The Perception and Production of Epenthetic Vowels in Non-native Clusters in Japanese:
Phonetic and Phonological Influences

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Master of Arts
in Linguistics in the
University of Canterbury
by
Wakayo Mattingley

University of Canterbury
2016

Acknowledgements

There are many people I would like to thank for helping me to come so far. The writing of this thesis was a challenging journey. Without those people who supported and encouraged me during the entirety of my research, I could not have completed my thesis. First and foremost, I would like to express my sincere gratitude to my supervisors. To Professor Beth Hume, I would like to thank her for her constant support, continued encouragement and insightful discussions as my primary supervisor throughout the project. And also to my second supervisor, Dr Kathleen Hall, who helped me from Canada by attending Skype meetings and e-mailing, giving me immense support and a lot of invaluable feedback. I cannot thank you enough for the time and energy that both you spent for me. I would like to thank Dr Heidi Quinn and Dr Kevin Watson for encouraging me to keep going for the last few years. I would also like to extend sincere thanks to Dr Viktoria Papp for her help and advice along the way. I would like to thank Dr Lynn Clark, Professor Jen Hay, Jacqui Nokes, Emma Parnell, and especially Dr Péter Rácz for statistics support. I would also express my gratitude to the PhD students for supporting me as well as being great friends. Ksenia Gnevsheva taught me how to use E-prime and gave me a lot of advice while I was writing the thesis. Darcy Rose did proofreading for my poor English, also giving me many useful tips for data analysis and writing. Xuan Wang helped me a lot when I was confused with analysing data using R. They also listened to me when I need to talk and brought laughter to me during the hard times. Mineko Shirakawa gave me useful tips for every aspect. I would like also like to thank Kylie Fitzgerald, Masako Ishii, Dan Jiao, Matthias Heyne, Jihyun Lee, and Keyi Sun. Without their support, this study could not have been done. In addition, I would like to thank all the people at the Department of Linguistics and the New Zealand Institute of Language, Brain and Behaviour at the University of Canterbury for giving me invaluable advice at the socio meeting and providing a helpful atmosphere during my research.

I would also like to thank Tommy Shirakawa, an amazing speaker, who helped me to create the stimuli used in the experiment. I am grateful to CCEL and Wilkinson’s Language School who helped me by allowing me to recruit from among their students. I wish to sincerely thank all who participated in my research for your cooperation and people who introduced me the participants. Without their voluntary participation, this study could not have been done.

I would like to thank my parents for their support and for encouraging me to pursue my long academic journey. Finally, I sincerely thank my husband, Steven Mattingley, for his understanding and encouragement for the last several years. None of the work would have been possible without your support and cooking!

Abstract

This thesis investigates the quality of epenthetic vowel that native speakers of Japanese tend to produce and perceive between unfamiliar sequences of consonants. Research on perceptual epenthesis in Japanese has revealed the high back [ɯ] to be the vowel commonly perceived in illicit consonant sequences. However, loanword studies suggest that there are three epenthetic vowels, which reflect phonotactic restrictions on certain consonant + vowel sequences. That is, the quality of epenthetic vowel is predictable from the preceding consonantal environment. In this study, I tested to what extent the response patterns in perceptual and production experiments are consistent with native phonotactics, and how phonetic properties of the listeners’ native language play a role in speech perception. This thesis first investigates the potential influence of the preceding consonant environment on perception and production of illicit consonant clusters. Second, the current study considers the effect of all vowel categories in Japanese, including allophonic variation of the Japanese high vowel [ɯ] — the high vowel undergoes devoicing when it occurs between voiceless obstruents — on the perception of illicit consonant sequences. This study thus integrates perceptual and production experimental work in an investigation of the contextual environments that contribute to predicting the quality of epenthetic vowels in Japanese.

In the perception experiment, a same-different AX discrimination task was employed, in order to determine whether native speakers of Japanese are able to tell the difference between licit [VC_1VC_2V] (C=consonant, V=vowel) and illicit [VCCV] pairs (e.g., [apata]-[apta]) when they listen to pre-recorded pseudo-word stimuli. In each trial, participants were asked to judge whether a pair of stimuli were the same or different. The experiment enabled us to test whether Japanese listeners perceive an illusory vowel between consonants in an illicit sequence and whether the vowel percept differs according to a given phonological environment. The results show that to some extent, the preceding consonant does influence the vowel perceived, yet there is a bias toward perceiving [ɯ] in voiceless consonantal contexts, a result not predicted by the language’s phonotactic patterns. Additionally, it was found that the order that the stimuli were presented to subjects influences epenthesis in perception. Japanese listeners were less accurate in identifying whether members of a pair were same-different with the [aCVCa-aCCa] order than with the [aCCa-aCVCa] order.

In the production experiment, a read-aloud task was employed. Speech production data was collected using the same pseudo-words as in the perception experiment though in this experiment the stimuli were presented to subjects orthographically. The results showed that for some preceding environments, the findings are relatively consistent with expectations based on the language's phonotactics, but this was not the case for all contexts. The results also revealed that there was variability across speakers as to which vowels they epenthesized after particular consonants.

The current series of studies revealed that the quality of epenthetic vowels was not merely influenced by the phonotactics of the native language in speech perception and production. Instead, other factors interact in a complex way during speech perception and production.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

The goals of this study are to examine first, the influence of native language phonotactics on the perception and production of vowel epenthesis in Japanese, especially focusing on the potential influence of preceding consonants, and second, the impact of phonetic properties of vowels on perception.

Many researchers have discussed loanword adaptation in Japanese, as well as the factors contributing to vowel epenthesis (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011; Irwin, 2011; Kaneko, 2006; Monahan, Takahashi, Nakao, & Idsardi, 2009; Peperkamp & Dupoux, 2003; Shoji & Shoji, 2014; Smith, 2005; Yazawa, Konishi, Hanzawa, Short & Kondo, 2015; among others). Loanword and production studies on epenthesis have shown that native speakers of Japanese have a tendency to insert vowels of three different types {i, o, u}, and that the type that is selected depends on the quality of the preceding consonant (Hirayama, 2003; Irwin, 2011; Katayama, 1998; Shoji &Shoji 2014, Yazawa et al. 2015).

Results from research on perceptual epenthesis differ in some ways from the above. One contributing factor may be due to the fact that perception studies have considered only a subset of the vowel qualities and preceding consonantal contexts examined in loanword and production studies. For example, Dupoux et al. (1999) and Dupoux et al. (2011) show that native speakers of Japanese are highly likely to perceive an illusory epenthetic vowel [ɯ] in stimuli containing consonant sequences that are illicit in their native language. However, the analyses do not take into account phonotactic patterns relating to the quality of the preceding consonants. In addition, only a subset of the language's five vowel qualities are considered: only the high front and/or high back vowels [i, ɯ] were presented to listeners. Although their study concludes that the phonotactics of a listener's native language affects speech perception, such an explanation tends to overlook the influence of the quality of the preceding consonant on epenthetic vowels in Japanese, as argued by many scholars (e.g., Irwin, 2011; Shoji & Shoji, 2014).

Monahan et al. (2009) considered the influence of preceding consonants, and focused on the perception of potential illusory vowels [ɯ, o] after alveolar [t d] and velar [k g] stops.

They interpret their results as indicating that Japanese listeners did not perceive the contextually predicted vowel [o] after alveolars, nor did they perceive an illusory epenthetic [ɯ] in this context. The high back vowel was, however, perceived after velar stops, consistent with loanword studies. The authors conclude that native language phonology alone cannot explain the perception of non-native speech.

Mattingley, Hume and Hall (2015) extended research on perceptual epenthesis, taking into account consonantal context and the full range of Japanese vowel qualities. The study investigated to what extent perceptual epenthesis is influenced by the quality of the preceding consonant and to what extent native phonotactic patterns constrain the process. Consistent with loanword studies, [ɯ] was perceived after labials and velars, while [i] was predominantly selected as the epenthetic vowel after the alveopalatal affricate. Yet, the mid back vowel [o] was not perceived much after the alveolar stop [d] as had been expected from loanword studies. Rather, the listeners were strongly biased to perceive [ɯ] after [d] even though *[dɯ] is an illicit phonotactic sequence in native Japanese. This result differs from Monahan et al. (2009).

The patterns reported in Mattingley et al. (2015) may be due to the use of different tasks than those of Monahan et al. (2009). Consequently, it may be that subjects were accessing different levels of knowledge. Monahan et al.'s listeners performed an AX discrimination task which possibly accesses an auditory level of discrimination, compared to the identification task used in this the Mattingley et al. (2015) study (based on Boomershine, Hall, Hume, & Johnson, 2008). The current study addresses this issue by using an AX discrimination task thereby allowing for the results to be more directly compared to those of Monahan et al. (2009).

In addition to work on loanword adaptation and perceptual epenthesis, work on vowel devoicing is important for this study. A study on the perception of spoken Japanese words by Cutler, Otake and McQueen (2009) shows that the vowel devoicing context makes speech segmentation and word recognition more difficult than the context which does not allow devoicing. As such, we might expect subjects to have difficulties perceiving vowels in devoicing contexts (between voiceless consonants). However, the role of vowel devoicing in perceptual vowel epenthesis is unlikely to be a decisive factor for illusory epenthetic vowels, since research has shown that Japanese listeners perceive the illusorily vowel [ɯ] even in non-devoicing contexts (Dupoux et al, 2011; Monahan et al., 2009). The current study will nonetheless investigate the effect of voicing type on perceptual epenthesis in the pseudo-stimuli.

