JOURNAL ARTICLE

Epenthetic vowel production of unfamiliar medial consonant clusters by Japanese speakers

Wakayo Mattingley[1], Kathleen Currie Hall[2] and Elizabeth Hume[3]

[1] Department of Linguistics, University of Canterbury, NZ

[2] Department of Linguistics, University of British Columbia, CA

[3] Office of Academic Affairs, The Ohio State University, US

Corresponding author: Wakayo Mattingley (wakayo.mattingley@pg.canterbury.ac.nz)

Existing nativized loanword studies have traditionally suggested that there are three epenthetic vowels in Japanese, which reflect both phonotactic restrictions and articulatory properties of certain consonant-vowel sequences in the language. Recent findings, however, call this tri-partite epenthesis pattern into question: First, several studies suggest that this epenthesis pattern is not true in the realm of perception and is not completely regular in production, and second, the relevant phonotactic restrictions seem to be weakening even outside of epenthesis contexts. This paper therefore investigates the extent to which the spontaneous choice of epenthetic vowels in the *production* of Japanese conforms to the traditional tri-partite pattern. Epenthesis was induced by presenting pseudo-word stimuli of the form of [aCCa] (C = a voiced consonant) to subjects orthographically. The findings suggest that indeed, the production pattern does not fully conform to what is generally reported for nativized loanwords; in particular, the traditionally "default" vowel [ɯ] is used by our participants frequently in all contexts, including the two where [o] or [i] is usually reported. That said, we also show that there is considerable variability across speakers as to which vowel is epenthesized, especially in the palatal context, and this variability includes tokens of vowels similar to all possible lexical vowels of Japanese.

Keywords: vowel epenthesis; loanword adaptation; phonotactics; Japanese

1. Introduction

Vowel epenthesis is a common repair strategy used by speakers to adapt loanwords that contain unfamiliar phonological structure in the borrowing language (e.g., Davidson, 2006; Fleischhacker, 2001; Hall, 2011; Kabak & Idsardi, 2007; Kang, 2011; Uffmann, 2006). Such is the case in Japanese where vowel epenthesis serves to make non-native structures more native-like (e.g., Hirayama, 2003; Itô, 1989; Smith, 2006; Kubozono, 2015). For example, the English word 'pipe' [paɪp] is commonly pronounced as [paɪpɯ] with [ɯ] occurring in word-final position, as consonants other than [N] do not occur word-finally in Japanese (Kubozono, 2015). The adaptation of unfamiliar consonant sequences by epenthesis in Japanese has served as a test case for studying the influence of native speech experience on speech perception and production (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Monahan, Takahashi, Nakao, & Idsardi, 2009; Peperkamp & Dupoux, 2003; Shoji & Shoji, 2014; Sperbeck, 2012; Yazawa, Konishi, Hanzawa, Short, & Kondo, 2015). For example, Dupoux et al. (1999) found that native Japanese listeners perceive an illusory vowel, [ɯ],[1] between sequences of consonants that are illicit in Japanese. This finding

---

[1] Dupoux et al. (1999) use an epenthetic [u] in transcription.

is consistent with research showing that speech perception is constrained by a listener's phonotactic knowledge; that is, non-native sound sequences are generally assimilated perceptually to licit sequences in the listener's native language (e.g., Best, 1994, 1995; Best & Strange, 1992; Dupoux et al., 2001; Dupoux et al., 2011; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak, 2003).

Japanese epenthesis has also received attention in the literature because three separate epenthetic vowels have been observed, [i, o, ɯ], with the choice among them being dependent on the quality of the preceding consonant (e.g., Hirayama, 2003; Irwin, 2011; Katayama, 1998; Kubozono, 2001; Lovins, 1975; Otaki, 2012). In nativized loanwords, the vowel [ɯ] has been shown to occur after labial, alveolar (except alveolar stops), and velar[2] consonants and is generally considered to be the default epenthetic vowel (Hirayama, 2003; Shoji & Shoji, 2014; Kubozono, 2015). Meanwhile, the vowel [i] is epenthesized after the palatal affricates [tɕ] and [dʑ], and [o] is used after the alveolar stops [t] and [d] (e.g., Kaneko, 2006; Kubozono, 2015).

This basic pattern of the distribution of the epenthetic vowel is corroborated by the findings of Yazawa et al. (2015). They investigated whether patterns of English speech production by Japanese learners of English are similar to the phonology of loanword epenthesis in Japanese, considered in relation to the level of English proficiency of the speakers. They analyzed speech corpus data of Japanese participants reading the Aesop fable "The North Wind and the Sun" in English, recorded in 2009 as part of the J-AESOP (Asian English Speech cOrpus Project) Corpus. They used a combination of an automatic annotator and manual/visual inspection of the formant space in comparison to 'typical' lexical vowels of Japanese to identify the quality of the epenthetic vowels. Their results showed that irrespective of learners' proficiency level, the quality of epenthetic vowels is similar to the patterns in loanword phonology. That is, an epenthetic vowel has a quality close to [o] after [t] and [d], [i] after [tɕ] and [dʑ], and [ɯ] when it follows any other consonant.

Typically, the explanation for this basic distributional pattern has relied on a combination of perceptual and phonotactic observations. The occurrence of [ɯ] as the 'default' epenthetic vowel is consistent with the view that the epenthetic vowel is the perceptually least salient in the language (Byarushengo, 1976; Fleischhacker, 2001; Kang, 2003; Kenstowicz, 2007; Shinohara, 1997; Steriade, 2001, 2008), as [ɯ] is considered to be the shortest vowel and the most susceptible to weakening and deletion in Japanese, common properties of perceptually weak vowels (e.g., Hirayama, 2003; Sagisaka & Tokuhara, 1984 as cited in Irwin, 2011; Shoji & Shoji, 2014; Kubozono, 2015). Meanwhile, the use of [i] after palatals is arguably because the front vowel [i] shares similar articulatory and perceptual properties with these consonants (Kubozono, 2015). Finally, the use of [o] after alveolar stops is presumably because neither of the other two vowels ([ɯ] or [i]) is phonotactically licit in this position in native Japanese (i.e., *[tɯ], *[dɯ], *[ti], and *[di]) (Kaneko, 2006; Kubozono, 2015). Kubozono (2015) suggests that the choice of [o] is also associated with perceptual properties; inserting [o] after alveolar stops keeps the original consonants, while inserting [ɯ] after alveolar stops could result in the alveolars changing to affricates [ts] and [dz] due to an allophonic rule in Japanese (Hirayama, 2003; Irwin, 2011; Kubozono, 2015).

There are, however, several potential problems with these explanations of the choice of epenthetic vowel. First, recent perceptual epenthesis studies suggest that the distribution of epenthetic vowels in *perception* differs from that observed in nativized loanwords and the Yazawa et al. (2015) corpus study. Mattingley, Hume, and Hall (2015) asked

---

[2] After the voiceless velar [k], [i] appears in some older loanwords, e.g., 'cake' is pronounced as [keiki], though more recent loanwords take [ɯ].

Japanese-speaking listeners to identify what vowel, if any, occurred between the two Cs in VCCV forms, where the place of articulation of the first C varied. In the labial, alveolar stop, and velar contexts, they perceived an epenthetic vowel between the two Cs 60–70% of the time, and in the palatal context, they perceived an epenthetic vowel 98% of the time. The choice of epenthetic vowel, however, did not always match the distribution described above. In the palatal context and the velar context, the perceptual epenthetic vowels were largely as expected; of the tokens where an epenthetic vowel was perceived at all, 94% were the expected [i] in the palatal context and 94% were the expected [ɯ] in the velar context. In the labial context, however, only 84% of perceived vowels were the expected [ɯ], with the rest being fairly equally distributed among [a], [e], and [i]. In the alveolar stop context, the discrepancy was even more extreme, with only 16% of perceived vowels being the expected [o], and 71% being [ɯ] instead, even though *[dɯ] is an illicit phonotactic sequence in native Japanese. It should be noted that there was no control group for language in this study; thus it is possible that these perceptual effects could be driven not by Japanese-specific patterns but rather by more general acoustic characteristics of the stimuli. Interestingly, Monahan et al. (2009) found that while Japanese-speaking listeners do perceive an illusory [ɯ] in velar VCCV contexts, they were able to discriminate alveolar VCCV sequences from similar sequences with either a medial [ɯ] or [o], suggesting that they perceived these as VCCV sequences without an epenthetic vowel (they did not test labial or palatal contexts).

There is also evidence from a different type of production study than that used in Yazawa et al. (2015) that suggests that the productive choice of epenthetic vowel may be different from that in lexicalized loanwords. Shoji and Shoji (2014) used a writing production experiment to examine patterns of vowel epenthesis in hypothetical loanwords from nonce words spelled in orthographic Latin script; native Japanese speakers transcribed the nonce words in Japanese characters, which forced them to either delete or epenthesize in every case. They found that in palatal contexts, where [i] might be expected, [i] occurred only 34% of the time in word-initial clusters, with [ɯ] occurring another 23.7% of the time; in word-final clusters, [i] occurred 85.6% of the time, while [ɯ] occurred 12.2% of the time. They also found that in alveolar-stop contexts, where [o] might be expected, [o] occurred only 45.6% of the time and [ɯ] occurred 32.32% of the time in word-initial clusters; [o] occurred 91.1% of the time and [ɯ] 5.6% of the time in word-final clusters. In velar contexts, where [ɯ] might be expected, [ɯ] did in fact occur 71% of the time in word-initial clusters, while [i] was used 15.6% of the time, and [ɯ] was used 95.6% of the time in word-final clusters. They do not report the results of labial contexts separately, but do find the expected [ɯ] used in their 'other' contexts, which included labials, >90% of the time in both word-initial and word-final clusters. Thus, while the traditionally expected vowels were the most frequent in any given context, in several contexts, they accounted for fewer than half of the actual tokens.

Third, there is separate evidence that the phonotactic constraints governing loanwords are changing in non-epenthetic contexts. Pintér (2015, p. 121–122) points out that while older loanwords with /ti/ sequences were typically adapted as [tɕi] (e.g., 'team' adapted as [tɕi:mɯ]), more recent loanwords are adapted more faithfully with [ti] (e.g., 'party' adapted as [pa:ti:]). Kubozono (2015, p. 325) notes that a single loanword can also have multiple, age-related adaptations (e.g., 'tissue' being adapted as either [tʃitʃʃɯ] or [tetʃʃɯ] by older speakers but as [tiʃʃɯ] by younger speakers). That said, [t] and [d] have been observed to occur before [i] in at least some loanwords since at least 1950 (Bloch, 1950 cites examples like *vanity case* [vaniti] and *caddy* [kjadi:]). A similar change is reported for /tu/ sequences; despite the traditional constraint against [tɯ] in Japanese, Pintér (2015) reports that this sequence is in fact possible in more recent loanwords, e.g., 'Bantu'

adapted as [bantɯː]. As Pintér (2008, p. 112) also points out, even the official stance of the Japanese National Language Committee changed between 1954 and 1991; in 1954, [tɯ] was not acknowledged as a possible written syllable of Japanese, but in 1991, it was acknowledged though not officially supported. Thus, the phonotactic motivations for epenthesizing only [o] after alveolars and [i] after only palatals may be eroding. Indeed, an examination of the Balanced Corpus of Contemporary Written Japanese (National Institute for Japanese Language and Linguistics, 2011) indicates that all five lexical vowels can occur after both [d] and [dʑ] in loanwords (defined as words of foreign origin other than Chinese); see also discussion in Hall (2009, 2013) showing that loanwords are eroding the predictability of several pairs of Japanese phonemes.

