and computing the percentage of tokens that contained an epenthetic vowel within the final cluster; epenthetic vowels produced at other word positions were ignored. The means were based on productions in the isolated normal-rate and slow-rate conditions.

Figure 5 compares the percentages in voiced-voiced vs. voiceless-voiceless clusters. The Japanese data clearly show that vowel epenthesis is more frequent in voiced environments than in voiceless environments. An ANOVA with Voicing as a within-subjects factor and Speaker as a random factor showed a highly significantly effect of Voicing $F(1,25) = 54.3$; $p < .001$], in agreement with Hypothesis 5. A trend in the same direction was also found in the native English data, but the difference was not significant [$F(1,5) = 2.3$; n.s.].

Figure 6 compares the rate of vowel epenthesis for clusters whose first consonant has different places of articulation. The figure shows that the percentages for the Japanese speakers' productions vary considerably by place, confirmed by a significant overall effect of Place for the Japanese data [$F(2,50) = 20.2$; $p< .001$]. To further determine which of the three places of articulation revealed significant differences, a post-hoc analysis was conducted using Tukey's HSD test. The test showed the following pattern of significant differences at the level: labial > velar > alveolar. That is, rate of epenthesis was higher within /-ps/ and /-bz/, i.e., following labials, than within /-ks/ and /-gz/, following velars, consistent with Hypothesis 7a. The alveolar clusters /-ts/ and /-dz/, however, showed a lower rate of vowel epenthesis than the velar clusters, inconsistent with Hypothesis 7a. This pattern is presumably due to the existence of the affricates /ts/ and /dz/ in Japanese, which makes it unlikely for epenthetic vowels to occur within this sequence. In the native English speakers' data, /-ts/ and /-dz/ clusters exhibited no epenthetic vowels at all, while a small percentage of the other clusters exhibited epenthesis. The differences, however, were not statistically significant [$F(2,10) = 1.2$; n.s.].

Figure 7 compares the means for stop-stop clusters (e.g., /-bd/ and /-gd/) vs. stop-fricative clusters (e.g., /-bz/ and /-gz/). No clear difference was found between these types of clusters, either for the Japanese data [$F(1,25) = 0.3$; n.s.] or the English data [$F(1,5) = 1.5$; n.s.]. Thus, manner of articulation of the second consonant in a cluster did not affect the rate of epenthesis. As we will see later, however, varying the manner of the preceding consonant (relevant to Hypothesis 6) leads to a different result.

Focusing on vowel epenthesis in word-final position offers a unique opportunity to examine vowel epenthesis, not only within consonant sequences, but also following singleton coda consonants that occur word-finally. Given that most consonants are not permitted in coda position in Japanese, it is expected that English words that end with a consonant would often be produced by native Japanese speakers with a word-final epenthetic vowel. To investigate this, we examined data from 12 words which end with singleton consonants: *raindrop*, /-p/, *admit* /-t/, *sneak* /-k/, *stretch* /-ʧ/, *photograph* /-f/, *actress* /-s/, *tribe* /-b/, *skid* /-d/, *clog* /-g/, *oblige* /-ʤ/, *thrive* /-v/, and *spies* /-z/. Figures 8, 9, and 10 show the rate of word-final vowel epenthesis as a function of the final consonant's voicing, place, and manner, respectively.

Figure 8 shows that word-final vowel epenthesis is more frequent following voiced consonants than voiceless consonants. The effect is clearly evident for both native Japanese and native English speakers' productions; in fact, the magnitude of the difference is slightly greater for the English speakers' productions. The differences were significant for both the Japanese data [$F(1,25) = 13.1$; $p = .001$] and the English data [$F(1,5) = 29.4$; $p = .003$]. This result is similar to that in Figure 5, and further confirms Hypothesis 5.

Figure 9 shows a tendency in the Japanese data for the rate of epenthesis to decrease as the place of articulation moves backward from labial to alveolar to velar. The overall effect of Place was found to be marginally significant [$F(2,50) = 3.2$; $p = .49$]. Tukey's HSD test found a significant difference between labial and velar stops at the $p = .05$ level (labial > velar), but no difference between these and alveolar stops. Similar to the result in Figure 6, the trend observed in Figure 9 is in agreement with Hypothesis 7a, and opposite from Hypothesis 7b. The native English speakers' data did not show a similar trend as the Japanese data. Consonant place did not have a significant effect on rate of epenthesis [$F(2,10) = 0.7$; n.s.].