The present study examines the role of the preceding consonant environment on perceptual and production epenthesis using a perception and production experiment with native speakers of Japanese as subjects. This study considers the effects of all vowel categories in Japanese on the perception and production of illicit consonant sequences. To study perception, a perceptual AX discrimination experiment was conducted which tested whether the sequence of first consonant and vowel influence Japanese-speaking listeners when discriminating between [aCVCa] and [aCCa] pairs. As mentioned before, there is a discrepancy regarding which vowel is perceived after the alveolar stop [d] between the studies of Monahan et al. (2009) and Mattingley et al. (2015). This is possibly due to methodological differences. Therefore, in this study, an AX discrimination test was used. The perceptual study provides empirical evidence for (i) the vowel quality perceived in word-medial consonant sequences by Japanese listeners when there is no medial vowel present, and (ii) the influence of the quality of preceding consonants. It is hypothesized that the quality of the preceding consonant influences listeners’ perception. In illicit word-medial consonant sequences listeners would be biased toward hearing the particular vowel that is expected according to Japanese phonotactics. Note that the specific prediction will be discussed in section 3.2.7. Using the AX task in this thesis enables us to examine potential differences in responses in two ways. It is possible to measure accuracy of the performance but also reaction time of the performance; that is, how quickly listeners responded. Even if listeners are able to discriminate between licit and illicit pairs and thus have high rates of accuracy, some pairs might be more difficult to discriminate than others. Reaction time gives us a way to measure these differences.

Next, a production study was conducted. I also explored the influence of the quality of the preceding consonant on epenthetic vowels in Japanese. In the production experiment, speakers were orthographically presented with the same word-medial consonant sequences used in the perception study. Since, with few exceptions, word-medial non-homorganic clusters do not occur in native Japanese words, speakers are expected to insert a vowel between the two consonants. I examined to what extent the response patterns in the experiments are consistent with Japanese native phonotactics. The proposed research is designed to provide a broader approach to the study of Japanese vowel insertion in order to obtain a clearer picture of the factors influencing it.

The structure of this thesis is as follows. Chapter 2 reviews relevant information regarding Japanese phonology and previous research on Japanese epenthesis. This review is followed by a discussion of previous research on perceptual and production epenthesis. The research questions and predictions are also presented. Chapter 3 presents the methodology and the results from Experiment 1: the AX discrimination experiment that tested the influence of native language phonotactics on perceptual epenthesis, followed by discussion. Chapter 4 presents the methodology and results from Experiment 2: speech production experiment, which investigates the preceding consonantal context and vowel duration that may influence the choice of epenthetic vowels, followed by discussion. Chapter 5 considers similarities and differences between the perceptual and production experiments, and discusses the implications of the results and presents the conclusion.

Chapter 2
Background

The aim of this chapter is to review existing studies on Japanese phonology and vowel epenthesis in Japanese. Since this thesis concerns whether the phonological properties of listeners’ native phonology influence speech perception and production, in section 2.1 I review aspects of Japanese phonology that are relevant to this thesis. Previous studies on loanword epenthesis in Japanese will be discussed in section 2.2. Previous studies on perceptual epenthesis in Japanese will be provided in section 2.3, while section 2.4 focuses on background for the production study. Section 2.5 presents the research questions and predictions.

2.1 Japanese Phonology

Modern Japanese has five phonemic vowels: high front [i], high back [ɯ], mid front [e], mid back [o], and low central [a] (e.g., Akamatsu, 2000; Shibatani, 1990; Tsujimura, 1996; Vance, 1987, 2008), as shown in Figure 2.1. As can be seen, Japanese vowels are relatively centralised in a vowel chart when compared to cardinal English vowels; Japanese vowels appear in boxes. According to Vance (2008), the Japanese high front vowel [i] is similar to the English high front vowel [i]. For the high back vowel [ɯ], the lips are compressed in careful speech, however, in normal speech tempo, compression of the lips is quite weak or totally absent. The tongue position of Japanese [ɯ] is quite centralised. The Japanese mid front vowel [e] is placed between English [e] and [ɛ]. The mid back vowel [o] is weakly rounded and falls between English [o] and [ɔ]. For the Japanese vowel [a], the tongue position is between English [a] and [ɑ]. The study of vowel openness in Japanese by Kawahara, Erickson and Suemitsu (in press) showed that front vowels [e] and [i] are more open than back vowels [o] and [ɯ], respectively.

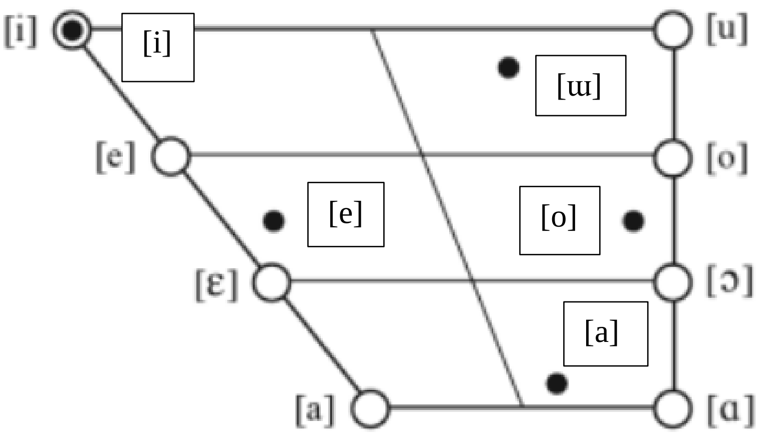

Figure 2.1. Vowel spaces for Japanese compared to cardinal English vowels (Vance, 2008: p54). Japanese vowels are in boxes.

Each Japanese vowel quality has a vowel length distinction between short and long vowels (Itô & Mester, 2003; Vance, 2008). The contrast between short and long vowels in minimal pairs is shown in (1).

(1) /obasan/ [ob**a**saɴ] ‘aunty’ vs. /obaasan/ [ob**a**ːsaɴ] ‘grandmother’
/ego/ [**e**go] ‘ego’ vs. /eego/ [**e**ːgo] ‘English’
/ozisan/ [odʑ**i**saɴ] ‘uncle’ vs. /oziisan/ [odʑ**i**ːsaɴ] ‘grandfather’
/joko/ [jok**o**] ‘side’ vs. /jokoo/ [jok**o**ː] ‘rehearsal’
/kɯki/ [k**ɯ**ki] ‘stem’ vs. /kɯɯɹki/ [k**ɯ**ːki] ‘air’

According to vowel duration studies by Han (1962, cited in Shoji & Shoji, 2014) and Yoshida (2006), the vowel [ɯ] is the shortest vowel in Japanese and the vowel [a] is the longest, as shown in Table 2.1. Among the five vowels, the high back vowel [ɯ] attracts less accent (Yoshida, 2006) and has the lowest sonority value (Hardison & Saigo, 2010; Katayama, 1998).

Table 2.1

Comparison of Two Studies in Terms of Duration of Japanese Vowels

Study	Order of vowels	Contexts
Han (1962)	Longest [a] > [e] > [o] > [i] > [ɯ] Shortest	Unknown
Yoshida (2006)	Longest [a] > [o] > [e] > [i] > [ɯ] Shortest	A female speaker of Tokyo Japanese, Accented vowel, Voiceless
	Longest [a] > [i] > [o] > [e] > [ɯ] Shortest	The same speaker, Unaccented vowel, Voiceless

In terms of vowel properties, the most remarkable phonetic characteristic of Japanese vowels is probably ‘vowel devoicing’. Table 2.2 provides examples of vowel devoicing. In Japanese, especially the Tokyo dialect of Japanese which is often regarded as ‘standard Japanese’, the high vowels /i, ɯ/ undergo devoicing when they occur between voiceless obstruents or in word-final position, and are not accented (Shibatani, 1990; Tsuchida, 1997; Vance, 1987). In some dialects of the Kyusyu area, the vowel is completely dropped (Shibatani, 1990).

Table 2.2

Examples of Vowel Devoicing

(a) /hashi/	‘chopsticks’	[haɕʃi̊]
(b) /akɯ/	‘open’	[akɯ̊]
(c) /kɯtɯ/	‘shoes’	[kɯ̊tsɯ]

According to Kondo (2005), the vowel devoicing process is virtually obligatory in phonetically and phonologically preferred environments even in accented syllables. High vowels in one devoicing environment in a word like (a) and (b) in Table 2.2 are almost constantly devoiced but not in consecutive devoicing environments such as (c). The vowel devoicing process makes the vowel shorter in duration, but the duration of the preceding consonant remains unaffected (Kondo, 2005).

In addition to vowels, some aspects of the phonology of Japanese consonants are important for the present study. Table 2.3 presents the consonantal phonemes of Japanese. The descriptions in Table 2.3 are based on Itô & Mester (2003) and Vance (1987).