These results combine to raise questions about the current state of epenthesis in Japanese. Specifically, it would appear that the use of [o] after alveolar stops and [i] after palatals, i.e., the typical pattern in loanword epenthesis, is not currently a clear-cut pattern in perception, production, and other areas of loanword adaptation. This suggests that either perception and production factors are not the explanatory causes of the lexicalized loanword epenthesis patterns or that there may similarly now be a change in the epenthesis adaptation patterns themselves. The current study is designed to be a first step in understanding the current state of loanword epenthesis patterns from a production perspective. It involves a production task that tests the full range of epenthetic vowels used to break up VCCV sequences in Japanese; in order to maintain maximal control over the stimuli, however, these sequences are simply presented as nonce words, rather than being loanwords in any real sense (i.e., they are not associated with meaning or claimed to come from a particular foreign source language); we note, however that it has generally been claimed that loanwords and nonce words in Japanese tend to follow similar grammatical patterns. As Kawahara (2012, p. 1194) points out, "both loanwords and nonce words show default accentuation patterns in Japanese (e.g., Katayama, 1998; Kawahara & Kao, 2012; Kubozono, 1996, 2006, 2008; Labrune, 2012; McCawley, 1968; Shinohara, 2000), [and] neither native words nor Sino-Japanese words allow voiced geminates, while both loanwords and nonce words allow them (Itô & Mester, 1995, 1999, 2008)." Kawahara goes on to demonstrate that Japanese speakers judge both loanwords and nonce words as being more natural when they follow Lyman's Law, even though Lyman's Law does not hold categorically outside of the native vocabulary. Thus, we have reason to believe that nonce words might be a reasonable proxy for loanwords.

To anticipate the results, some patterns were consistent with earlier studies, while there was considerable individual speaker variability in other cases. In particular, [ɯ] was consistently used by all 14 participants after both labials and velars, which is in line with reported patterns of loanword adaptation. However, after alveolar stops, only two participants used the expected [o], with the other 12 speakers either using [ɯ] or using some combination of both [o] and [ɯ], suggesting that the loosening of the phonotactic constraint against [dɯ] is possibly extending to epenthetic contexts. Perhaps most surprisingly, the palatal context induced a large range of epenthetic vowels; again, only two speakers consistently used the expected vowel, [i], in this context, while four used [ɯ], and the rest used some combination of almost every lexical vowel, including [e] and [a], which are not generally reported as being used epenthetically at all.

1.1. Background on Japanese

Before describing the details of the production experiment, we note the following relevant information regarding the Japanese phonological system. Modern Japanese has five phonemic vowel qualities, as shown in **Table 1** (e.g., Akamatsu, 2000; Shibatani, 1990;

Table 1: Vowel of Japanese. Adapted from Vance (2008).

	Front	Central	Back
Close	i, i:		ɯ, ɯ:
Mid	e, e:		o, o:
Open		a, a:	

Table 2: Consonants of Japanese. Adapted from Akamatsu (2000), Vance (2008).

	Bilabial		Alveolar		Alveolo-Palatal		Palatal	Velar		Uvular	Glottal
Plosive	p	b	t	d				k	g		
Nasal		m		n				(ŋ)		ɴ	
Fricative		(ɸ)	s	z	(ɕ)	(ʑ)	(ç)				h
Affricate			(ts)	(dz)	(tɕ)	(dʑ)					
Approximant						j			ɰ		
Liquid						ɾ					

Tsujimura, 1996; Vance, 1987, 2008). All vowels have a phonemic length contrast. In terms of vowel duration, [ɯ] is the shortest vowel in Japanese while [a] is the longest (Campbell, 1992; Han 1962, cited in Shoji & Shoji, 2014; Yoshida, 2006). Further, among the five vowels, the high back vowel [ɯ] is the least likely vowel to be accented (Yoshida, 2006).

   Table 2 presents the traditional consonantal phonemes of native Japanese; a conservative view of common allophones appear in parentheses. The alveolar consonants /t/, /d/, /s/, /z/, and the glottal fricative /h/ are palatalized to [tɕ], [dʑ], [ɕ], [ʑ], and [ç], respectively, when they occur before the high vowel /i/. Alveolar /t/, /d/, and glottal /h/ are also realized as [ts], [dz], and [ɸ], respectively, when they are followed by the high back vowel /ɯ/.

   It should be noted that this conservative analysis for Japanese allophonic status fails for Sino-Japanese and, especially crucial for the current study, loanwords. Pintér (2015, p. 125), for example, claims that the "innovative variety [of Japanese] … accommodates (almost) all logically possible CV combinations," including sequences like [ti], [di], [tɯ], and [dɯ], as mentioned above, and while he does not take a firm stance on the appropriate phonological representations of these sounds, he does treat the innovative forms as emergent contrasts, suggesting that they are not simply contextually predictable allophones.

   Japanese syllable structure is relatively simple, usually consisting of a consonant-vowel (CV) or vowel (V) sequence. Syllables are maximally CVC, but only a nasal or the first part of a geminate consonant is allowed in coda position (Tsujimura, 1996); e.g., [sim.bɯɴ] (CVC.CVC) 'newspaper,' [gak.koo] (CVC.CVV) 'school.' Otherwise, consonant clusters are illicit in word initial, medial, and final positions.

2. Methodology

In order to more thoroughly investigate the nature of epenthetic vowels in Japanese, a production study was carried out. Native speakers of Japanese were asked to produce nonsense words that were likely to trigger epenthesis across a variety of consonantal environments, and then acoustic analyses were conducted to determine the identity of the epenthetic vowel in each context for each speaker.

2.1. Speakers

Fourteen native speakers of Japanese[3] (10 female, 4 male) participated in the production experiment, conducted at University of Canterbury, in New Zealand. Participants were recruited from local English language schools via posted fliers at the schools, and were compensated with a $20 voucher. Participant age ranged from 21 to 46 (mean = 27.3). All participants had lived in an English-speaking country for less than one year, and were on a working holiday or studying English. No participants reported any speech or hearing disorders. They had all received English language education for six years in junior high and high school in Japan, since English is compulsory from age 12. Their total years living in foreign countries including non-English-speaking countries was less than three years.

2.2. Materials

There were two types of stimuli created, those for a control condition and those for an experimental condition. The control condition was intended to elicit natural examples of each speaker's regular production of each of the five Japanese lexical vowels, in an inter-consonantal context. The experimental condition was intended to elicit natural, spontaneous examples of epenthetic vowels. The structure of the pseudo-words for the control condition was $[aC_1VC_2a]$ where V was one of the five Japanese vowel qualities {a, e, i, o, ɯ}. The structure of the pseudo-words for the experimental condition was $[aC_1C_2a]$. In both conditions, consonants were selected from the set of voiced obstruents {b, d, g, d͡ʑ}, and C1 ≠ C2 (e.g., [b…d], [b…g], [b…d͡ʑ]).[4] The initial and final vowels of pseudo-words were always [a] in order to maintain uniformity across all stimuli. There were 60 control items with the form $[aC_1VC_2a]$ (12 consonant combinations * 5 vowels) and 12 experimental items of the form $[aC_1C_2a]$ (12 consonant combinations), for a total of 72 items with a voiced obstruent environment. Additionally, 24 pseudo-word fillers of the structure [VCCa] were included; in these, the initial V was never [a] and the C2 could include the nasals {m,n} in addition to {b, d, g, d͡ʑ}, for a total of 96 items. The fillers were intended to increase the variety of produced items in order to minimize participants' recognition of patterns in the stimuli. The 60 control items were repeated twice, while the 12 experimental items and 24 filler items were each repeated three times, to create a total of 228 trials. These trials were then divided evenly across two sessions, as shown in **Table 3**. A full list of production stimuli is given in Appendix A. The total possible number of epenthetic vowels for each speaker was 36 (12 experimental stimuli * 3 repetitions). For the vowels in the phonotactically licit control stimuli, {a, e, i, o, ɯ}, there were 24 instances of each (12 voiced consonantal environments * 2 repetitions) during the two sessions. Thus, there were 156 tokens of interest for each speaker. It should be noted,

Table 3: Example session schedule in production experiment.

Session	Name of list	Items			# of Trials
		Control	Experimental	Fillers	
A	List 1 Voiced	60	18	36	114
B	List 2 Voiced	60	18	36	114
Total		120	36	72	228

---

[3] Five additional participants were excluded either because they were highly bilingual or because they mis-read the stimuli.

[4] Following Mattingley et al. (2015), the flanking consonants were chosen to be voiced obstruents in order to avoid potential challenges that could arise in analyzing vowels between voiceless consonants since they tend to become devoiced in Japanese (Vance, 2008; Shaw & Kawahara, 2018).

however, that not all items were always produced as intended. Thus, these numbers reflect the *maximum possible* number of tokens per speaker.

2.3. Procedure

Stimuli were randomized and presented to each participant using E-prime software (Schneider, Eschman, & Zuccolotto, 2012). Each pseudo-word (e.g., *aguba*) was represented in Roman orthography (Hepburn system) and appeared in the following carrier sentence, presented in Japanese characters, e.g., *Kore mo* aguba *desu* "This is aguba, too" (**Figure 1**). The carrier sentence was in Japanese to help encourage participants to produce the stimuli using their Japanese phonology. A Tascam HD-P2 audio recorder with 44,100 samples/s, 16 bit/s, and a Beyer dynamic head-mounted microphone were used for recording, with speakers recorded individually in a sound-attenuated room at the University of Canterbury.

Note that we chose to present the stimuli orthographically rather than auditorily to avoid any additional interference from perception and to reduce bias on the part of the participants toward any particular epenthetic vowel (or lack thereof) based on clues from the stimulus. Smith (2006), for example, shows that Japanese loanwords may in fact have 'doublet' adaptations, one with epenthesis and one with deletion (e.g., 'Hepburn' as either [hep.pɯ.baːɴ] or [he.boɴ]), and argues that the deletion cases likely arise from perceptual factors while the epenthesis cases are more likely influenced by orthographic factors. Given our interest in epenthesis here, orthographic stimuli seemed preferable, though we acknowledge that this choice can have consequences for the results of loanword studies (e.g., Vendelin & Peperkamp, 2006). We discuss this matter further in Section 4.

The procedure was described to participants in Japanese. After seeing a stimulus on the computer screen, they were asked to say the whole sentence, including pronouncing the stimulus item as if it were a Japanese word.[5] If participants thought that they had misread an item, they were able to pronounce it one more time. The participant then pressed any key on the keyboard to display the next stimulus. Each participant produced a randomized list of 228 items during each session.
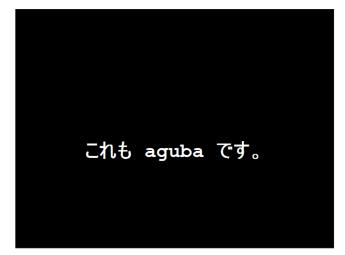

Figure 1: Example slide presented to speakers in the experiment.

---

[5] In retrospect, perhaps it would have made more sense to introduce the stimuli explicitly as 'loanwords' to be borrowed into Japanese, rather than as nonce Japanese words, given the research question. At the same time, the motivations given for the choice of epenthetic vowel are largely based in assumptions about native Japanese phonotactics and production, so presenting the materials this way should have maximized the use of Japanese phonology. Given that, the results, only some of which match traditional descriptions of Japanese epenthesis patterns, are somewhat more compelling.

2.4. Acoustic measurements

As mentioned above, not all of the produced items matched the intended targets. When participants did not produce a phonologically expected vowel or consonant, the token was excluded (decided upon both auditorily and acoustically) (e.g., [adɯba] misread as [abɯba], or [adaga] misread as [adoga]). For vowels, one of the authors manually checked values of formant frequency in words and compared the values with other vowels. **Table 4** summarizes where the discrepancies occurred, broken down by vowel and consonantal context. Recall that for each combination of lexical vowel and C1, there could have been six tokens per speaker, for a maximum of 84 tokens across the 14 speakers. For the epenthetic vowels (symbolized with a plain V in **Table 4** and subsequently), there could have been nine tokens per speaker, for a maximum of 126 tokens across the 14 speakers. As can be seen in **Table 4**, of the total possible 2184 tokens, the production task resulted in a 1971 recorded lexical and epenthetic vowel tokens that could be analyzed. Note that participants in fact inserted epenthetic vowels between two consonants in 100% of the experimental tokens where they were expected. However, only 463 of the 504 epenthetic stimuli were included, because the accompanying consonants were not always produced as expected (e.g., [abda] misread as [adVda]).