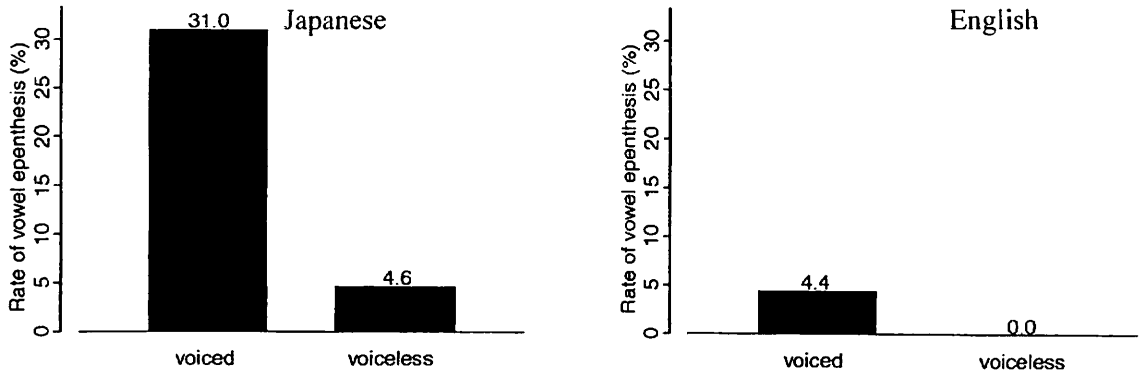


Figure 5. Rate of occurrence of vowel epenthesis as a function of voicing. "Voiced" data calculated from *cubes* /-bz/, *described* /-bd/, *adds* /-dz/, *legs* /-gz/, *bragged* /-gd/. "Voiceless" data calculated from *grapes* /-ps/, *accept* /-pt/, *magnets* /-ts/, *six* /-ks/, *looked* /-kt/.

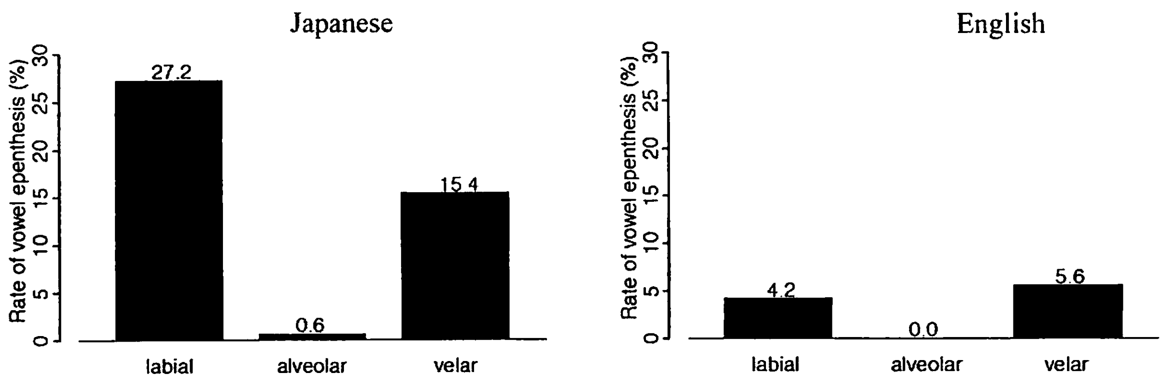


Figure 6. Rate of epenthesis as a function of preceding consonant's place. "Labial" data calculated from *grapes* /-ps/, *cubes* /-bz/. "Alveolar" data calculated from *magnets* /-ts/, *adds* /-dz/. "Velar" data calculated from *looked* /-kt/, *bragged* /-gd/.