Table 2.3

Japanese Consonants

	Bilabial	Alveolar	Alveolo-Palatal	Palatal	Velar	Uvular	Glottal
Plosive	p b	t d			k ɡ		
Nasal	m	n			(ŋ)	ɴ	
Flap		ɹ					
Fricative	(ɸ)	s z[1]	(ɕ) (ʑ)[2]	(ç)			h (ɦ)
Affricate		(ts) (dz)	(tɕ) (dʑ)				
Approximant				j	ɰ		

Some Japanese consonants vary allophonically depending on phonological environment; allophones appear in parentheses in Table 2.3. The alveolar consonants /t/, /d/, /s/, /z/ and the glottal fricative /h/ are palatalized when they occur before the high vowel /i/. However, Vance (1987) notes that [t] and [d] can occur before /i/, as in *asisuti* 'ice tea' and *dizeru* 'diesel'. He calls [ti], [di] sequences the 'innovative' variety while traditional allophonic CV sequences such as [tɕi], [dʑi] are called the 'conservative' variety. However, usage in the innovative variety is limited to loanwords. 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/. The phonological rules noted above are listed in (2).

(2) Distribution of Consonants

Palatalization	Examples		
/t/ → [tɕ] /_i	/mati/	[matɕi]	'town'
/d/→ [dʑ]/_i	/tidimɯ/	[tɕidʑimɯ]	'shorten'
/s/ → [ɕ] /_i	/hasi/	[haɕi]	'bridge'
/h/ → [ç] /_i/	/hito/	[çito]	'human'
Changing in place of articulation			
/h/→ [ɸ]/_ ɯ	/hukɯ/	[ɸɯkɯ]	'clothes'

---

[1] In modern Japanese, /z/ and /dz/ do not contrast with each other (Itô & Mester, 2003).

[2] According to Vance (1987), /ʑ/ is hardly produced in modern Japanese. It has merged with /dʑ/.

Affrication

/t/ → [ts] /_ ɯ /katɯ/ [katsɯ] ‘win’
/d/ →[dz]/_ ɯ /tedɯkɯri/[tedzɯkɯɹi] ‘handmade’

In addition to individual sounds, the way that Japanese combines sounds in sequences is important for this thesis research. Japanese syllable structure is very simple, consisting most often of a consonant-vowel (CV) or vowel sequence, with only a nasal or the first part of a geminate consonant allowed in coda position (Tsujimura, 1996). This is shown in (3). Otherwise consonant clusters are illicit in word initial, medial and final positions.

(3) [sim.bɯɴ] (CVCCVC) ‘newspaper’
[gak.koo] (CVCCVV) ‘school’

2.2 Previous Studies on Loanword Epenthesis in Japanese

In many languages, vowel epenthesis is typically used as a repair strategy for loanwords which contain syllable codas and consonant clusters that are illicit in source languages (Fleischhacker, 2001; Hall, 2011; Kabak & Idsardi, 2007; Kang, 2011; Uffmann, 2006). As is the case with many other languages, the Japanese language includes vowel epenthesis as a syllable modification strategy (Hirayama, 2003; Itô, 1989; Smith, 2006; Kubozono, 2015). For example, the English word ‘pipe’ [paɪp] becomes [paɪpɯ] in Japanese through the insertion of the vowel [ɯ] in word-final position (Hirayama, 2003); recall that the only consonant that can occur word-finally in Japanese is [ɴ].

The phonological adaptation of loanwords in Japanese has been investigated by many scholars (Hirayama 2003; Irwin, 2011; Katayama, 1998; Kubozono, 2001, 2015; Lovins 1975; Otaki, 2012; among others), and it has been found that the choice of epenthetic vowel is constrained by the quality of preceding consonant. These studies all agree that three different vowels [i, o, ɯ] can be inserted depending on the quality of the preceding consonant. Irwin (2011) investigated the history of Japanese loanwords. His data came from written texts, from the sixteenth-century to the present, and the donor languages include English among other languages. According to Irwin, among the five Japanese vowels, the high vowels [i] and [ɯ] are the most common epenthetic vowels. However, in the majority of situations, he claims that native speakers of Japanese are more likely to insert [ɯ] than [i]. The high front vowel [i] is typically inserted after the palato-alveolar affricates [tʃ], [dʒ] and the voiceless velar [k]. Irwin states that [ʃ] and [ʒ] from a donor language sometimes trigger [i], however, words adapted with [ʃi] and [ʒi] often have doublets with an epenthetic [ɯ]. For example, the English word ‘sash’ [sæʃ] becomes [saɕɕi] or [saɕɕɯ] in Japanese through the insertion of the vowel [i] or [ɯ]. In addition to epenthetic [i] and [ɯ], Irwin found that the mid back vowel [o] is epenthesized after alveolar stops [t, d]. The reason for the insertion of [o] after alveolar stops is likely that [tɯ], [dɯ], [ti], and [di] do not occur in the native Japanese syllable inventory. Although these later sequences are becoming permissible sequences in contemporary loanword pronunciations, the high back [ɯ] has not completely replaced [o] (Irwin, 2011). For example, with the English word, *straight* [streit], the Japanese borrowing is [sɯ.to.ree.to], but not *[sɯ.tɯ.ree.tɯ]. The epenthetic vowels based on research from Japanese loanwords are summarised as follows:

(4) Epenthetic vowels

(i) Ø → [i]/ {[tʃ, dʒ], [ʃ, ʒ] } _ or [k]_[3]
(ii) Ø → [o] / [t , d] _
(iii) Ø → [ɯ] / in all other contexts

In terms of distribution, some scholars state that the epenthetic vowel [ɯ] is the default vowel, the most unmarked and perceptually least salient (Hirayama, 2003; Shoji & Shoji, 2014; Kubozono, 2015). This could be because the vowel [ɯ] is phonetically the shortest vowel, and the most susceptible to weakening and deletion (Sagisaka & Tokuhara 1984 as cited in Irwin, 2011; Kubozono, 2015). As for the insertion of [i] after [tʃ], [dʒ], it allows the borrowed words to keep the palatal nature of the words in the source language (Hirayama, 2003). Additionally, the front vowel [i] shares similar articulatory and perceptual properties with these consonants (Kubozono, 2015). As mentioned above, [o] insertion is likely due to absence of alveolar stop + [i], [ɯ] sequences. Kubozono states 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 change these consonants to affricates [ts], [dz] due to a native assimilation rule (see § 2.1). Inserting [o] allows for the distinction to be maintained between [t] and [ts], which are distinctive phones in other

---

[3] Irvin (2011) states that velar fricative /x/ triggers the epenthetic vowel /i/ when donor languages were German and Dutch. Also the epenthetic vowel /i/ occurs after retroflex fricative /ʂ/ in Russian

languages including English. For example, the distinction between *ruuto* 'root, route' and *ruutsu* 'roots' can be maintained in Japanese by inserting different epenthetic vowels.

2.3 Studies of Perceptual Epenthesis

The influence of native speech experience and the native phonetic system on speech perception and production has been well investigated with different theories being proposed. Several studies discuss how native-language phonotactics influence the perception of non-native sounds. This research indicates that stimuli with non-native sound sequences are generally assimilated perceptually to licit sequences in the listener's native language (Best, 1994, 1995; Best & Strange, 1992; Dupoux et al., 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Dupoux et al., 2011; Hallé, Segui, Frauendelfer & Meunier, 1998; Kobak, 2003; Kabak & Idsardi, 2007).

For example, Dupoux et al. (1999) carried out a cross-linguistic perception study of consonant clusters (CC), comparing Japanese listeners with French listeners. While Japanese allows only a nasal or the first part of a geminate consonant in coda position, French allows a range of CC sequences. In order to investigate effects of native language, they created six audio files with differing lengths of the middle vowel [ɯ][4], yielding a continuum of stimuli from full vowel duration to no vowel. The participants were asked to judge whether a medial vowel [ɯ] was present or not in pseudo-words [VC(V)CV]. The study shows that when no medial vowel was present in the stimuli, native speakers of Japanese were highly likely to perceive an illusory epenthetic vowel [ɯ] in stimuli containing consonant sequences that are illicit in their native language. Japanese listeners also had difficulty discriminating between illicit (VCCV) and licit (VCVCV) pairs (e.g., [ebzo]-[ebɯzo]) in an ABX discrimination test, while French listeners did not. Their interpretation of this finding is that the speech perception process is constrained by phonotactic knowledge. However, their study was designed for listeners to perceive an epenthetic vowel [ɯ] in word-medial consonant clusters (e.g., [abɯge], [agɯmi], [akɯmo]) in order to investigate the role of phonotactics on perception. That is, they did not include two other distinct preceding consonant environments in their stimuli nor did they consider all vowel categories in Japanese.

---

[4] Dupoux et al. (1999, 2011) use an epenthetic [u] in transcription. In this thesis, [ɯ] will be used for the high back vowel.

In a follow-up cross-linguistic study, Dupoux et al. (2011) examined the perceptual epenthesis effect using three types of nonsense words as original stimuli: (1) VCCV, (2) VC[ɯ]CV and (3) VC[i]CV. They created seven audio files with differing lengths of the middle vowel for type (2) and (3), yielding a continuum of stimuli from full vowel duration to no vowel. As in their previous study, Japanese listeners showed a strong perceptual epenthesis effect in illegal consonant clusters. The epenthetic vowel was predominantly [ɯ] when no vowel was present, both when there was no original vowel as in (1), and when the vowel [ɯ] in (2) had been removed. It should be noted that, even when the vowel [i] in (3) had been removed, the consonant clusters with coarticulatory cues from [i] elicited [ɯ]-responses in 20% of the responses, and [i]-responses in 34.5 % of the responses.