Figure 2 illustrates examples of both a lexical (2a) and an epenthetic (2b) vowel in the context [ad_ba] as produced by participant M2; in both cases, the vowel quality was [ɯ]. The total word duration for the word with lexical [ɯ] is 0.33 s, while that for the word with epenthetic [ɯ] is 0.31 s.

The duration of all words and vowel tokens was measured, and the values for F1, F2, and F3 were extracted using Praat (Boersma & Weenink, 2014). Formant measurements were taken at the midpoint of the relevant vowel. While the focus of the analysis is on the quality of epenthetic vowels in [aCCa] sequences, vowels from the control pseudo-words were used for comparison, e.g., [aCVCa]. Specifically, the quality of a given epenthetic vowel (V) was determined by comparing it acoustically to the baseline vowels produced by each speaker. All vowel plots given below show normalized mean F1 and F2 values, with data ellipses enclosing 95% of the data for each lexical vowel [i, e, ɯ, o, a] and the epenthetic vowel (V), using the stat_ellipse() function in the ggplot2 package (Wickham, 2016) of R (R Core Team, 2017), which in turn is based on the dataEllipse() function in the car package (Fox & Weisberg, 2011). The formant values were z-score normalized using the Lobanov normalization procedure in NORM (Thomas & Kendall, 2007) to remove overall effects of speaker sex. All original formant values were measured in Hz.

Table 4: Lexical vowel and epenthetic vowel (V) tokens by preceding consonant. The maximum possible number of tokens of each lexical vowel in each context is 84; the maximum possible number of epenthetic vowels in each context is 126.

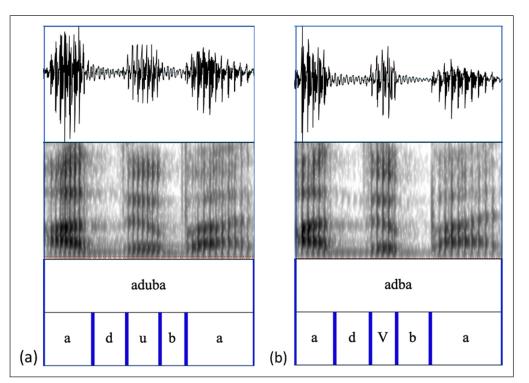| C_1	Vowel						Total
	a	**e**	**i**	**o**	**ɯ**	**V**	
[b]	77	80	78	74	76	107	**492**
[d]	76	74	78	81	67	114	**490**
[g]	79	52	61	84	77	121	**474**
[dʑ]	79	79	75	79	82	121	**515**
Total	**311**	**285**	**292**	**318**	**302**	**463**	**1971**

Figure 2: Examples of a production of **(a)** lexical [ɯ] and **(b)** epenthetic [ɯ] by speaker M2 in the [ad__ba] context.

2.5. Statistical analyses

All data were analyzed using R (R Core Team, 2017). Linear mixed-effects models were created using the lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) packages, in which the normalized F1 value, the normalized F2 value, or the vowel duration was predicted from vowel quality for each of the preceding consonantal contexts, across separate analyses. The random intercepts in the analyses were Speaker and Word, and the fixed effect was Vowel. In all cases, the question of interest is whether the epenthetic vowel is similar to any particular lexical vowel; hence, the epenthetic vowel was always set to be the baseline value.

3. Results

3.1. Overall acoustic characteristics of the production tokens

3.1.1. Characteristics of the control (lexical) vowels

We start by considering the acoustic characteristics of the vowels in the control condition, to establish a baseline of the vowel characteristics against which the epenthetic vowels can be compared. First, we consider vowel duration. **Table 5** shows that the ranking of vowels produced by speakers, from longest to shortest mean duration, is [a], [e], [o], [i], and [ɯ], consistent with earlier vowel duration studies (Campbell, 1992; Han, 1962, cited in Shoji & Shoji, 2014). The standard deviation of vowel durations ranges from 17 to 19 ms, and is consistent across vowels. Overall, [i] and [ɯ] are the shortest vowels, approximately the same as each other, and shorter than all of the others. An ANOVA on a linear mixed effect model with random slopes for speaker indicates that there is a significant effect of vowel quality on duration [$F(4, 52) = 78.471, p < 2.2e\text{-}16$]. A series of subsequent post-hoc t-tests indicate that [i] and [ɯ] are each significantly shorter than either [a] or [e] ($p < 0.02$), but that there are no other significant differences in vowel durations.

Table 5: Mean durations and standard deviations (*SD*), in milliseconds, for each vowel, from 14 speaker averages.

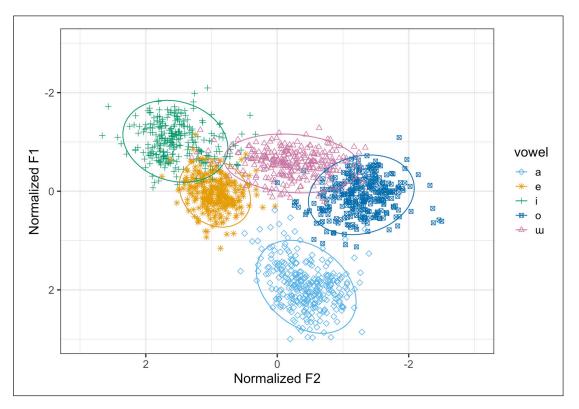|	Vowel Duration (ms)	
	Mean	*SD*
[a]	104	19
[e]	98	18
[o]	93	19
[i]	81	17
[ɯ]	81	19

Figure 3: Lobanov normalized mean F1 and F2 values for each lexical vowel from 14 speakers. Ellipses represent 95% data ellipses.

Next, we consider the quality of each vowel. **Figure 3** shows the overall normalized F1/F2 spaces for each lexical vowel across all 14 speakers and all four consonantal contexts, which on the whole are fairly well separated. It can be seen that [i] slightly overlaps with [e], and [ɯ] slightly overlaps with each of [e] and [o]. The high vowel [i] is higher and more fronted than any other vowel. The vowels [e] and [o] are similar in height, and the high vowel [ɯ] and low vowel [a] are similar in backness. Note that both of these latter vowels are actually more central than back. An articulatory study (Nogita, Yamane, & Bird, 2013) reported that the vowel conventionally described as the high back vowel in Japanese is, in fact, a rounded high central vowel [ʉ] in younger speakers. Participants in the current study were mostly under 35 years old, except for two participants in their 40s. Although the vowel [ɯ] is phonetically central, for readers' convenience, we use [ɯ] for this vowel, following the usual convention.

3.2. Epenthetic vowels in the labial context

3.2.1. F1 and F2 analyses

We begin by looking at the quality of the epenthetic vowel in the labial context where, according to the traditional statement of the distributional pattern of epenthesis, we would expect it to be similar to the lexical vowel [ɯ]. The current results largely support that expectation. **Figure 4** shows lexical vowels from [bVC]-forms and epenthetic vowels from [bC]-forms for all speakers; ellipses are 95% data ellipses. For this context, the epenthetic vowel /V/ (represented by black circles) largely overlaps with the lexical vowel /ɯ/ (pink triangles), though both /V/ and /ɯ/ also slightly overlap with the space of /o/ (dark blue crossed squares). The number of tokens of each vowel are: [a] = 77, [e] = 80, [i] = 78, [o] = 74, [ɯ] = 76, and V = 107. As can be seen, a few productions of the epenthetic vowel are outside of the bound of [ɯ]; specifically, there are three tokens that seem to clearly fall in the space of [i] and two that seem to clearly fall in the space of [a]. Overall, however, most of the tokens are clearly in the high-back vowel space, and overlap entirely with the space for [ɯ].

Two linear mixed-effect models, to predict either F1 or F2 values from vowel quality, were fit to the 492 tokens of lexical and epenthetic vowels in the labial context, shown in **Tables 6** (F1) and **7** (F2). The intercept in each case was set to be the epenthetic vowel V. Thus, we can see for each vowel whether the model predicts its formant values to be significantly different from those of the epenthetic vowel. For the F1 model (**Table 6**), each of the other vowels is predicted to have a significantly different F1 value than the epenthetic vowel V, with the notable exception of [ɯ] ($t = -0.919$, $p = 0.379$). The vowels [a], [e], and [o] are predicted to have statistically significant F1 values that are
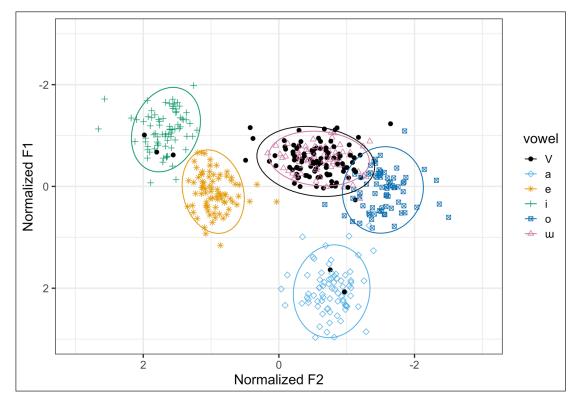

Figure 4: Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel from 14 speakers in the labial context. Ellipses are 95% data ellipses. The number of tokens of each vowel are: [a] = 77, [e] = 80, [i] = 78, [o] = 74, [ɯ] = 76, and V = 107. The epenthetic vowel /V/ largely overlaps with the lexical vowel /ɯ/, though some productions occur outside of the [ɯ] space.

Table 6: Effect estimates and *p*-values on predictors for the F1 value of vowels in the labial
context. Vowels with a significantly different F1 value from the epenthetic vowel V are marked
with asterisks.

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	−0.469	0.045	−10.536	6.73e-06	***
Vowel a	2.52	0.067	37.492	2.96e-12	***
Vowel e	0.580	0.067	8.707	5.19e-06	***
Vowel i	−0.644	0.067	−9.603	2.19e-06	***
Vowel o	0.556	0.068	8.197	6.71e-05	***
Vowel ɯ	−0.062	0.067	−0.919	0.379	

Table 7: Effect estimates and *p*-values on predictors for the F2 value of vowels in the labial
context. Vowels with a significantly different F2 value from the epenthetic vowel V are marked
with asterisks.

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	−0.502	0.115	−4.381	0.0006	***
Vowel a	−0.282	0.156	−1.813	0.094	.
Vowel e	1.443	0.152	9.491	2.82e-07	***
Vowel i	2.195	0.156	14.093	4.90e-09	***
Vowel o	−1.052	0.156	−6.747	1.64e-05	***
Vowel ɯ	−0.102	0.156	−0.657	0.523	

higher than that of the epenthetic vowel (i.e., to be significantly lower vowels), while
[i] is predicted to have a statistically significant F1 value that is lower than that of the
epenthetic vowel (i.e., to be significantly higher). For F2 (**Table 7**), the results also show
a significant effect of most vowels except for [ɯ] (*t* = −0.657, *p* = .523). That is, no
significant difference in formants was found between the epenthetic vowel and [ɯ] in this
context, while the epenthetic vowel is decidedly different from most of the other vowels.[6]
The vowels [i] and [e] are predicted to have significantly higher F2 values than the
epenthetic vowel (i.e., to be significantly more fronted vowels), while [o] (and to some
extent, [a]) is predicted to have a lower F2 value than the epenthetic vowel (i.e., to be a
more backed vowel). In short, the quality of the epenthetic vowel is extremely similar to
[ɯ] in both height and backness, and not particularly similar to any other vowel.