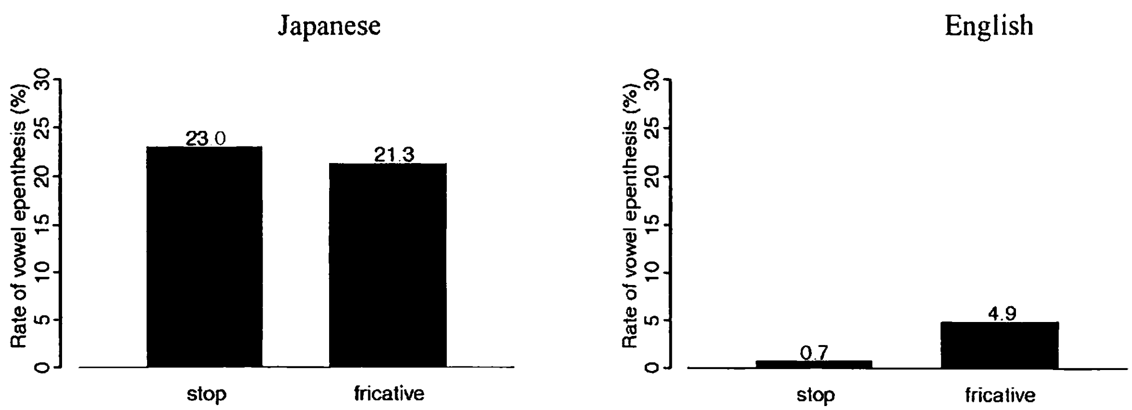


Figure 7. Rate of epenthesis as a function of following consonant's manner. "Stop" data calculated from *accept* /-pt/, *described* /-bd/, *looked* /-kt/, *bragged* /-gd/. "Fricative data calculated from *grapes* /-ps/, *cubes* /-bz/, *six* /-ks/, *legs* /-gz/.

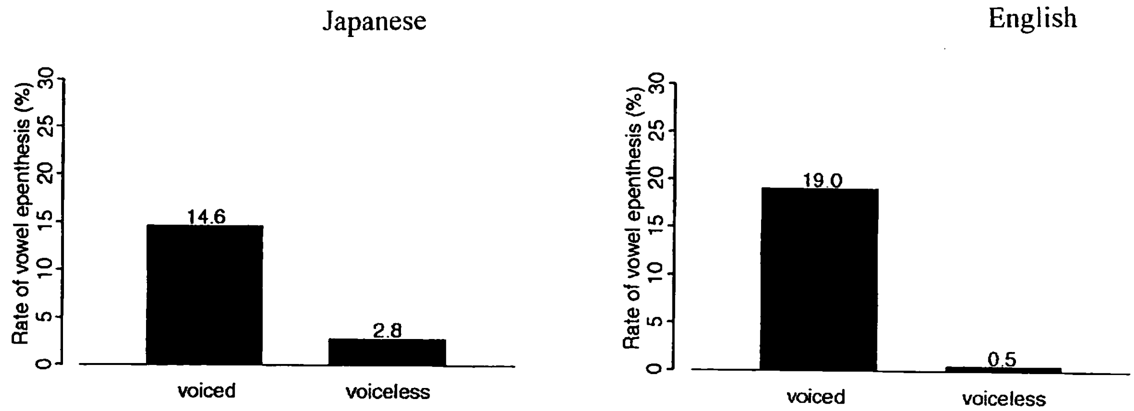


Figure 8. Rate of word-final epenthesis as a function voicing. "Voiced" data calculated from *raindrop* /-p/, *admit* /-t/, *sneak* /-k/, *stretch* /-ʧ/, *photograph* /-f/, *actress* /-s/. "Voiceless" data calculated from *tribe* /-b/, skid /-d/, clog /-g/, oblige /-ʤ/, *thrive* /-v/, and spies /-z/.

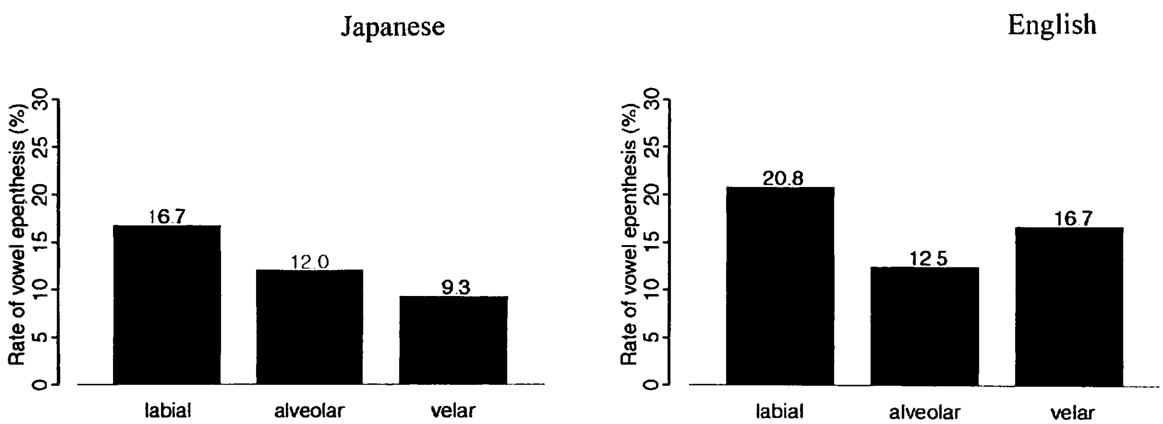


Figure 9. Rate of epenthesis as a function of consonant place. "Labial" data calculated from *raindrop* /-p/, *tribe* /-b/. "Alveolar" data calculated from *admit* /-t/, *skid* /-d/. "Velar" data calculated from *sneak* /-k/, *clog* /-g/.