The study by Monahan, Takahashi, Nakao, and Idsardi (2009) examined the relationship between the mid back vowel [o] and preceding phonological environments. Recall that in Japanese phonotactics, [o] can follow the alveolar stops [t d], while [ɯ] and [i] cannot. They analysed perceptual epenthesis of the illusory vowels [ɯ][5] and [o] after [t d] and velar stops [k g] between native speakers of English and Japanese using an AX discrimination task, in which a participant hears a pair of stimuli (A and X) in a trial and decides whether X is the same as A, or different. They interpret their results as suggesting that Japanese listeners did not perceive an illusory epenthetic [ɯ], the most common epenthetic vowel in Japanese, nor the contextually predicted vowel [o] after coronal consonants; rather, Japanese listeners were able to discriminate, for example, [etoma] from [etma], and [etɯma] from [etma], respectively. However, Japanese listeners did perceive an illusory vowel [ɯ] after velar stops. That is, Japanese listeners performed significantly more poorly in discriminating between [egɯma - egma] and [ekɯma - ekma] than English listeners did. These findings suggest that an illusory epenthetic [ɯ] does not always occur with non-native consonant sequences and the perceptual illusory vowel effect may be influenced by the interaction of epenthetic contextual environment and vowel category. Japanese listeners are likely to be sensitive to the phonological environment in which an epenthetic vowel occurs. Therefore, Monahan et al. (2009) suggest that native language phonology alone cannot explain the perception of non-native speech.

---

[5] Monahan et al. (2009) use an epenthetic [u] in transcription.

Mattingley, Hume, and Hall (2015) examined the influence of preceding consonants on the perception of word-medial consonant sequences by native speakers of Japanese. The stimuli included consonant sequences that do not occur in Japanese (i.e., non-homorganic consonant clusters). More specifically, the stimuli consisted of pseudo-words of the form $[aC_1(V)C_2a]$ with consonants selected from the set of voiced obstruents {b, d, g, dʑ} where C1≠C2, giving a total of 12 different consonantal combinations. (V) represented one of five Japanese vowel qualities {a, e, i, o, ɯ} or no vowel (e.g., [abada], [ageba], [agba]). Listeners were asked to identify the vowel heard between the consonants of the pseudo-word. The results suggested that, to some extent, perceptual epenthesis in Japanese is constrained by native phonotactics. It was found that when no vowel was present, [ɯ] was perceived as the epenthetic vowel after [b] and [g], and [i] was predominantly selected as the epenthetic vowel after the palatal affricate [ʥ]. However, Japanese listeners did not make use of the mid back vowel [o] after the alveolar stop [d]. In this context the vowel was predominantly identified as [ɯ], which raises the question of whether the domain of the default vowel [ɯ] is spreading to beyond what would be predicted by Japanese native phonotactics. This finding conflicts with the findings of Monahan et al. (2009), which may be due to the use of different methodologies. Monahan and colleagues investigated the relationship between perceptual epenthesis and native language phonology using an AX discrimination task, whereas an identification task was used in Mattingley et al. (2015). According to Gerrit & Schouten (1998), participants are accessing the phonemic level of knowledge or linguistic knowledge during categorical perception (e.g., using an identification task). Werker & Logan (1985) also argue that AX discrimination tasks access acoustic information rather than higher level phonetic or phonological knowledge. Thus, the AX discrimination task might require an acoustic level of knowledge to discriminate differences. On the other hand, participants would need to access the phonological level of knowledge in the identification task. This is an important issue for the proposed research.

Some scholars suggest that perceptual salience and similarity are crucial factors in loanword adaptation (Fleischhacker, 2001; Kang, 2003; Kenstowicz, 2007; Shinohara, 1997; Steriade, 2001). For example, Fleischhacker (2001) argues that perceptual similarity plays a fundamental role in loanword adaptation since in some languages, the location of the epenthetic vowel varies depending on its auditory similarity to the input. For example, in

English loanwords in Hindi,[6] prothesis occurs *before* voiceless sibilant+stop (ST) clusters (e.g., [ɪskul] ‘school’), whereas vowel epenthesis occurs in the *middle* of obstruent+sonorant (OR) clusters (e.g., [pɪlɪz] ‘please’). In addition, when a cluster is STR such as in the English word *screw*, prothesis occurs (e.g., [ɪskru]). Experimental studies show that native English speakers judged these distinctive epenthesis patterns to be more similar to their non-epenthesized inputs, respectively.

Steriade also supports a view that loanword adaptation is largely driven by perceptual factors. According to Steriade (2001, 2008), speakers have knowledge of the perceptibility of phonological contrast. In terms of epenthesis, Steriade argues that the choice of epenthetic segment is based on speakers’ judgments of relative similarity between an individual segment and no segment (i.e., Ø). That is, the segment most confusable with Ø is expected to be inserted in a given context. Steriade uses vowel epenthesis as evidence that schwa is cross-linguistically preferred since the vowel is arguably the closest to no epenthesis at all due to its short duration and variability in quality compared to other vowels.

Thus, it is often argued that the quality of the epenthetic vowel should be the one which has low salience or a minimal perceptual/auditory difference from the source form. However, why should low salience be of concern in epenthesis? This perspective may be in part due to the observation that one motivation of vowel epenthesis is to facilitate communication (Hume, Hall, Wedel, Ussishkin, Adda-Dekker, & Gendrot, 2013). As mentioned above, the phonological process of vowel epenthesis typically breaks up unfamiliar sequences of consonants. As a consequence, vowel epenthesis might make it easier for non-native speakers to perceive and produce non-native sounds than it would be with the original form or other modifications. This is consistent with Kuijpers, Donselaar & Cutler (1996) who show that in an experiment of auditory word recognition, words with epenthesis were processed more accurately and rapidly than words with deletion. In another study, vowel epenthesis facilitates the perceptibility of the liquid consonant in a liquid-obstruent cluster in spoken Dutch words (Donselaar, Kuijpers & Cutler, 1999). In their lexical decision and phoneme identification tasks, listeners detected the phoneme targets in the words with epenthesis faster than forms without epenthesis, even though the form without epenthesis was the more canonical form of the word. From their findings, Donselaar et al.

---

[6] Fleischhacker cited Hindi data from Broselow (1992) and Singh (1985).

suggest that speakers are motivated to epenthesize vowels to help the listeners. Additionally, Hume (2016) argues that the quality of epenthetic vowel is the one that “contributes the most to successful message transmission while having the least negative impact on system efficiency” (p.7). Thus, low salience in epenthesis is motivated by both phonetic and cognitive perspectives.

2.4 Production Studies of Epenthesis

The influence of the native phonetic system on speech production has been the subject of considerable debate in the literature; however, most studies of non-native speech production that are related to epenthesis have focused on second language learning perspectives (e.g. Broselow & Finer, 1991; Lin, 2003; Sperbeck, 2010). Although a large and growing body of literature has investigated vowel epenthesis in Japanese, there are few empirical production studies of epenthesis from phonetic and phonological perspectives.

One study by Kobayashi (2000) makes a number of observations about English loanwords in Japanese. He states that if one of the five Japanese vowels is inserted after the last consonant in English words like, ‘cup’, ‘net’, ‘kick’, ‘cab’ ‘head’ and ‘dog’, a high vowel [ɯ] is more appropriate to insert than any other vowel because it maintains the closest link between the underlying lexical representation and surface form. According to Kobayashi, this vowel insertion is related to the articulation of the tongue and lips. He claims that the vowel [ɯ] is the most neutral vowel sound in the Japanese vowel inventory since the tongue and lips move less than for the other four vowels. Therefore, [ɯ] is claimed to be the vowel that is easier and faster to produce and process. However, there are two other vowels which are used in epenthesis, depending on the preceding consonant. The high front vowel [i] is typically inserted after the palato-alveolar affricates [tʃ], [dʒ]; it is phonetically natural to insert the front vowel [i] which shares a similar place of articulation with these consonants. To explain the insertion of [o] after alveolar stops, he suggests it is to maintain the features of the preceding consonant as [+alveolar, +plosive]. As noted above, if the high vowels [i] and [ɯ] were inserted after alveolar stops, the stops may change to affricates. In fact, in some loanwords, the epenthetic vowel [ɯ] is used after the voiceless alveolar stop [t], with [tɯ] becoming an affricate, as in *tsɯin* ‘twin’, *tsɯii* ‘tree’.

In another study, Shoji and Shoji (2014) examined patterns of vowel epenthesis in Japanese loanwords from English using writing production experiments. They hypothesized that the high back unrounded vowel [ɯ] would be epenthetic in most of Japanese loanwords in their experiment. They state, ‘the epenthetic vowel [ɯ] is the most unmarked and perceptually the least salient among Japanese vowels’ (p.3). Other vowels [i] and [o] are hypothesized to be context-dependent epenthetic vowels. The palato-alveolar affricates, [tʃ] and [dʒ], in the source words would be pronounced as [ʨ] and [ʥ] in the process of loanword adaption. After these consonants, [i] is typically inserted because the consonant and vowel share similar articulations. The vowel [o] typically occurs after alveolar stops [t d] since the selection of vowels is constrained by the preceding consonants, as discussed earlier. It is noteworthy that Shoji and Shoji (2014) found epenthetic vowels other than those that had been hypothesized. For example, there were 23.7% of [ɯ]-responses after [tʃ dʒ] and 32.7% after [t d] in word-initial consonant clusters. This raises the question of whether native speakers of Japanese are making more use of the high back vowel [ɯ] irrespective of the preceding consonantal environment.