3.2.2. Duration analyses

Figure 5 presents a set of box plots for the vowel durations of the lexical vowels (in
[bVC]-forms) and epenthetic vowels (V in [bC]-forms) for all 14 speakers. The epenthetic
vowel tends to be shorter than the other vowels; in raw terms, then, it is most similar to
[ɯ]. The results of a linear mixed-effect model predicting the vowel duration from vowel
quality are shown in **Table 8**. The intercept was again set to be the epenthetic vowel V.
Thus, we can see for each vowel whether the model predicts its duration to be significantly
different from that of the epenthetic vowel. The results show that [a], [e], and [o] are

---

[6] Note that the difference between the epenthetic vowel and [a] in F2 does not quite reach statistical signifi-
cance assuming an alpha of 0.05 (*t* = −1.813, *p* = 0.094). Looking at Figure 4, this is not surprising, given
that the span of F2 values for the epenthetic vowel encompasses the range of those of [a]. It is clearly the
case, however, that the epenthetic vowel V is not [a], given their significant F1 differences.
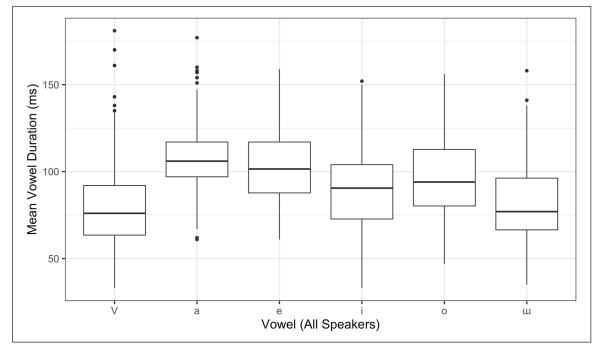
Figure 5: Boxplots of vowel durations for all speakers in the labial context.

Table 8: Effect estimates and *p*-values on predictors for vowel duration in the labial context. Vowels with a significantly different duration value from the epenthetic vowel V are marked with asterisks.

	Estimate	*SE*	*t* value	Pr(>|t|)	
(Intercept)	82.54	5.88	14.03	1.90e-11	***
Vowel a	25.98	4.15	6.27	5.75e-05	***
Vowel e	19.70	4.10	4.80	0.0005	***
Vowel i	8.11	4.14	1.96	0.076	
Vowel o	13.42	4.16	3.22	0.008	**
Vowel ɯ	−0.09	4.15	−0.02	0.98	

predicted to have significantly different values than the epenthetic vowel ($p < 0.05$). The epenthetic vowel V is not quite significantly different from [i] assuming an alpha value of 0.05 ($t = 1.96$, $p = 0.076$), and it is also not significantly different from [ɯ] ($t = −0.02$, $p = 0.98$). Thus, it can be seen that the vowel inserted after a labial consonant is most similar in duration to the shortest vowel, which is [ɯ].

3.3. Epenthetic vowels in the velar context

3.3.1. F1 and F2 Analyses

As is the case in the labial context, we expect from prior descriptions that the quality of the epenthetic vowel in the velar context [g] will be similar to that of the lexical vowel [ɯ], and once again, the current results largely corroborate that expectation. **Figure 6** shows the overall vowel space for the lexical vowels from [gVC]-forms and epenthetic vowels from [gC]-forms for all 14 speakers. The number of tokens are: [a] = 79, [e] = 52,[7]

---

[7] The number of [e] is small compared to other vowels. This is because some participants had a tendency to interpret the <e> as having an effect on the consonants and hence misread the given words (e.g., <ageda> [ageda] was read [adʑeda]).

[i] = 61, [o] = 84, [ɯ] = 77, and V = 121. As can be seen in **Figure 6**, the epenthetic vowel space V (again represented by black circles) is almost entirely overlapping with the vowel space of [ɯ] (pink triangles). Again, however, there are a few productions of the epenthetic vowel in both the [o] space and the [i] space.

As before, linear mixed-effect models predicting F1 or F2 from vowel quality were fit to the 474 tokens of lexical and epenthetic vowels in the velar context. Results are shown in **Tables 9** (F1) and **10** (F2). The intercept is the epenthetic vowel V. For F1 (**Table 9**), each vowel was significantly different from the epenthetic V, except for [ɯ] ($t = -0.255$, $p = .805$). The vowels [a], [e], and [o] are predicted to have significantly higher F1 than the epenthetic vowel, while [i] is predicted to have a significantly lower F1 value
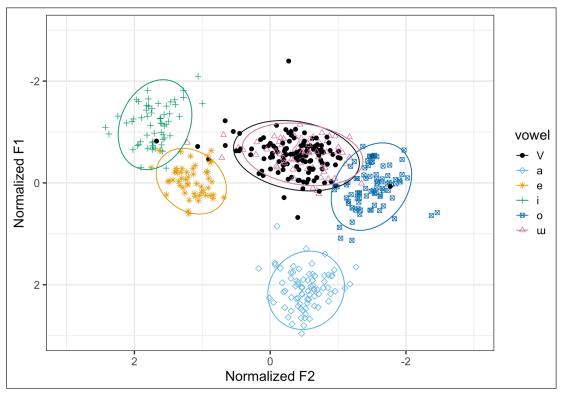

Figure 6: Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel from 14 speakers in the velar context. Ellipses are 95% data ellipses. The number of tokens of each vowel are: [a] = 79, [e] = 52, [i] = 61, [o] = 84, [ɯ] = 77, and V = 121. The epenthetic vowel space V is almost entirely overlapping with the vowel space of [ɯ], though some productions are scattered outside of the [ɯ] boundary.

Table 9: Effect estimates and *p*-values on predictors for the F1 value of vowels in the velar context. Vowels with a significantly different F1 value from the epenthetic vowel V are marked with asterisks.

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	−0.534	0.058	−9.152	3.94e-06	***
Vowel a	2.633	0.079	33.450	2.36e-10	***
Vowel e	0.488	0.084	5.826	0.0001	***
Vowel i	−0.632	0.082	−7.739	1.73e-05	***
Vowel o	0.612	0.078	7.873	3.45e-05	***
Vowel ɯ	−0.020	0.079	−0.255	0.805	

Table 10: Effect estimates and *p*-values on predictors for the F2 value of vowels in the velar context. Vowels with a significantly different F2 value from the epenthetic vowel V are marked with asterisks.

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	−0.348	0.122	−2.851	0.013	*
Vowel a	−0.180	0.170	−1.063	0.308	
Vowel e	1.518	0.171	8.856	7.17e-07	***
Vowel i	2.049	0.171	12.011	2.51e-08	***
Vowel o	−1.159	0.165	−7.019	8.41e-06	***
Vowel ɯ	−0.105	0.170	−0.620	0.546	

Figure 7: Vowel durations in the velar context for all speakers.

than the epenthetic vowel. For F2 (**Table 10**), as in the labial context, the results show a significant effect of each vowel except for [a] ($t = -1.063, p = 0.308$) and [ɯ] ($t = -0.620, p = 0.546$), with [e] and [i] being significantly fronter and [o] being significantly backer. Overall, there is no significant difference between the epenthetic vowel and [ɯ] for F1 and F2 values in the velar context, while the epenthetic vowel is significantly different from all other vowels in at least one dimension.

3.3.2. Duration analyses

Figure 7 presents a set of box plots showing the durations for lexical vowels (in [gVC]-forms) and epenthetic vowels (V in [gC]-forms) for all 14 speakers. Again, the epenthetic V in this context is quite short, and appears to be most similar to [i] or [ɯ]. The linear mixed-effect model results for duration in **Table 11** confirm this observation; the values for [a], [e], and [o] are all significantly different from that of the epenthetic vowel ($p < 0.01$), while [i] ($t = 0.57, p = 0.58$) and [ɯ] ($t = 0.89, p = 0.40$) were not significantly different.

When taken together, the results based on the duration and quality of the epenthetic vowel in the velar context indicate that it is most similar to [ɯ].

Table 11: Effect estimates and *p*-values on predictors for vowel durations in the velar context. Vowels with a significantly different duration value from the epenthetic vowel V are marked with asterisks.

	Estimate	*SE*	*t* value	Pr(>|t|)	
(Intercept)	73.78	6.11	12.08	6.33e-10	***
Vowel a	25.85	3.98	6.49	0.0001	***
Vowel e	26.70	4.23	6.31	6.35e-05	***
Vowel i	2.37	4.12	0.57	0.58	
Vowel o	16.58	3.93	4.22	0.003	**
Vowel ɯ	3.55	4.00	0.89	0.40	

3.4. Epenthetic vowels in the alveolar context

3.4.1. F1 and F2 analyses

Figure 8 presents a vowel plot of both the lexical vowels from [dVC]-forms and the epenthetic vowels from [dC]-forms for all 14 speakers. Recall that based on previous accounts, the epenthetic vowel is predicted to be most similar to [o] in this context. The number of tokens of each vowel are: [a] = 76, [e] = 74, [i] = 78, [o] = 81, [ɯ] = 67, and V = 114.

As can be seen in **Figure 8a**, the vowel space of the epenthetic vowel overlaps substantially with both the high vowel [ɯ] and the mid vowel [o], indicating that there may be variability as to which vowel is epenthesized, a result that is strikingly different from that seen in the labial and velar contexts.

A closer look at the results reveals that much of this variability can be attributed to variability across speakers (see **Figure 8b**). Only two speakers (M1 and M3) had epenthetic vowels that clearly matched the expected pattern of being [o] (see **Figure 9a**). Another seven (F1, F4, F9, F10, F13, F15, and M2) had a clear [ɯ] in this context instead (see **Figure 9b**). Finally, five speakers (F5, F7, F12, F16, and M4) had epenthetic vowels that either alternated between the two categories or spanned both categories.[8] Impressionistically, two speakers, F7 and F16, seem to sometimes use [o] and sometimes use [ɯ]; their lexical vowels are clearly separated in acoustic space, and they have epenthetic vowels that sometimes fall into one category and sometimes into the other. Another two speakers, F5 and F12, have lexical [o] and [ɯ] categories that are quite similar to each other. Their epenthetic vowels span this entire space, making it harder to determine whether they are sometimes using [o] and sometimes [ɯ], or simply have one epenthetic vowel category that is somewhere in between. Finally, M4 is particularly difficult to categorize; his lexical vowels are fairly distinct, and most of his epenthetic vowels seem to be within the [o] category, but there are three tokens that are closer to his [ɯ] category; all three come from the word [addʑa],

---

[8] As Kubozono (2015) points out, when faced with loanword adaptation in the context of an alveolar stop followed by /u/, Japanese speakers have two obvious choices, namely, to use the vowel closest to /u/ and therefore adapt with [ɯ], or to use [o], which is typically what actually happens. The reason for this choice is generally explained as a tension between being faithful to the vowel versus faithful to the consonant. The dilemma arises because using [ɯ] in this context would lead one to expect accompanying affrication, following the general allophonic rule in Japanese turning alveolar stops into affricates before [ɯ], such that the sequences surface as [tsɯ] or [dzɯ]. By changing the vowel to [o], the original consonant quality as [t] or [d] can be preserved. Indeed, Kubozono gives examples of both strategies being used, though the use of [ɯ] is treated as being exceptional. For our speakers, however, the use of [ɯ] was not typically accompanied by this affrication; that is, the tokens reported here as having [ɯ] were in fact sequences of [dɯ] and not [dzɯ]. All tokens were checked manually by the first author, who is a native speaker of Japanese, and any that were questionable were verified by the third author, who is a native speaker of English.
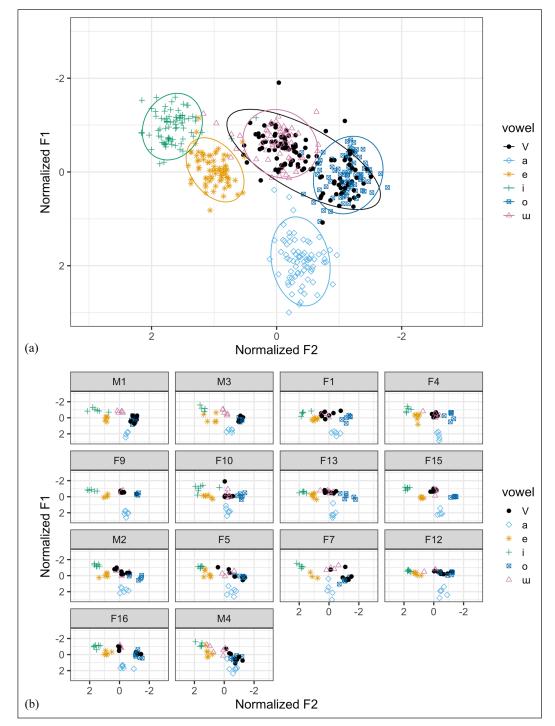
Figure 8: (a) Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel, averaged across 14 speakers, in the alveolar context. The number of tokens of each vowel are: [a] = 76, [e] = 74, [i] = 78, [o] = 81, [ɯ] = 67, V = 114. Ellipses are 95% data ellipses. **(b)** Individual F1 and F2 plots for each speaker in the alveolar contact. The first two speakers appear to use epenthetic [o]; the next seven appear to use [ɯ]; and the final five appear to use either both [o] and [ɯ] or a vowel that is somewhere between these two lexical vowels.

suggesting that the *following* context may be influencing the quality of the epenthetic vowel as well.