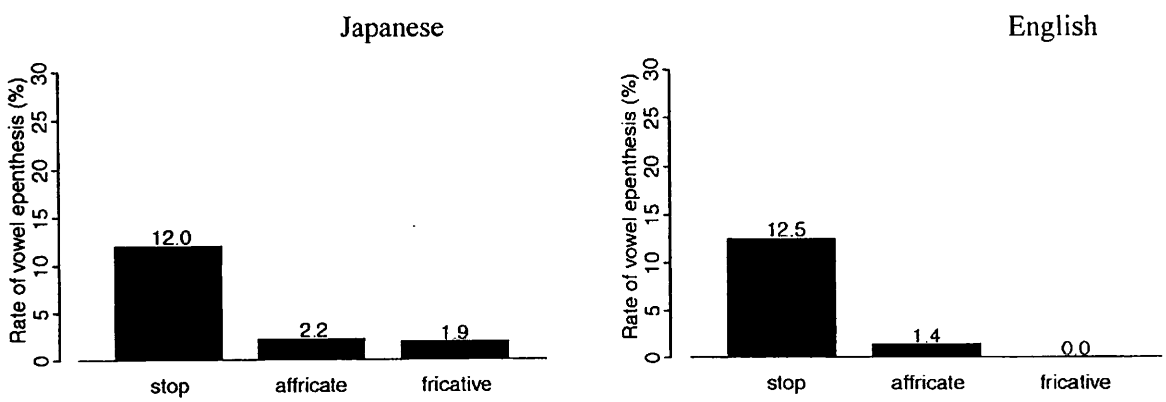


Figure 10. Rate of epenthesis as a function of consonant manner. "Stop" data calculated from *admit* /-t/, *skid* /-d/. "Affricate" data calculated from *stretch* /-ʧ/, oblige /-ʤ/. "Fricative" data calculated from *actress* /-s/, *spies* /-z/.

Figure 10 indicates that stops are more likely followed by epenthetic vowels than are affricates or fricatives, in both native Japanese and native English speakers' productions. The effect of Manner was found to be significant for the Japanese data [$F(2,50) = 9.6$; $p < .001$]. Tukey's HSD test found the following pattern of significant differences at the $p = .01$ level: stops > affricates = fricatives, confirming Hypothesis 6. The English data did not show a significant effect of Manner [$F(2,10) = 4.1$; n.s.], even though the magnitude of the effect was comparable to that of the Japanese data. This result, together with the result from Figure 7, suggest that manner of the preceding consonant, but not the following consonant, has an effect on the rate of occurrence of epenthetic vowels.

In sum, data from Figures 5-10 suggest that native Japanese speakers' tendency to produce epenthetic vowels is affected by phonetic environment, namely, voicing of the surrounding consonants, place of articulation of the preceding stop, and manner of articulation of the preceding consonant. The data also reveal that epenthetic vowels occasionally witnessed in native English speakers' productions are also subject to some of these phonetic effects.

| Hypothesis | Description | Result |
|---|---|---|
| ---- *Word familiarity / loanword frequency (Section 3.2)* ---- | | |
| Hypothesis 1 | Word familiarity | Confirmed (small trend) |
| Hypothesis 2a,2b | Presence/frequency of loanwords | Inconclusive |
| | | |
| ----*Speaking condition (Section 3.3)* ---- | | |
| Hypothesis 3 | Speaking rate | Confirmed |
| Hypothesis 4a,4b | Word vs. sentence production | Neither confirmed |
| | | |
| ---- *Phonetic environment (Section 3.4)* ---- | | |
| Hypothesis 5 | Voicing | Confirmed |
| Hypothesis 6 | Manner | Confirmed |
| Hypothesis 7a,7b | Place | 7a confirmed |

Table 1. Summary of the hypotheses concerning the effect of various factors on rate of occurrence of epenthetic vowels, and their outcomes.

## 4 Discussion and Conclusion

To sum up, this study provided an empirical investigation of the occurrence of epenthetic vowels in native Japanese speakers' production of English words. Auditory and spectrographic analysis by several phonetically trained judges confirmed that epenthetic vowels occur frequently within consonant clusters and following word-final consonants. The seven hypotheses in Section 1.2 concerning various factors that might influence the rate of occurrence of epenthetic vowels were each examined. Table 1 provides a summary of the results.