More recently, Yazawa, Konishi, Hanzawa, Short & Kondo (2015) 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 analysed speech corpus data of Japanese participants reading the Aesop fable “The North Wind and the Sun”. The results showed that the higher the proficiency, the fewer the epenthetic vowels. However, 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 d], [i] after [tʃ dʒ], and [ɯ] when it follows any other consonant. Note that these findings are different from the perception studies we discussed above. That is, research on perceptual epenthesis in Japanese suggests that Japanese listeners did not always perceive the mid back vowel [o] after the alveolar stop (Mattingley et al, 2015; Monahan et al. 2009).

2.5 Research Questions and Predictions

The overall goals of this thesis are to investigate, in relation to vowel epenthesis in Japanese: (1) the influence of native phonotactics on the perception and production of epenthesis in non-native clusters; (2) the phonetic properties of the epenthetic vowels; and (3) similarities and differences between the behaviour of subjects in the perceptual and production experiments. Based on previous literature, since word-medial [CC] does not occur in native Japanese words, with few exceptions, it is expected that Japanese listeners would perceive and insert a vowel between the two consonants. The current study predicts that listeners are more likely to be less accurate and/or be slower discriminating contrasting pairs [aCVCa] and [aCCa] when the medial vowel is the particular vowel that is expected according to the quality of the preceding consonant context, as shown in Table 2.4. This prediction is made under the assumption that when a vowel category, expected according to Japanese phonotactics, is presented to listeners, they will have a greater tendency to perceive an illusorily epenthesized vowel in [aCCa]. As for speech production, the current study predicts that the quality of the preceding consonant will influence the choice of epenthetic vowel in a way similar to that predicted for perception (i.e., consistent with the language's phonotactic patterns).

Table 2.4

Predictions

Preceding Consonant	Perception		Production
	AX Accuracy	AX Reaction Time	
Labial	more errors with [ɯ]	slower with [ɯ]	mostly [ɯ]
Velar	more errors with [ɯ]	slower with [ɯ]	mostly [ɯ]
Alveolar	more errors with [o]	slower with [o]	mostly [o]
Palatal	more errors with [i]	slower with [i]	mostly [i]

Chapter 3
Perception Experiment

3.1 Introduction

Research on perceptual epenthesis in Japanese has revealed high back [ɯ] to be the vowel commonly perceived in illicit consonant sequences (Dupoux et al, 1999: Dupoux et al 2011; Monahan et al. 2009). However, as noted above, loanword studies suggest that there are three epenthetic vowels, which reflect phonotactic restrictions on certain consonant + vowel sequences (Hirayama, 2003; Irwin, 2011; Katayama, 1998; Shoji & Shoji 2014). Expanding previous perception studies, this thesis investigates the extent to which perceptual epenthesis in Japanese is also constrained by the language’s phonotactic patterns. In particular, I seek to determine to what extent the preceding consonant influences perceptual epenthesis, reflecting native phonotactics.

3.2 Methodology

3.2.1 Stimuli

The stimuli consisted of pseudo-words with a consonant cluster in the middle of the word. The structure of the pseudo-words was [aC_1(V)C_2a] where (V) was either one of the five Japanese vowels {a, e, i, o, ɯ} or no vowel. The consonants were selected from either the set of voiced obstruents {b, d, ɡ, d͡ʑ} or their voiceless counterparts {p, t, k, t͡ɕ}, and C1≠ C2. The initial and final vowels of the pseudo-words were always [a] in order to have a uniform context across all stimuli (e.g., /abada/, /ageba/, /akta/).

The stimuli for the perception experiment were collected by recording a 23-year-old male native speaker of Japanese reading the pseudo-words. He was born in Japan and has lived in New Zealand since he was eight. He is fluent in both Japanese and English and has had no linguistic training. He spoke Japanese at home when he lived in NZ. A Tascam HD-P2 audio recorder with 44,100 samples/s, 16 bit/s and Beyerdynamic head-mounted microphone were used for recording, which took place in a sound-attenuated room at the University of Canterbury.

The stimuli were produced in the carrier sentence written in Japanese *hiragana* characters, *Koremo* _____ *desu*. ‘This is ____, too.’ PowerPoint slides were used to display stimuli with one slide for each sentence. To ensure the speaker identified and pronounced the stimuli correctly, sample Japanese words were given to illustrate each vowel and consonant combination in a practice section. The key words for each vowel that were presented to him before the recording started were: [a] *a*ki ‘autumn’, [e] *e*ki ‘station’, [i] *i*ki ‘breath’, [o] *o*ki ‘offing’, and [ɯ] *u*ki ‘bob’. The IPA symbols did not appear on the screen. Only one pseudo-word corresponded to an actual word in Japanese: [akita], which is a prefecture name in Japan. The speaker was asked to say each stimulus and the carrier sentence as naturally as possible when it appeared on the computer screen. He repeated each sentence three times and was asked to maintain the same tempo across readings.

Production recordings were analysed acoustically using Praat phonetic software (Boersma & Weenink, 2014) (hereafter Praat). For each stimulus type, two recordings were manually selected from the three repetitions, giving consideration to clarity of production and the duration of the words. Finally, the target words were extracted from the carrier sentence.

Six items of the form aCCa needed to be re-recorded due to problems with the sound quality. The re-recorded items had higher intensity than the other stimuli recorded in the earlier session. In order to ensure consistency across stimuli, the intensity of the re-recorded items (6*two recording files) was modified using Praat. This was done by determining the mean intensity of the ten aCVCa-forms recorded earlier and modifying the intensity of the other six stimuli to match.

There were 60 full vowel sound files (12 consonant combinations * 5 vowels), which were used as the control stimuli. There were also 12 original no-vowel files in aCCa-forms, for a total of 72 audio files for each voicing type, as shown in Table 3.1.

Table 3.1

Pseudo-words with and without Consonant Clusters and Number of Experimental Stimuli in Each Condition

	VC$_1$C$_2$V	**VC$_1$VC$_2$V stimuli**					
	no vowel	**[a]**	[e]	[i]	[o]	[ɯ]	
	aCCa	**aCaCa**	**aCeCa**	**aCiCa**	**aCoCa**	**aCɯCa**	
C$_1$ =bilabial [b]	abda abga abdʑa	abada abaga abadʑa	abeda abega abedʑa	abida abiga abidʑa	aboda aboga abodʑa	abɯda abɯga abɯdʑa	
C$_1$ =alveolar [d]	adba adga addʑa	adaba adaga adadʑa	adeba adega adedʑa	adiba adiga adidʑa	adoba adoga adodʑa	adɯba adɯga adɯdʑa	
C$_1$=velar [g]	agba agda agdʑa	agaba agada agadʑa	ageba ageda agedʑa	agiba agida agidʑa	agoba agoda agodʑa	agɯba agɯda agɯdʑa	
C$_1$=alveo-palatal [dʑ]	adʑba adʑda adʑga	adʑaba adʑada adʑaga	adʑeba adʑeda adʑega	adʑiba adʑida adʑiga	adʑoba adʑoda adʑoga	adʑɯba adʑɯda adʑɯga	
Subtotal	12	12	12	12	12	12	72
C$_1$ =bilabial [p]	apta apka aptɕa	apata apaka apatɕa	apeta apeka apetɕa	apita apika apitɕa	apota apoka apotɕa	apɯta apɯka apɯtɕa	
C$_1$ =alveolar [t]	atpa atka attɕa	atapa ataka atatɕa	atepa ateka atetɕa	atipa atika atitɕa	atopa atoka atotɕa	atɯpa atɯka atɯtɕa	
C$_1$=velar [k]	akpa akta aktɕa	akapa akata akatɕa	akepa aketa aketɕa	akipa akita akitɕa	akopa akota akotɕa	akɯpa akɯta akɯtɕa	
C$_1$=alveo-palatal [tɕ]	atɕpa atɕta atɕka	atɕapa atɕata atɕaka	atɕepa atɕeta atɕeka	atɕipa atɕita atɕika	atɕopa atɕota atɕoka	atɕɯpa atɕɯta atɕɯka	
Subtotal	12	12	12	12	12	12	72
Grand Total	24	24	24	24	24	24	144

3.2.2 Acoustic Characteristics of the Stimuli

In order to determine the acoustic characteristics of the stimuli, vowels from the [aCVCa] stimuli were analysed using Praat. The duration of each target vowel was measured and the mean values for F1, F2 and F3 for the stimulus vowels were extracted using a Praat script. All measurements were taken at the midpoint of the marked segment. All extracted formants were checked manually to ensure the validity of the values.