This division of the speakers into subsets was based on both visual inspection of their individual vowel plots and statistical analyses of each individual speaker (the complete results of which are shown in Appendix B). That said, we acknowledge that the amount of
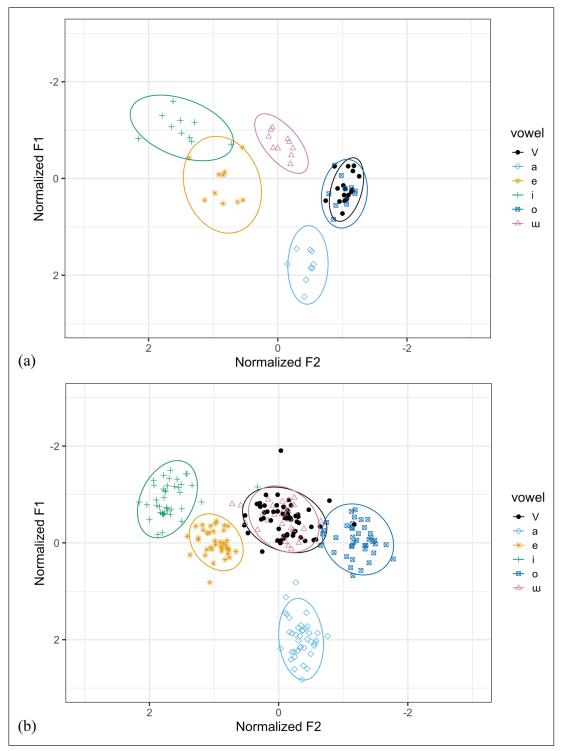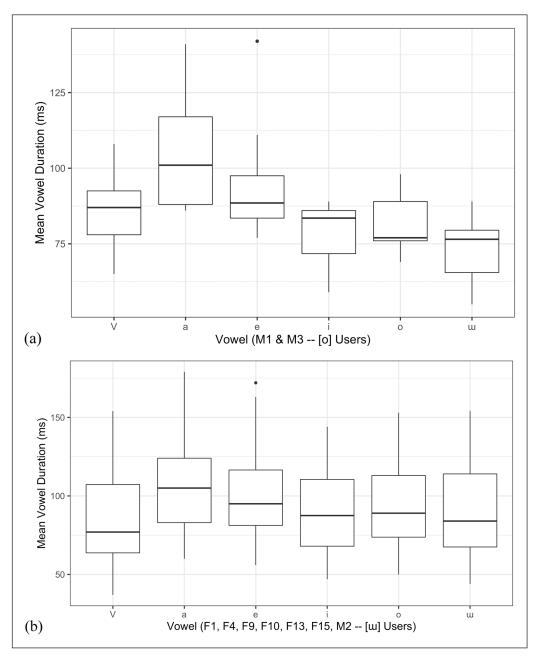
Figure 9: (a) Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel
in the alveolar context from (a) two speakers [M1, M3] who appear to primarily use [o] as the
epenthetic vowel and **(b)** seven speakers [F1, F4, F9, F10, F13, F15, M2] who appear to primarily
use [ɯ] as the epenthetic vowel. Ellipses are 95% data ellipses. In (a), the number of tokens for
each vowel is: [a] = 9, [e] = 10, [i] = 10, [o] = 11, [ɯ] = 10, and V = 18. In (b), the number of tokens
for each vowel is: [a] = 39, [e] = 38, [i] = 40, [o] = 40, [ɯ] = 35, and V = 56.

intra-speaker replication is quite small, and the experiment was not designed to examine
this kind of individual variability (see also discussion in Senn, 2014); we are not trying to
make strong claims about the relative strength or frequency of any of these sub-patterns.
Instead, the key point here is that the alveolar-stop context resulted in a considerable

degree of both intra- and inter-speaker variability in a way that is quite different from that of the labial and velar contexts discussed above.

3.4.2. Duration analyses

We now turn to an examination of the durations of the epenthetic vowel in the alveolar-stop context. For the two sub-groups of participants that seemed to consistently produce an epenthetic vowel quality similar to one of the lexical vowels, the duration results are also consistent with their producing that same lexical vowel. In particular, it is notable that the epenthetic vowel is not simply similar to the shortest vowel (regardless of quality), but rather seems to match the duration of the lexical vowel that it is similar in quality to. That is, for the two speakers for whom the epenthetic vowel is similar in quality to [o] in the alveolar context, the duration of the epenthetic vowel is most similar to [e] or [o], and is in fact both significantly longer than [i] and [ɯ], and significantly shorter than [a] (see **Figure 10a**
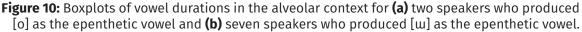

Figure 10: Boxplots of vowel durations in the alveolar context for **(a)** two speakers who produced [o] as the epenthetic vowel and **(b)** seven speakers who produced [ɯ] as the epenthetic vowel.

Table 12: Effect estimates and *p*-values on predictors for vowel duration in the alveolar context
for two groups. Vowels with a significantly different duration value from the epenthetic vowel
V are marked with asterisks.

(a) Two speakers who produced [o]

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	86.94	5.71	15.23	0.011	*
Vowel a	18.35	5.15	3.57	0.0007	***
Vowel e	7.66	4.97	1.54	0.129	
Vowel i	−9.97	4.98	−2.00	0.05	*
Vowel o	−5.07	4.82	−1.05	0.30	
Vowel ɯ	−13.44	4.97	−2.71	0.009	**

(b) Seven speakers who produced [ɯ]

	Estimate	SE	t value	Pr(>|t|)	
(Intercept)	85.35	10.35	8.24	3.94e-05	***
Vowel a	23.29	5.65	4.12	0.002	**
Vowel e	18.26	5.67	3.22	0.008	**
Vowel i	5.30	5.57	0.95	0.360	
Vowel o	11.01	5.64	1.95	0.076	
Vowel ɯ	5.65	5.72	0.99	0.342	

and **Table 12a**). For the seven speakers for whom the epenthetic vowel is similar
in quality to [ɯ], the duration is also short and consistent with both [i] and [ɯ],
although also not quite statistically significantly different from [o] (see **Figure 10b** and
Table 12b).

3.5. Epenthetic vowels in the palatal context

3.5.1. F1 and F2 analyses

Finally, we turn to the palatal context, [dʑ]. Recall that the expected epenthetic
vowel based on earlier studies would be similar to the lexical vowel [i]. However,
this was not found to be the case for the majority of speakers in this study. In fact,
only two speakers produced a vowel similar to [i], while four produced epenthetic
vowels similar to [ɯ], and for the others there was a great deal of variability.
As above, we describe the results in terms of two factors, formant frequency and
vowel duration. Vowel quality results are described in terms of observable trends
in the data; they were not analyzed statistically since the dataset for each pattern
is small. Statistical results for each individual speaker in this context are given in
Appendix C.

   Figure 11a shows the individual epenthetic and lexical vowels across all speakers in the
palatal context. As can be seen, the epenthetic vowels are broadly distributed across the
entire vowel space, overlapping with each of the other five vowel categories.[9] Based on a

---

[9] Note that the data ellipse for the epenthetic vowel presents itself somewhat unintuitively, in that it seems
to have been unaffected by the epenthetic tokens that are similar to lexical [a]. As far as we can tell, this
is simply a function of the non-normal distribution of these data in this context. The primary point of this
graph is simply that the epenthetic vowel can take on any lexical vowel quality in this context; the ellipse
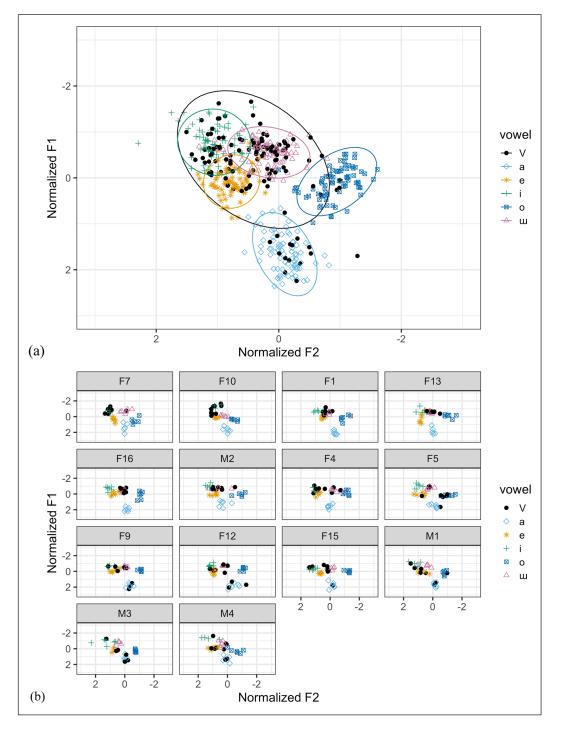helps visually summarize the data, but is not used analytically.

Figure 11: Normalized mean F1 and F2 values for each lexical vowel and individual epenthetic vowels from **(a)** all 14 speakers in the palatal context and **(b)** for each individual speaker in the palatal context.

combination of visual inspection of the individual plots in **Figure 11b** and the statistical analyses in Appendix C, speakers generally fall into one of three groups, as discussed just below: (a) epenthetic-[i]; (b) epenthetic-[ɯ]; or (c) variable.

Of the 14 speakers, two (F7, F10) appeared to consistently produce an epenthetic [i] in this context, the vowel observed after palatals in prior studies. As shown in **Figure 12**, the acoustic space of the epenthetic vowel for these two speakers completely overlaps with that of lexical vowel [i].

Figure 12: Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel from two speakers (F7, F10) who appear to epenthesize [i] in the palatal context. Number of vowel tokens displayed: [a] = 11, [e] = 11, [i] = 10, [o] = 10, [ɯ] = 12, V = 17. Ellipses are 95% data ellipses.
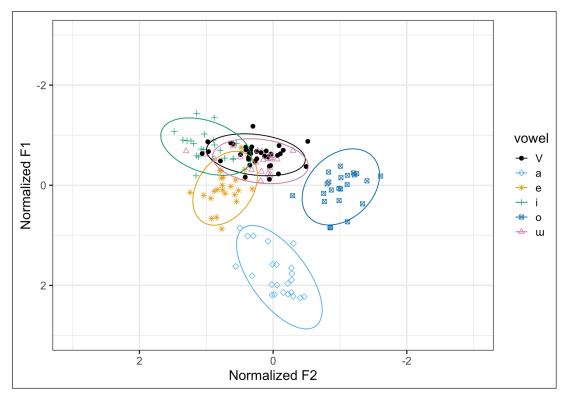

Figure 13: Normalized mean F1 and F2 values for each lexical vowel and the epenthetic vowel from four speakers (F1, F13, F16, M2) who appear to epenthesize [ɯ] in the palatal context. Number of vowel tokens: [a] = 23, [e] = 24, [i] = 23, [o] = 23, [ɯ] = 23, V = 36. Ellipses represent 95% confidence intervals.