When all the factors influencing vowel epenthesis are collectively examined, it appears that voicing of the neighboring consonants has by far the strongest effect. The ratio between rate of epenthesis in voiced vs. voiceless environments is 6.7:1 in Figure 5 and 5.2:1 in Figure 8 for the Japanese speakers' productions. The effect of other factors such as consonant place and manner, speaking rate, and word familiarity seems to be smaller in magnitude. It is possible that the voicing effect is so strong that it overshadows the effect of other factors. If so, then partialing out the voicing effect could reduce variability, and could provide a clearer picture of the extent to which other factors affect vowel epenthesis. Nevertheless, this study suggests that, aside from voicing, factors such as speaking rate and word familiarity exert some influence on the results, suggesting the need to control these variables when designing methods to evaluate and train L2 learners' ability to accurately produce L2 syllable structure.

It is noteworthy that epenthetic, non-lexical vowels were identified not just in native Japanese speakers' productions of English words, but also in native English speakers' productions. Finding epenthetic vowels in the native-speaker productions is in itself not surprising, since it is not uncommon for natives to release, for example, word-final voiced stops, e.g., *bad* pronounced as /bæd/, in certain speaking styles, resulting in a non-lexical vowel-like segment word-finally. This opens the question as to whether or not epenthetic vowels produced by native English speakers are equivalent to epenthetic vowels produced by L2 learners in terms of their acoustic-phonetic characteristics and their perceptual consequences. Resolving this issue requires additional research. In addition, it is also worth noting that some of the factors examined had comparable effects in both Japanese and English, e.g., the effect of voicing (Figure 8) and manner (Figure 10), and that it is possible for these effects to be understood in terms of aerodynamic and articulatory tendencies. In other words, the insertion of epenthetic vowels in Japanese-accented English likely stems from language-specific structural differences between English syllables and Japanese syllables, but exactly where and how often epenthetic vowels are actually produced appear to be subject to language-general phonetic principles.

Finally, it should be reiterated that distortion of syllable structure by L2 learners is a problem that is found not just in production, but also in perception. Studies have demonstrated that native Japanese speakers have reduced ability to count syllables in spoken English words (Erickson et al. 1999) and reduced ability to discriminate stimuli such as /ebuzo/ and /ebzo/ which violates native-language syllable structure constraints (Dupoux et al. 1999). These results suggest that native Japanese speakers have difficulty in both production and perception of English syllables. Additional research is needed to examine how closely the production and perception difficulties match. Research is also necessary for developing techniques to train L2 learners' ability to cope with L2 syllable structure.

## References


Amano, S. and Kondo, T. (2000). *Nihongo-no Goi-Tokusei (Lexical Properties of Japanese* ,7, Sanseido, Tokyo).

Blevins, J. (1995). The syllable in phonological theory. In Goldsmith, J. A. (ed.), *The Handbook of Phonological Theory*, pp. 206 - 244. Cambridge, MA: Blackwell Publishers.

Byrd, D. (1993). 54,000 American Stops. *UCLA Working Papers in Phonetics, 83.*

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1568-1578.

Erickson, D., Akahane-Yamada, R., Tajima, K., and Matsumoto, K.F. (1999). Syllable counting and mora units in speech perception. *In Proceedings of the XIVth International Congress of Phonetic Sciences*, 1479-1482. San Francisco.

Hancin-Bhatt, B. and Bhatt, R. M. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition, 18*, 341-387.

Katayama. M. (1998). Optimality Theory and Japanese Loanword Phonology. Doctoral dissertation. University of California, Santa Cruz.

Lovins, J. B. (1973). Loanwords and the Phonological Structure of Japanese. Doctoral dissertation. University of Chicago.

Otaka, T. (2000). Listeners' representations of within-word structure: A cross-linguistic study of Japanese and English. Paper presented at the Seventh Conference on Laboratory Phonology, Nijmegen, the Netherlands, June, 2000.

Ross, S. (1994). The ins and outs of paragoge and apocope in Japanese English interphonology. 10: 1-24.

Saunders, N. J. (1987). Morphophonemic variation in clusters in Japanese English. *Language Learning* 37: 247-272.

Tajima, K., Erickson, D., and Akahane-Yamada, R. (2000). Training Japanese to perceive syllables in spoken English. In *Proceedings of the 2000 Spring Meeting of the Acoustical Society of Japan*, Chiba, Japan, 347-348.

Tajima, K., Port, R. F., and Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25: 1-24.

Zue, V. W., Seneff, S. and Glass, J. (1990). Speech database development at MIT: TIMIT and beyond. *Speech Communication*, 9, 351 - 356.