The number of vowels with duration measured differed in the voiced and voiceless consonantal contexts. There are two tokens of each word. All 120 tokens of vowels in the voiced condition were used. In the environment of preceding voiceless consonants, /i/ and /ɯ/ underwent devoicing between the two consonants. Measuring the duration of the devoiced high vowels [i] after the voiceless affricate [tɕ] was technically difficult because the boundary between the voiceless vowel and preceding consonant was not clear. Therefore, the duration of the six tokens containing the vowel [i] with the affricate consonant were excluded, leaving only 114 vowel tokens in the voiceless condition.

Figure 3.1 (two plots) shows differences in duration across vowel qualities. In these plots, if the notches of any two box plots do not overlap, the two medians tend to be significantly different with 95% confidence level (McGill, Tukey, & Larsen, 1978). The mean duration of [ɯ] was the shortest in length among the five vowels; it was 77.16 ms (median = 80 ms) in the voiced consonant context and 40.91 ms (median = 38.5 ms) in the voiceless consonant context.

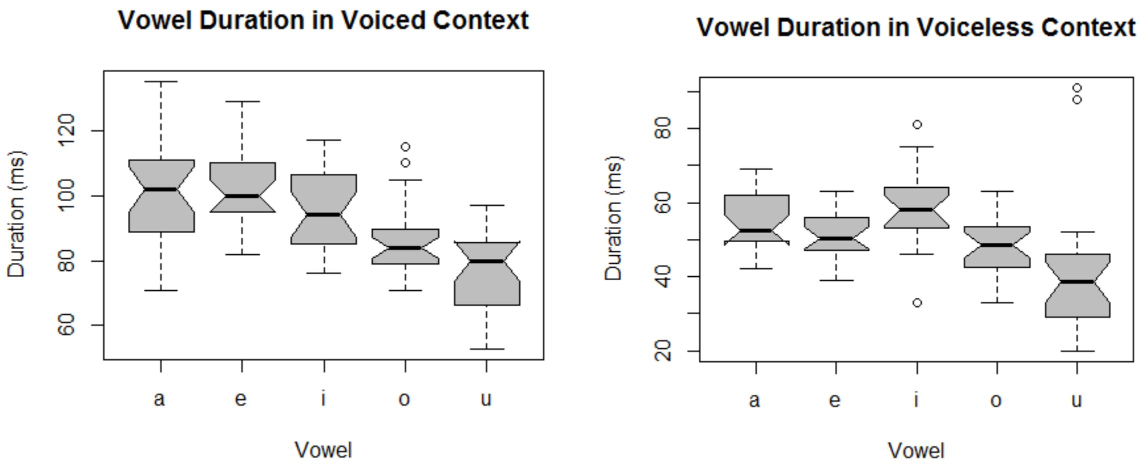

Figure 3.1. Boxplots of five vowels for the speaker in voiced and voiceless contexts.

An analysis of variance (ANOVA) showed an effect of vowel for the voiced condition [$F(4, 115) = 16.5$, $p < .001$]. A Tukey post-hoc test showed that there was a significant effect of vowel on duration among [ɯ] and other vowels (except [o]): [a] ($p < .001$), [e] ($p < .01$), [i] ($p < .001$), [o] ($p = .151$). For the voiceless context, ANOVA showed an effect of vowel

[*F*(4, 109) = 8.432, *p* < .001] and there was also a significant effect of vowel on duration among [ɯ] and other vowels (again, except [o]): [a] (*p* <.001), [e] (*p* < .05), [i] (*p* < .001), [o] (p = 0.144).The finding that the high vowel [ɯ] is the shortest vowel is consistent with vowel duration studies by Han (1962, cited in Shoji & Shoji, 2014) and Yoshida (2006).

The stimulus vowels' mean F1/F2/F3 values are shown in Table 3.2. Formant values were not normalised. Since there were no voicing bars for the devoiced high vowels /i/ and /ɯ/ between voiceless consonants, the F1/F2 formants extracted automatically by Praat script were not reliable. Therefore, the high vowel F1/F2 formants in the voiceless context were excluded for the plotting figure (Figure 3.2).

Table 3.2

Mean F1/ F2/F3 Formant Values and Standard Deviations for Voiced and Voiceless Contexts

Number of Tokens	Vowel	F1		F2		F3	
		mean	SD	mean	SD	mean	SD
48	[a]	617.1	53.2	1428.9	129.7	2448.6	132.1
48	[e]	429.4	29.2	1972.5	101.0	2724.0	104.1
24	[i]	283.4	17.1	2334.1	90.1	3157.3	171.3
48	[o]	439.6	23.8	1000.8	143.2	2789.1	214.6
24	[ɯ]	334.0	15.3	1458.1	195.4	2710.1	111.2

Non-normalised ellipse plots in Figure 3.1 show the overall F1/F2 spaces with mean values and 2.0 standard deviations for each lexical vowel from the speaker. Figure 3.2 shows that the F1/F2 space for the stimulus vowels is consistent with the Japanese vowel space presented in Vance (2008). The high front vowel [i] is higher and fronter than other vowels. The other high vowel [ɯ] is quite centralised. The mid front vowel [e] and mid back vowel [o] are similar in terms of height. The vowels [a] and [ɯ] are almost equal in backness (see also Chapter 2 for details of vowel space in Japanese).

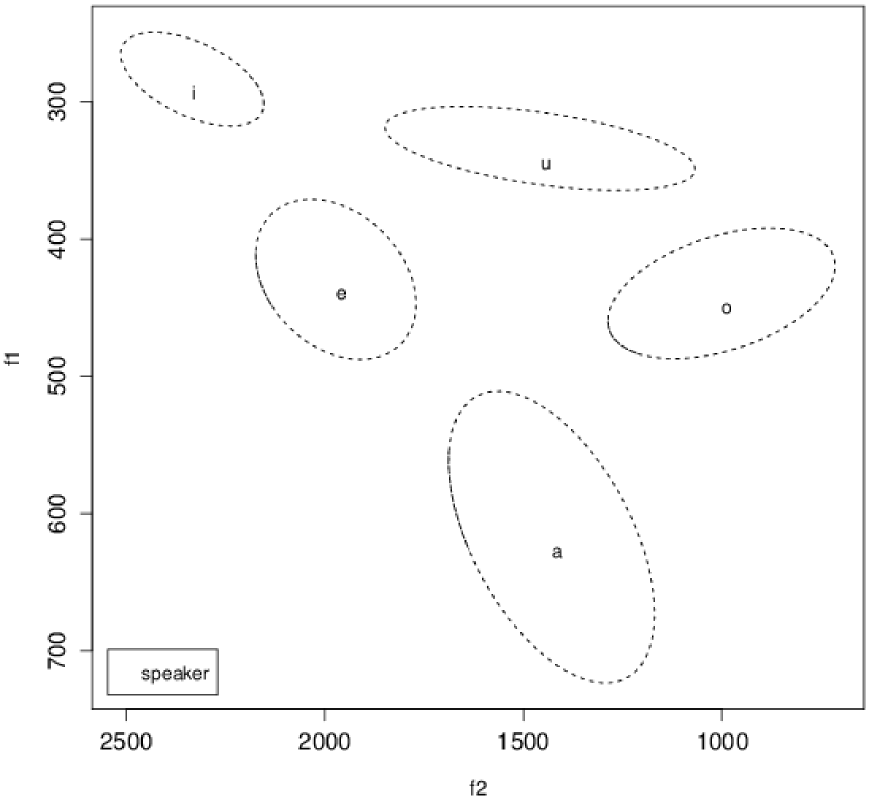

Figure 3.2. F1 and F2 ellipse plots showing means and 2.0 standard deviations from the mean for the speaker (note the values of devoiced high vowels /i/ and /ɯ/ were excluded.)

Figure 3.3 shows two examples of waveforms and spectrograms of the stimuli ‘aputa’ [apɯta] and ‘apta’ [apta] produced by the speaker.

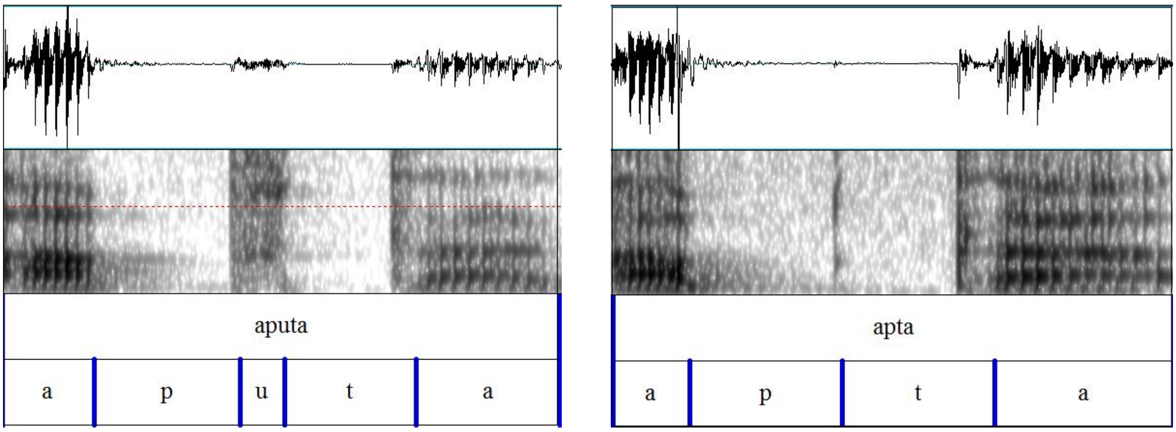

Figure 3.3. Spectrogram and waveform for the productions of ‘aputa’ [apɯta] and ‘apta’ [apta].