Table 13: A list of the results in palatal context, by speaker within the 'variable' group.

Speaker	Epenthetic Vowel in Context
	[dʑ]
F4	[e], [i], [ɯ]
F5	[a], [e], [o], [ɯ]
F9	[a], [e], [i], [ɯ]
F12	[a], [e], [ɯ]
F15	[a], [i], [ɯ]
M1	[a], [e], [i], [o]
M3	[a], [e], [i]
M4	[a], [e], [ɯ]

Four speakers (F1, F13, F16, M2) consistently produced an epenthetic vowel that is most consistent with their lexical [ɯ] in this context, as shown in **Figure 13**. Although the epenthetic vowel ellipsis extends slightly into the region of [i] and [e], it mostly overlaps with the acoustic space of [ɯ]. Notably, the lexical [ɯ] vowel in this context appears to be fronted as compared to its production in the other contexts, such that it also overlaps with lexical [i] and [e].

Finally, the remaining eight speakers (F4, F5, F9, F12, F15, M1, M3, M4) were not sorted into either of the two groups above, since each of these speakers produced epenthetic vowels of multiple different qualities. Statistical analysis did not support that the speakers produced any specific vowel in the palatal context. Vowel quality results are described in terms of observable trends in the data in **Figure 11** and are summarized in **Table 13**. Again, these results are not meant to be definitive descriptions of what these speakers might *always* do, but instead are intended to highlight the wide range of variability in this context, including the use of [a] and [e].

3.5.2. Duration analyses

Turning to duration, the vowel durations for the epenthetic vowel in the palatal context overall were also quite variable—epenthetic vowels were as short and as long as the shortest and longest lexical vowels, which isn't surprising given that there were tokens of the epenthetic vowel in this context that matched each possible lexical vowel quality. **Figure 14** provides boxplots and **Table 14** provides a summary of vowel durations in the palatal context for the two groups of speakers who were relatively consistent in their productions in this context, i.e., the two speakers who produced [i] and the four speakers who produced [ɯ]. These groups of participants were also more consistent with their vowel durations; both groups produced consistently short vowels, similar in duration to the lexical vowel they seemed to be producing.

For speakers who appear to use the vowel [i] as epenthetic after the palatal [dʑ], the plot in **Figure 14(a)** and **Table 14(a)** show that their epenthetic vowel is quite short, similar to their lexical [i]. A linear mixed effect model showed that [a], [e], and [o] are predicted to have significantly larger duration values than the epenthetic vowel ($p < 0.01$). Both [i] ($t = -0.12, p = 0.90$) and [ɯ] ($t = 1.97, p = 0.07$) were not significantly different from V, though [ɯ] was close.

For speakers who appear to use the vowel [ɯ] as epenthetic after the palatal [dʑ], the plot in **Figure 14(b)** and **Table 14(b)** show that the shortest vowel of this group is the
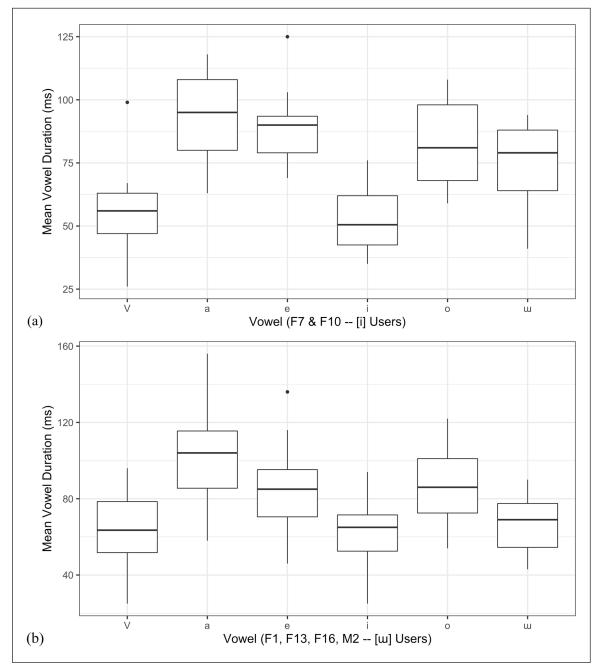
Figure 14: Boxplots of vowel durations for **(a)** the two speakers who appear to epenthesize [i] in the palatal context and **(b)** the four speakers who appear to epenthesize [ɯ] in the palatal context.

epenthetic vowel [i]. A linear mixed effect model showed [a], [e], [o] are predicted to have significantly different duration values than the epenthetic vowel ($p < 0.01$). Both [i] ($t = -0.33$, $p = 0.77$) and [ɯ] ($t = 0.65$, $p = 0.53$) were not significantly different from V.

In summary, for the two groups that produced a single epenthetic vowel quality in the palatal context, their epenthetic vowel was consistently short, and not significantly different in either case from the duration of the lexical vowel they seemed to use. In both cases, though, this meant that the epenthetic vowel was not significantly different from either [i] or [ɯ].

Table 14: Effect estimates and *p*-values on predictors for vowel duration in the palatal context for (a) the group who produced [i] and (b) the group who produced [ɯ]. Vowels with a significantly different duration value from the epenthetic vowel V are marked with asterisks.

(a) Two speakers who produced [i]

	Estimate	SE	*t* value	Pr(>|t|)	
(Intercept)	54.12	7.69	7.04	7.09e-05	***
Vowel a	37.12	10.17	3.68	0.003	**
Vowel e	35.14	10.17	3.45	0.005	**
Vowel i	−1.26	10.30	−0.12	0.90	
Vowel o	27.63	10.30	2.68	0.02	*
Vowel ɯ	19.89	10.10	1.97	0.07	

(b) Four speakers who produced [ɯ]

	Estimate	SE	*t* value	Pr(>|t|)	
(Intercept)	63.08	7.57	8.34	7.26e-05	***
Vowel a	37.93	7.16	5.30	0.0002	***
Vowel e	21.79	7.12	3.06	0.01	*
Vowel i	−2.18	7.16	−0.30	0.77	
Vowel o	24.63	7.16	3.44	0.006	**
Vowel ɯ	4.64	7.16	0.65	0.53	

4. Discussion

The current study was designed to directly test the nature of epenthesis as a strategy to break up unfamiliar consonant clusters in Japanese using orthographic input. The experiment revealed two primary results of interest. First, the production of consonant clusters consistently yielded epenthetic vowels when [aCCa] pseudo-word stimuli were presented (all 504 target tokens included epenthetic vowels, although some tokens were not included in the analysis because the accompanying consonants were misread). This suggests that the general Japanese phonotactic restriction against consonant clusters influences the production of such clusters, as expected. Second, the results are only partially consistent with previous studies in terms of *which* epenthetic vowel is used. The baseline hypothesis would be that the quality of epenthetic vowels would be similar to patterns found in lexicalized loanword phonology, i.e., [ɯ] after labials and velars, [o] after alveolar stops, and [i] after palatals. After labial and velar consonants, the present study did indeed find that the quality of the epenthetic vowel was quite consistently [ɯ], as has been found in studies of Japanese loanword adaptation (Hirayama, 2003; Kubozono, 2015; Yazawa et al., 2015). However, the current results diverge from those predicted from loanword studies and those found in Yazawa et al. (2015) in the alveolar and palatal contexts. We observed that while some speakers produced [o] in the alveolar context and [i] in the palatal context, such speakers were actually in the minority. Instead, the epenthetic vowel [ɯ] was commonly used in each of these contexts, although there was also a great degree of variability among individuals in the type of vowel that was inserted, as we discuss further below.

The epenthetic vowels produced by each speaker in each context are summarized in **Table 15**. Note that only one speaker, F7, was close to producing all and only the expected vowel qualities in each context, and even she produced unexpected tokens

Table 15: A summary of the results for each speaker, by context (Shading marks vowels that
were produced as predicted given prior literature. Note that speakers F15 and M3 in the labial
context used unpredicted [a] only once each, and similarly, and speakers F13 and F15 in the
velar context used unpredicted [o] only once each. Speaker F9 used [i] three times in the labial
and one time in the velar context).

Speaker	Age	[b]	[g]	[d]	[dʑ]
F1	23	[ɯ]	[ɯ]	[ɯ]	[ɯ]
F4	22	[ɯ]	[ɯ]	[ɯ]	[e], [i], [ɯ]
F5	21	[ɯ]	[ɯ]	[o] and [ɯ]	[a], [i], [o], [ɯ]
F7	46	[ɯ]	[ɯ]	[o] and [ɯ]	[i]
F9	30	[ɯ] and [i]	[ɯ] and [i]	[ɯ]	[a], [e], [i], [ɯ]
F10	27	[ɯ]	[ɯ]	[ɯ]	[i]
F12	21	[ɯ]	[ɯ]	[o] and [ɯ]	[a], [e], [ɯ]
F13	22	[ɯ]	[ɯ] and [o]	[ɯ]	[ɯ]
F15	31	[ɯ] and [a]	[ɯ] and [o]	[ɯ]	[a], [i], [ɯ]
F16	25	[ɯ]	[ɯ]	[o] and [ɯ]	[ɯ]
M1	22	[ɯ]	[ɯ]	[o]	[a], [e], [i], [o]
M2	40	[ɯ]	[ɯ]	[ɯ]	[ɯ]
M3	24	[ɯ] and [a]	[ɯ]	[o]	[a], [e], [i]
M4	28	[ɯ]	[ɯ]	[ɯ] for [addʑa]; [o] for others	[a], [e], [i], [ɯ]

of [ɯ] in the alveolar context. Interestingly, F7 is also the oldest speaker among the
participants (age = 46), which might be suggestive that older speakers tend to have a
more conservative pattern. However, the next oldest speaker, M2 (age = 40), consistently
produced [ɯ] in all contexts. In fact, three speakers in the current study (F1, F13, and M2)
consistently used [ɯ] as the epenthetic vowel in all contexts; F1 and F13 were among the
younger speakers overall (ages 23 and 22, respectively). The remaining eleven speakers
had varying patterns. As above, we present this summary not to put too much weight on
the 'pattern' produced by any given speaker, but rather to showcase the striking contrasts
between (1) the regularity of the expected epenthetic [ɯ] in the labial and velar contexts,
(2) the tendency for both expected [o] and unexpected [ɯ] to be used in the alveolar
context, and (3) the striking irregularity in the palatal context (though note that even in
this highly variable context, 10 of the 14 speakers did use at least some instances of [ɯ]).

Recall from Section 1 that other recent studies have also investigated these epenthesis
patterns, and the conflicting results across these prior studies were in part what motivated
the current study. Interestingly, the current results do not match any of those other results,
although there are some similarities.

Yazawa et al. (2015) conducted a production study where the expected tri-partite
pattern was in fact largely found. One difference between that study and the current one
is that Yazawa et al. used a text-reading task in which participants read the Aesop fable
"The North Wind and the Sun" in English, whereas pseudo-words were used in the current
study. Many of the contexts in this passage where epenthesis would be expected were
word final rather than word medial, as in the current study, and it is possible that the
epenthesis patterns are simply not the same across these two kinds of contexts.

Additionally, the Yazawa et al. (2015) participants (who were all English learners) were
likely in 'English mode,' given the task, whereas the current participants were specifically
instructed to produce the words as if they were Japanese. *A priori*, this difference might

be expected to have biased the results in the opposite direction—that is, one might think that Japanese nonce words would be more likely to follow the traditional tri-partite pattern than English words being produced in English. That said, there are at least two possible reasons for the actual results. One would be that the specific English words that were produced with epenthetic vowels might have lexicalized loanword counterparts that happen to follow the more traditional pattern. Alternatively, the epenthesis in the Yazawa et al. study might in some sense have been more 'naturalistic,' in that it happened while participants were reading out a full passage, with their attention presumably directed toward fluent English production more generally. In the current study, on the other hand, attention was focused on illicit consonant clusters in individual nonce words. It is possible that in the more naturalistic setting, epenthesis patterns similar to those in lexicalized Japanese loanwords were more likely to arise spontaneously, while in the more targeted setting, participants were more likely to apply meta-awareness of some sort (e.g., trying to be consistent across their productions of multiple nonce words with different consonantal contexts).

It is also potentially important to note that, while most of the Yazawa et al. (2015) participants "produced at least one epenthetic vowel," there were only 518 total epenthetic tokens in their data, despite having 71 participants each reading a passage with more than 60 likely opportunities for epenthesis to occur; that is, epenthesis occurred in only around 12% of the places where it might have, suggesting that their participants were very much English-like in their productions. The current study had approximately the same number of actual tokens of epenthetic vowels (463), but these were concentrated in the productions of only 14 speakers, and our participants did in fact epenthesize in 100% of the expected contexts. Thus, the apparent conformity of the tripartite pattern in Yazawa et al. is diluted across a wide set of speakers and contexts. Of the four contexts tested here, all fourteen speakers produced the expected vowel at least some of the time in the labial and velar contexts, and more than half of them produced the expected vowel at least some of the time in the alveolar and palatal contexts. If the rate of epenthesis had been lower in the current study, it is possible that the 'aberrant' instances would have been more sparsely represented, and the epenthesis patterns have looked more as expected.

Shoji and Shoji (2014) also conducted a production study whose results do not match the Yazawa et al. (2015) ones, but their results only partially line up with the current ones. Recall that the production task in Shoji and Shoji was orthographically based: Their participants were provided with nonsense words written in Latin script (e.g., *consuch, zod, bkautu*) and had to *re-write* them in Japanese characters. Interestingly, they did find that the tri-partite pattern held for the vast majority of their *word-final* epenthetic contexts, but that it broke down in the word-initial contexts (again suggesting that epenthesis patterns depend at least partially on word position). Of course, in the current study, the consonant clusters targeted by epenthesis were word-medial, making it harder to directly compare the results. That said, in both the palatal and the alveolar stop word-initial contexts, Shoji and Shoji did find, as do we here, that [ɯ] is an extremely common epenthetic vowel, being used around 25% of the time in each context (more than any other vowel except the traditionally expected ones, which occurred 35–45% of the time). Thus, there is at least some converging evidence that Japanese speakers treat [ɯ] as a good candidate epenthetic vowel in the alveolar-stop context in nonce word production.

In terms of perception, both Monahan et al. (2009) and Mattingley et al. (2015) used VCCV stimuli, analogous to those used in the current study. It should be noted that the VCCV stimuli in both studies contained stop release bursts. The presence of the stop release has been known to influence speech perception and facilitate non-native speakers to perceive illusory vowels (e.g., Daland, Oh, & Davidson, 2019). Monahan et al. tested only

alveolar-stop and velar contexts, using an AX discrimination task. For the velar contexts, they did indeed find that Japanese listeners seemed to perceive an illusory epenthetic vowel, specifically, [ɯ], as would be expected. But they found that in the alveolar-stop contexts, neither [o] (as would be expected if perception mirrored traditional production) nor [ɯ] (as would be expected if the illusory vowel were always the 'default' vowel) was perceived as an illusory vowel. Instead, Japanese listeners seem to have perceived these sequences faithfully as VCCV sequences, on par with English listeners. Mattingley et al., on the other hand, did test the same full set of contexts as the current study, using an identification task. They found that in the labial, velar, and palatal contexts, an illusory vowel was heard the majority of the time (though that varied from 70% for the labial and velar contexts to 98% for the palatal context), *and* that the illusory vowel usually matched the traditional prediction (i.e., [ɯ] for labial and velar and [i] for palatal). In the alveolar-stop context, they found that 60% of tokens were identified with an illusory epenthetic vowel (echoing Monahan et al.'s finding that illusory vowels are somewhat less likely in this context), but that within that 60%, the majority (70%) were [ɯ] and only 16% were [o]. Thus, again, at least for the alveolar-stop context, there seems to be evidence from perception as well that [ɯ] is a viable epenthetic vowel.

The current study was intended to describe *what* the production patterns are, rather than to try to explain *why* the patterns might differ from those traditionally reported. Especially given the varying results across the various recent production and perception studies, it would be premature to try to come up with an explanation for any given set of results. That said, the independent observations that the typical phonotactic constraints in loanword adaptation are loosening (e.g., Pintér, 2008, 2015; Hall, 2013; Kubozono, 2015) suggest at least one pathway of change. Specifically, if there is no longer a phonotactic constraint against the sequence [dɯ] (at least in loanwords), then there is no reason not to use the default epenthetic vowel [ɯ] in this context just as in the labial and velar contexts, exactly as seen in the current results.

Interestingly, even in the older and more conservative stages of Modern Japanese, there has not been a phonotactic restriction against [dʑɯ] sequences; the only historical phonotactic restriction after the palatals was that [e] could not occur. The ostensible reason for the use of epenthetic [i] after palatals simply comes from an articulatory or perceptual closeness between the palatals and [i] (Kubozono, 2015).[10] But, as Kubozono (2015, p. 330) points out, "[t]his raises the interesting question of why the palatoalveolar fricative [ɕ][11] usually takes /ɯ/[12] rather than /i/." Perhaps, then, the use of [ɯ] in the other three contexts has enabled it to be used after palatal affricates as well. This possibility would not, of course, explain the huge amount of intra- and inter-speaker variability in terms of *other* vowels used epenthetically in this context (such as [e] and [a]), but we are somewhat reluctant to theorize too much about their use given that other studies have not found similar trends.

That said, one potential explanation for the variety in the palatal context comes from the influence of English orthography; see, e.g., Vendelin and Peperkamp (2006) for general discussion. In English orthography, the letter <j> is most often followed by <u> (247 words in the IPHOD corpus, Vaden, Halpin, & Hickok, 2009); next by <a>, <o>, or <e> (around 180 words each in the IPHOD corpus), and least often by <i> (only 46 words in the IPHOD corpus). Given that (1) the current participants were in an English-speaking

---

[10] Note that [i] is also the most frequent vowel, both in terms of type and token frequency, after palatals both in the Corpus of Spontaneous Japanese (National Institute for Japanese Language and Linguistics, 2008) and in words tagged specifically as loanwords in the Balanced Corpus of Contemporary Written Japanese (National Institute for Japanese Language and Linguistics, 2011).
[11] Transcribed in Kubozono (2015) as /ʃ/.
[12] Transcribed in Kubozono (2015) as /u/.

country at the time of participation, (2) the target words were written using Romanization instead of Japanese characters, and (3) most of the participants showed at least some influence of English orthography by misproducing at least some of the <ge> or <gi> sequences as starting with [dʑ], it is not outside the realm of possibility that they could have chosen an epenthetic vowel in the <j> context based on what they thought was likely given English orthography. Interestingly, however, in terms of English *phonology*, [dʒ] is most often followed by [i] or [ɪ] (613 words in the IPHOD corpus),[13] followed by [e] or [ɛ] (304 words), or another vowel (fewer than 100 words each). And, the current participants did certainly produce the orthographic <j> as [dʑ] (indeed, it was the most accurately produced of the consonantal contexts, with 94% of possible tokens being produced with the correct consonant; the next most accurate was the labial context, with 90%). Furthermore, there's no particular evidence that the current participants relied on English orthographic frequency patterns for any of the other contexts; in the IPHOD corpus, <b> is most frequently followed by <a>, while both <d> and <g> are most frequently followed by <e> (and in fact all three are *least* often followed by <u>), so if English orthography is playing a role here, it is unclear as to why it would do so *only* for <j>. Other potential explanations for the unusual behaviour in the palatal context (the following consonant, priming from adjacent stimuli, the age or gender of speakers, etc.) would similarly be limited in their ability to uniquely predict variability in this context, because such factors were consistent across all contexts.

Another possibility to consider is that the apparent epenthetic vowels that are seen in the current study might in fact not be truly epenthetic but rather the result of gestural mistiming between the production of the first and second consonant in the sequence.[14] If this were the case, then it would make sense that the quality of many of the vowels across the contexts was similar, as it would not be phonologically governed, but rather based on articulatory facts. Davidson (2010) considers the difference between full epenthetic vowels and transitional vocoids in cases where a schwa was inserted between two consonants word-initially by English and Catalan speakers. To diagnose the difference between these scenarios, she examines duration, F1, and F2. A transitional vocoid would likely be shorter than a lexical vowel, have a lower F1 value (because it would be produced not with a tongue height target but rather as a temporary lowering between the surrounding stop closures), and have a lower F2 value in this context (because it would be likely to anticipate the upcoming [a] "even more than for normal vowel-to-vowel coarticulation," as the [a] would be the next upcoming vowel target; Davidson, 2010, p. 283). As noted in the above results sections, however, none of these characteristics were found. That is, the epenthetic vowels were never significantly different from the lexical vowel they were most similar to in any of these three dimensions. Thus, we believe that the epenthetic vowels found in our tokens are indeed inserted vowels, with accompanying vowel targets, and not simply insufficient attempts to produce CC clusters.

In terms of the larger implications of this work, while the task in the current paper was a non-word production task intended to provide insight into how illicit consonant clusters might be adapted by native Japanese speakers in a loanword situation, the stimuli in the experiment were not themselves loanwords, and so we cannot claim that loanword adaptation patterns are changing. That is, our stimuli were not associated with meaning and did not originate from some foreign source that might be known to the participants. Thus, some factors that likely influence the way that loanwords are adapted were absent from the current study—in particular, participants did not have acoustic models of the

---

[13] Though it is important to note that [dʒi] and [dʒɪ] sequences in English are generally spelled with <g>.
[14] We are grateful to an anonymous reviewer for bringing up this possibility.

words, nor did they have any knowledge of ways in which they might be related to other related loanwords, and there was no larger social context that might influence the adaptation. This is why we cannot claim that loanwords would necessarily follow the same trends as those seen here, but we can speculate that loanword adaptation may be undergoing changes such that loanwords with medial CC clusters in which the first member is a (voiced) alveolar stop or a palatal may no longer follow the clear tri-partite pattern traditionally reported in the literature. Supporting this speculation is the fact that, as was noted in Section 1, Japanese loanwords and nonce words have been shown to follow similar phonological patterns (Kawahara, 2012 and references therein). And in fact, at least one online English/Japanese dictionary tool[15] uses ドゥ [duɯ] instead of ド [do] for English loanwords whose final consonant is [d] in Kana. For example, 'food' /fu:d/ is represented as 'フードゥ' and 'read' /ri:d/ is represented as 'リードゥ'. All English words that end with [d] or [t] are written in katakana with [duɯ] ドゥ and [tuɯ] トゥ instead of using [do] ド and [to] ト.

5. Conclusion

The current paper reports on a production experiment that directly tests the nature of epenthesis as a strategy to break up unfamiliar consonant clusters in Japanese. All fourteen of the speakers in the experiment consistently produced an epenthetic vowel in VCCV sequences, following the expected phonotactic patterns of syllable structure in Japanese. The *choice* of epenthetic vowel, however, was not always as expected. The expected vowel [ɯ] was used in the labial and velar contexts, but also by many speakers in many tokens in both the alveolar-stop and the palatal contexts, where [o] and [i] would have been expected, respectively. Although a full explanation for the variability of the results is beyond the scope of the paper, the results do suggest that the independent loosening of the phonotactic constraint against [duɯ] sequences may be affecting epenthesis strategies typically assumed to be governed by this constraint. We thus predict that in new loanwords with consonant clusters (at least, word-medial clusters with voiced obstruents), we may increasingly see [ɯ] being used as an across-the-board epenthetic vowel, rather than the quality continuing to be governed by preceding context.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** List of items for production experiment. DOI: https://doi.org/10.5334/labphon.158.s1
- **Appendix B.** Results of linear mixed-effects models predicting F1 and F2 for individual speakers in the alveolar context. DOI: https://doi.org/10.5334/labphon.158.s2
- **Appendix C.** Results of linear mixed-effects models predicting F1 and F2 for individual speakers in the palatal context. DOI: https://doi.org/10.5334/labphon.158.s3

Acknowledgements

---

[15] "英辞郎 on the WEB" at https://eow.alc.co.jp.