Differences can be observed between the two stimuli in the figures. In the waveform of ‘aputa’, the vowel [ɯ] is observed after a tiny burst of [p], nevertheless the waveform shows no clear periodic waves. The spectrogram does not show a voice bar. These observations indicate that the vowel is devoiced. On the other hand, the waveform and spectrogram of ‘apta’ show a tiny burst and there is no vowel between the consonants. It should be noted that some stops in the current study’ stimuli have burst releases in the VCCV context while some do not.

3.2.3 Pairs of Words

The perceptual experiment employed a same-different AX discrimination task in which a participant hears a pair of stimuli (A and X) in a trial and decides whether X is the same as A, or different.

In the current experiment, there are two types of stimuli, A and B. While A is a licit sequence stimulus [aC_1VC_2a], B is an illicit consonant sequence stimulus [aC_1C_2a] in Japanese phonotactics. Items were presented in four types of pairs: <AB>, <BA>, <AA> and <BB>. The *different* pairs are licit-illicit pairs, <AB> and <BA>; [aC_1VC_2a] vs. [aC_1C_2a], [aC_1VC_2a] vs. [aC_1C_2a]. These pairs differed in whether they had consonant sequences or not. The *same* pairs are either licit pairs <AA> or illicit pairs <BB>; [aC_1VC_2a] vs. [aC_1VC_2a], [aC_1C2a] vs. [aC_1C_2a]. Identical recordings were not used for the *same* pairs. For the licit pairs, the medial vowel (V) is the same across the two stimuli. In all pairs, V is one of the five Japanese vowels {a, e, i, o, ɯ}, and the consonants are the same for both stimuli in a given pair. For example, while [abada] vs. [abda] and [abda] vs. [abada] are *different* pairs, [abada] vs. [abada] and [abda] vs. [abda] are *same* pair stimuli. A sample set of stimuli is shown in (1).

(1) Sample of AX discrimination stimuli: C_1= [b], V= [a], C_2= {d, g, dʑ}

(a) Different pairs: <AB> [aC_1VC_2a] vs. [aC_1C_2a]; <BA> [aC_1C_2a] vs. [aC_1VC_2a]

<AB> [abada] vs. [abda]
<BA> [abda] vs. [abada]

<AB> [abaga] vs. [abga]
<BA> [abga] vs. [abaga]

<AB> [abadʑa] vs. [abdʑa]

<BA> [abdʑa] vs. [abadʑa]

(b) Same pairs : <AA> [aC_1VC_2a] vs. [aC_1VC_2a]; <BB> [aC_1C_2a] vs. [aC_1C_2a]

<AA> [abada] vs. [abada]

<BB> [abda] vs. [abda]

<AA> [abaga] vs. [abaga]

<BB> [abga] vs. [abga]

<AA> [abadʑa] vs. [abadʑa]

<BB> [abdʑa] vs. [abdʑa]

Participants listened to <AA>, <AB>, and <BA> pairs one time each, and <BB> pairs were presented five times each, as shown in Table 3.3. Thus, the number of same and different pairs were balanced. Each participant listened to 240 pairs for each voicing type. There were thus 480 trials altogether for each participant.

Table 3.3

Sample Items using /a/ for Each Participant in Voiced Consonant Contexts

	Same Pairs				Different Pairs			
C_1	AA		BB		AB		BA	
bilabial [b]	ba-ba	abada-abada abaga-abaga abadʑa-abadʑa	b-b	abda-abda abga-abga abdʑa-abdʑa	ba-b	abada-abda abaga-abga abadʑa-abdʑa	b-ba	abda-abada abga-abaga abdʑa-abadʑa
coronal [d]	da-da	adaba-adaba adaga-adaga adadʑa-adadʑa	d-d	adba-adba adga-adga addʑa-addʑa	da-d	adaba-adba adaga-adga adadʑa-addʑa	d-da	adba-adaba adga-adaga addʑa-adadʑa
velar [g]	ga-ga	agaba-agaba agada-agada agadʑa-agadʑa	g-g	agba-agba agda-agda agdʑa-agdʑa	ga-g	agaba-agba agada-agda agadʑa-agdʑa	g-ga	agba-agaba agda-agada agdʑa-agadʑa
alveo-palatal [dʑ]	dʑa-dʑa	adʑaba-adʑaba adʑada-adʑada adʑaga-adʑaga	dʑ-dʑ	adʑba-adʑba adʑda-adʑda adʑga-adʑga	dʑa-dʑ	adʑaba-adʑba adʑada-adʑda adʑaga-adʑga	dʑ-dʑa	adʑba-adʑaba adʑda-adʑada adʑga-adʑaga
Subtotal	12		12		12			12
	*5 vowels {a, e, i, o, ɯ}		no vowel but 5 repetitions		*5 vowels {a, e, i, o, ɯ}		*5 vowels {a, e, i, o, ɯ}	
Total	60		60		60		60	

3.2.4 Participants

The participants were 21 native speakers of Japanese (16 female, 5 male), living in Christchurch, New Zealand, who were tested at the University of Canterbury. They were recruited at local language schools and via the researcher's acquaintances. One participant was excluded from the analysis due to the fact that she had lived in the United States of America for a total of nine years from the ages of two to fifteen. Because her English proficiency might have influenced the experiment tasks, her data were excluded.

The remaining 20 participants ranged in age from 21 to 46 (mean 27.1 years). All of the participants had been in New Zealand for less than two years, and were on a working holiday or studying English. Only one person was a university student. They had all received English language education for six years in junior high and high school in Japan, since English is a compulsory subject from age 12 in Japanese education. They spoke English as a foreign language and their total years living in foreign countries including non-English-speaking countries was less than three years. No participants reported any speech or hearing disorders, except one person. The hearing in her left ear was not as clear as in her right ear, but she reported that it did not affect her daily life and she did not need a hearing aid. In fact, her results did not differ from the others. Participation was voluntary. Participation was voluntary. Participants received a 20 dollar shopping voucher for participating in the experiments.

Thirteen participants spoke the Tokyo dialect (sometimes referred to as 'standard' Japanese) in everyday speech. Of these participants, eight spoke only the Tokyo dialect. The other participants reported that they shifted between the Tokyo dialect and their regional dialects depending on who they were talking to and what the situation was (for the list of dialects, see Appendix A).

3.2.5 Overall Procedure

The research design consisted of two experiments: perception and production, both of which were completed by all participants (see Chapter 4 for the production experiment). The perception experiment was an AX discrimination task, while the production experiment consisted of reading pseudo-words in a carrier phrase. Every participant completed two sessions of approximately 45 minutes each. The sessions were completed on different days in order to avoid participant fatigue and unreliable results. On each day, the participant completed both the AX discrimination task and the production task, in that order. Each task consisted of a practice session and an experimental trial section. The experimental section for each task was divided into two blocks: voiced consonant context and voiceless consonant context. Participants had the opportunity to take a break after each block. An example of the session schedule is presented in Appendix B. Experimental blocks were randomized across participants and days. All but one participant attended the first session on one day and the second session on another day. Only one participant completed all the tasks in one day. The intervals between the first session and second session ranged from 3 hours to one week, except one participant who had a 10 days interval between the two sessions due to unavoidable circumstances. The mean length of the interval between the two sessions was approximately 2.8 days (67 hours).

A background questionnaire was distributed to each participant after the conclusion of the second session, in order to avoid the questions influencing their performance. This questionnaire asked for basic demographic information and language experience (see Appendix C).

3.2.6 Procedure: AX Discrimination Task

The perception experiment was divided into two sections, with a practice and experimental task per session. Participants were given eight practice trials that included both voiced and voiceless consonant stimuli, which were real stimuli from the experiment task. Then there was a question-and-answer session to ensure that they understood the procedure before the experiment began. The experimental task consisted of two blocks per session. For the experimental task, four blocks of stimuli were created: two voiced and two voiceless lists, balancing conditions across lists. The experimental task was designed so that two lists were completed per session: one voiced and one voiceless, as shown in Table 3.4.

Table 3.4

Sample Session Schedule in the AX Discrimination Experiment

Session A	Session B
Instruction	Instruction
Practice	Practice
Question & Answer Session	Question & Answer Session
Block 1: Voiced Consonant Stimuli List 1	Block 1: Voiceless Consonant Stimuli List 2
Break	Break
Block 2: Voiceless Consonant Stimuli List 1	Block 2: Voiced Consonant Stimuli List 2

For example, on the first day, a participant listened to a list of voiced consonant stimuli in the first block and a list of voiceless stimuli in the second block. This is called *Session A* in Table 3.4. On the second day, the participant listened to the remaining two lists of stimuli in reverse order, as shown in *Session B*. While half of the participants did A on the first day and B on the second day, the other half completed B on the first day and A on the second day.