Competing Interests

The authors have no competing interests to declare.

References

Akamatsu, T. (2000). *Japanese phonology: A functional approach*. München: Lincom Europa.

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

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In C. Goodman & H. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge: The MIT Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception: Standing at the crossroads. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross- language research* (pp. 171–204). Baltimore: York Press.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*, 305–330.

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, *26*(1), 86–125. DOI: https://doi.org/10.2307/410409

Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer [Computer program]. Version 5.4. Retrieved from http://www.praat.org/

Byarushengo, E. R. (1976). Strategies in loan phonology. In *Annual Meeting of the Berkeley Linguistics Society*, *2*, 78–88. DOI: https://doi.org/10.3765/bls.v2i0.2304

Campbell, N. (1992). *Segmental elasticity and timing in Japanese speech* (pp. 403–418). IOS Press.

Daland, R., Oh, M., & Davidson, L. (2019). On the relation between speech perception and loanword adaptation: Cross-linguistic perception of Korean-illicit word-medial clusters. *Natural Language and Linguistic Theory*, *37*(3), 825–868. DOI: https://doi.org/10.1007/s11049-018-9423-2

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137. DOI: https://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*(2), 272–288. DOI: https://doi.org/10.1016/j.wocn.2010.01.001

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578. DOI: https://doi.org/10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and cognitive processes*, *16*(5–6), 491–505. DOI: https://doi.org/10.1080/01690960143000191

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*, 199–210. DOI: https://doi.org/10.1016/j.jml.2010.12.004

Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics*, *7*, 71–116.

Fox, J., & Weisberg, S. (2011). An {R} Companion to Applied Regression, Second Edition. Thousand Oaks, CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. Columbus, OH: The Ohio State University Doctoral dissertation.

Hall, K. C. (2013). Documenting phonological change: A comparison of two Japanese phonemic splits. In S. Luo (Ed.), *Actes du congrès annuel de l'Association canadienne de linguistique 2013/Proceedings of the 2013 Annual Conference of the Canadian Linguistic Association*.

Hall, N. (2011). Vowel Epenthesis. In M. van Oostendorp, C. Ewen, E. Hume & K. Rice (Eds.), *Comparison to phonology* (pp. 1576–1596). Oxford: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0067

Hallé, P. A., Segui, J., Frauenfelder, U. H., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of experimental psychology: Human perception and performance, 24*(2), 592–608. DOI: https://doi.org/10.1037//0096-1523.24.2.592

Han, M. (1962). The feature of duration in Japanese. *The study of sounds, 10*, 65–80.

Hirayama, M. (2003). Contrast in Japanese vowels. *Toronto Working Papers in Linguistics, 20*.

Irwin, M. (2011). *Loanwords in Japanese*. Philadelphia: John Benjamins Publishing Company. Retrieved from http://www.canterbury.eblib.com.au/patron/FullRecord.aspx?p=717677. DOI: https://doi.org/10.1075/slcs.125

Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language & Linguistic Theory, 7*(2), 217–259. DOI: https://doi.org/10.1007/BF00138077

Itô, J., & Mester, A. (1995). Japanese phonology. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 817–838). Oxford: Blackwell.

Itô, J., & Mester, A. (1999). The phonological lexicon. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics* (pp. 62–100). Oxford: Blackwell.

Itô, J., & Mester, A. (2008). Lexical classes in phonology. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 84–106). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780195307344.013.0004

Kabak, B. (2003). *The perceptual processing of second language consonant clusters*. Newark, DE: University of Delaware Doctoral dissertation.

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech, 50*(1), 23–52. DOI: https://doi.org/10.1177/00238309070500010201

Kaneko, E. (2006). Vowel selection in Japanese loanwords from English. In *Proceedings LSO Working Papers in Linguistics*, 49–62. Madison: Linguistics Student Organization, University of Wisconsin-Madison. Retrieved from http//www.ling.wisc.edu/lso/wpl/6/kaneko.pdf

Kang, Y. (2003). Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology, 20*(2), 219–273. DOI: https://doi.org/10.1017/S0952675703004524

Kang, Y. (2011). Loanword phonology. *The Blackwell companion to phonology*. In M. v. Oostendorp, C. J. Ewen, E. Hume & K. Rice (Eds.), *The Blackwell Companion to Phonology Volume IV: Phonological Interfaces* (pp. 1003–1026). Oxford & Malden, Mass: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781444335262.wbctp0095

Katayama, M. (1998). *Optimality Theory and Japanese loanword phonology*. Santa Cruz, California: University of California Doctoral dissertation.

Kawahara, S. (2012). Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment studies. *Lingua, 122*(11), 1193–1206. DOI: https://doi.org/10.1016/j.lingua.2012.05.008

Kawahara, S., & Kao, S. (2012). The productivity of a root-initial accenting suffix, [-zu]: judgment studies. *Natural Language and Linguistic Theory*, *30*(3), 837–857. DOI: https://doi.org/10.1007/s11049-011-9132-6

Kenstowicz, M. (2007). Salience and similarity in loanword adaptation: A case study from Fijian. *Language Sciences*, *29*(2), 316–340. DOI: https://doi.org/10.1016/j.langsci.2006.12.023

Kubozono, H. (1996). Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin (Phonetic Society of Japan)*, *211*, 71–82.

Kubozono, H. (2001). Epenthetic vowels and accent in Japanese: Facts and paradoxes. In J. Van de Weijer & T. Nishihara (Eds.), *Issues in Japanese Phonology and Morphology* (pp. 113–142). Berlin, New York: Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110885989.111

Kubozono, H. (2006). Where does loanword prosody come from? A case study of Japanese loanword accent. *Lingua*, *116*(7), 1140–1170. DOI: https://doi.org/10.1016/j.lingua.2005.06.010

Kubozono, H. (2008). Japanese accent. In: S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 165–191). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780195307344.013.0007

Kubozono, H. (2015). Loanword phonology. In H. Kubozono (Ed.), *The handbook of Japanese language and linguistics: Phonetics and phonology* (pp. 313–361). Berlin: Mouton de Gruyter.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. J*ournal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Labrune, L. (2012). *The phonology of Japanese*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199545834.001.0001

Lovins, J. B. (1975). *Loanwords and the phonological structure of Japanese*. Bloomington: IULC.

Mattingley, W., Hume, E., & Currie Hall, K. (2015). The influence of preceding consonant on perceptual epenthesis in Japanese. In *Proceedings of the 18th International Congress of Phonetic Sciences* (ICPhS) *in Glasgow*.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese*. The Hague: Mouton.

Monahan, P. J., Takahashi, E., Nakao, C., & Idsardi, W. J. (2009). Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. In S. Iwasaki, H. Hoji, P. M. Clancy & S.-O. Sohn (Eds.), *Japanese/Korean Linguistics*, *17*, 391–405. Stanford, CA: CSLI Publications.

National Institute for Japanese Language and Linguistics. (2008). *The Corpus of Spontaneous Japanese*.

National Institute for Japanese Language and Linguistics. (2011). *The Balanced Corpus of Contemporary Written Japanese*.

Nogita, A., Yamane, N., & Bird, S. (2013). The Japanese unrounded back vowel /ɯ/ is in fact rounded central/front [ʉ – ʏ]. *Ultrafest VI Program and Abstract Booklet*, 39–42.

Otaki, Y. (2012). A phonological account of vowel epenthesis in Japanese loanwords: Synchronic and diachronic perspectives. *Phonological studies*, *15*, 35–42.

Peperkamp, S., & Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In *Proceedings of the 15th International Congress of Phonetic Science* (ICPhS) *in, Barcelona*, 367–370.

Pintér, G. (2008). *Asymmetrical segment distributions in Japanese*. Doctoral dissertation, Kobe University.

Pintér, G. (2015). The emergence of new consonant contrasts. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 121–165). Boston: De Gruyter Mouton.

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Sagisaka, Y., & Tokuhara, Y. (1984). *Kisoku ni yoru onsei gōsei no tame no on'in jikanchō seigyo kisoku* [Phoneme duration control for speech synthesis by rule]. Denshi tsūshin gakkai ronbunshi [The Transactions of the Institute of Electronics, Infromation and Communication Engineers A], *67*(7), 629–636.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.

Senn, S. (2014). Mastering variation: Variance components and personalised medicine. *Statistics in Medicine, 35*, 966–977. DOI: https://doi.org/10.1002/sim.6739

Shaw, J. A., & Kawahara, S. (2018). The lingual articulation of devoiced /u/ in Tokyo Japanese. *Journal of Phonetics, 66*, 100–119. DOI: https://doi.org/10.1016/j.wocn.2017.09.007

Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.

Shinohara, S. (1997). *Analyse phonologique de l'adaptation japonaise de mots étrangers*. Thèse pour le doctorat. Université de la Sorbonne nouvelle Paris III.

Shinohara, S. (2000). Default accentuation and foot structure in Japanese: Evidence from Japanese adaptations of French words. *Journal of East Asian Linguistics, 9*(1), 55–96. DOI: https://doi.org/10.1023/A:1008335811086

Shoji, S., & Shoji, K. (2014). Vowel Epenthesis and Consonant Deletion in Japanese Loanwords from English. In *Proceedings of the Annual Meetings on Phonology, 1*(1). DOI: https://doi.org/10.3765/amp.v1i1.16

Smith, J. L. (2006). Loan phonology is not all perception: Evidence from Japanese loan doublets. In T. J. Vance & K. A. Jones (Eds.), *Japanese/Korean Linguistics, 14*, 63–74. Palo Alto: CLSI.

Sperbeck, M. (2012). The production and perception of English consonant sequences by Japanese-speaking learners of English. In *Proceedings of Meetings on Acoustics, 9*(1), 060005. Acoustical Society of America. DOI: https://doi.org/10.1121/1.3651080

Steriade, D. (2001). Directional asymmetries in place assimilation: A perceptual account. In: E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (pp. 219–250). New York: Academic Press.

Steriade, D. (2008). The phonology of perceptibility effects: The P-map and its consequences for constraint organization. In S. Inkelas & K. Hanson (Eds.), *The nature of the word: Essays in honour of Paul Kiparsky* (pp. 151–180). Cambridge, MA: MIT Press. DOI: https://doi.org/10.7551/mitpress/9780262083799.003.0007

Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. Online Resource: http://ncslaap. lib. ncsu. edu/tools/norm

Tsujimura, N. (1996). *An introduction to Japanese linguistics*. Cambridge, MA: Blackwell Publishers.

Uffmann, C. (2006). Epenthetic vowel quality in loanwords: Empirical and formal issues. *Lingua, 116*(7), 1079–1111. DOI: https://doi.org/10.1016/j.lingua.2005.06.009

Vaden, K. I., Halpin, H. R., & Hickok, G. S. (2009). Irvine Phonotactic Online Dictionary, v. 2.0. Retrieved from www.iphod.com

Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany, N.Y.: State University of New York Press.

Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, UK: Cambridge University Press.

Vendelin, I., & Peperkamp, S. (2006). The influence of orthography on loanword adaptations. *Lingua*, 116, 996–1007. DOI: https://doi.org/10.1016/j.lingua.2005.07.005

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-3-319-24277-4_9

Yazawa, K., Konishi, T., Hanzawa, K., Short, G., & Kondo, M., (2015). Vowel epenthesis in Japanese speakers' L2 English. In *Proceedings of the 18th International Congress of Phonetic Science* (ICPhS) *in Glasgow*.

Yoshida, Y. (2006). Accents in Tokyo and Kyoto Japanese vowel quality in terms of duration and licensing potency. *SOAS Working Paper in Linguistics, 14*, 249–264.