Participants were tested individually using E-prime software, in a sound-attenuated room at the University of Canterbury. However, three sessions were conducted in pairs. That is, two participants were tested individually but at the same time in the same room. All of them were in the first session. Each participant was situated in front of a computer screen wearing SENNHEISER HD280 Professional headphones. All participants listened to the stimuli at the same volume level. The participants were presented with instructions written in Japanese on the computer screen and these instructions were also briefly explained to subjects verbally in Japanese before the experiment began. Participants were divided into two groups. All participants were told that they would listen to pairs of sounds. The participants in the first group were asked to judge whether a speaker repeated the same word or said a different word; they were to press <1> on the keyboard for same, and <0> for different. The second group of participants were asked to press the keys in reverse order (i.e. <0> on the keyboard for same, and <1> for different). Figure 3.4 illustrates the AX discrimination protocol per trial. The inter-stimulus interval was 500ms. Participants were instructed to respond within 2000ms otherwise their responses were not detected by E-prime. After the participants pressed one of the choices, the next stimulus (= next trial) was presented.

Trial 1 = First Stimuli Pair (e.g., [abda - abada])				Trial 2 = Second Stimuli Pair (e.g., [agida - agida])			
Stimulus 1	Inter-stimulus interval	Stimulus 2	Response	Stimulus 1	Inter-stimulus interval	Stimulus 2	Response
Word 1	500ms	Word 2	2000ms	Word 1	500ms	Word 2	2000ms
abda		abada		agida		agida	

Figure 3.4. AX discrimination design diagram for experimental trials

Participants received their accuracy score (% correct) and response time between each trial during the experiment to encourage them to do the task as accurately and quickly as possible. All stimuli were randomised and presented to listeners in a different order. Each participant was tested in a total of 240 experimental trials per session: 120 voiced stimuli trials and 120 voiceless stimuli trials. The perception tasks took a total of approximately 30 minutes per session.

3.2.7 Hypothesis and Predictions

In the perceptual experiment, we assume, based on previous literature, that Japanese listeners should perceive a vowel between between the two consonants, since [CC] is an illicit sequence in Japanese. Given this assumption, the following hypothesis will be tested with regards to illusory epenthetic vowels: If perceptual epenthesis in Japanese is constrained by native phonotactics in the context [aC_1C_2a], we would expect Japanese listeners to perceive [o] after alveolar stops, [i] after palatal affricates and [ɯ] elsewhere. Two measures will be used to evaluate this hypothesis: accuracy and reaction time.

As for accuracy, we predict that there will be more errors between the pseudo-word with no vowel and the corresponding pseudo-words with the expected illusory vowel than there will be with pseudo-words with other vowels. This prediction is made under the assumption that when an expected vowel category consistent with phonotactics is presented, listeners are likely to perceive an illusory vowel in [aCCa]. The predicted judgements and relative reaction times are given in Table 3.5. In the voiced labial context, we predict that Japanese listeners will have a greater tendency to perceive the two stimuli [abɯCa - abCa] to be more similar than other different pairs. Similarly, Japanese listeners will have a greater tendency to judge the two stimuli [agɯCa - agCa] to be more similar in the voiced velar context. In the voiced alveolar and palatal contexts, respectively, the stimuli [adoCa - adCa] and [adʑiCa - adʑCa] will be judged to be more similar than other different pairs.

Table 3.5

Predictions of Judgements and Relative Reaction Times for Each Preceding Consonantal Context

Context	more errors/slower	fewer errors/faster	Context	more errors/slower	fewer errors/faster
Labial	abɯCa - abCa	abaCa - abCa	Alveolar	adoCa - adCa	adaCa - adCa
		abeCa - abCa			adeCa - adCa
		abiCa - abCa			adiCa - adCa
		aboCa - abCa			adɯCa - adCa
	apɯCa - apCa	apaCa - apCa		atoCa - atCa	ataCa - atCa
		apeCa - apCa			ateCa - atCa
		apiCa - apCa			atiCa - atCa
		apoCa - apCa			atɯCa - atCa
Velar	agɯCa - agCa	agaCa - agCa	Palatal	adʑiCa - adʑCa	adʑaCa - adʑCa
		ageCa - agCa			adʑeCa - adʑCa
		agiCa - agCa			adʑoCa - adʑCa
		agoCa - agCa			adʑɯCa - adʑCa
	akɯCa - akCa	akaCa - akCa		atɕiCa - atɕCa	atɕaCa - atɕCa
		akeCa - akCa			atɕeCa - atɕCa
		akiCa - akCa			atɕoCa - atɕCa
		akoCa - akCa			atɕɯCa - atɕCa

As for reaction time, we predict them to be slower between the pseudo-word with no vowel and the corresponding pseudo-words with the expected illusory vowel than they will be with pseudo-words with other vowels. For example, (a) [abɯda vs. abda] and (b) [adoba vs. adba] would yield slower reaction times than (c) [abada vs. abda] and [adaba vs. adba], respectively. This is because the members of the pairs in (a) and (b) are assumed to be more perceptually similar than those in (c), and it would therefore take listeners longer to make a decision, even if they do eventually come to the correct decision.

The predicted relative reaction times are also given in Table 3.5. In voiced labial and velar contexts, the pairs [abɯCa - abCa] and [agɯCa - agCa] would yield slower reaction times than when the V in the pairs [abVCa - abCa] and [agVCa - agCa] is one of the four vowels {a, e, i, o}. In the voiced alveolar context, [adoCa - adCa] would yield slower reaction times than stimuli pairs in which one of unexpected vowels {a, e, i, ɯ} is in [adVCa]. For the voiced palatal context, the pair [adʑiCa - adʑCa] would yield slower reaction times than stimuli pairs in which one of unexpected vowels {a, e, o, ɯ} is in [adʑVCa].

3.2.8 Analysis

In order to measure performance across participants, both accuracy and mean reaction time in the AX discrimination task were analysed. First, data from each subject, collected by E-prime, was merged into one file using E-Merge. The merged data was transferred to an Excel spreadsheet.

In the AX discrimination task, 13 responses were not detected/recorded since the participants did not respond within the given timeframe. Therefore, these 13 responses were removed, leaving 9587 tokens out of 9600 tokens. Then, the distribution of the data was checked by histogram using the statistical analysis tool R (R Core Team, 2014), and reaction times more than 2 standard deviations above the mean were removed as outliers from further analysis (Figure 3.5). This is because the middle 85-95% of the observations in reaction time distributions tends to be more reliable responses to test hypotheses (Ratcliff, 1993). It is possible that responses with long reaction times are real responses — not outliers —, however, long responses might also reflect loss of attention, distraction, and a simple memory lapse regarding which button is pushed for which answers. Therefore, in the current research, 4.4% of observations were removed, and 9164 observations remained for analysing the accuracy and reaction times of responses. Due to the high percentage of accuracy across environments and subjects as shown in the result section, the data did not fit a normal distribution for accuracy rates. For the results of accuracy, a logistic regression statistical analysis was conducted as will be discussed in detail further below.

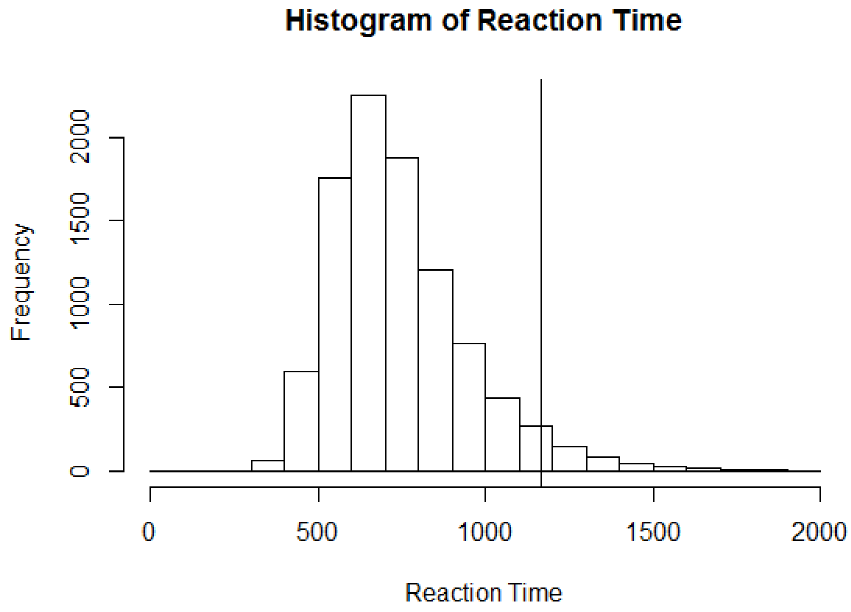

Figure 3.5. Reaction times with 2 standard deviations boundary for outlier removal.

Since a pilot AX discrimination study had shown high accuracy rates, it was expected that listeners in the current experiment would perform well in discriminating given pairs. This was in fact the case; only 632 pairs were judged inaccurately, for an overall percent correct of 93%. The current study will examine the details of these inaccurate judgments below in §3.3.1, but focus much of the rest of the discussion on the analysis of the reaction times of the correct pairs instead of the accuracy rates. In doing so, we would like to determine how quickly listeners are able to discriminate given pairs and to what extent reaction times differ between contrasting pairs in §3.3.2. The analysis of the reaction times of only the correct pairs follows standard practice in reaction time data analysis (e.g., Babel & Johnson, 2010; Davidson & Shaw, 2012; Pisoni & Tash, 1974).

3.3 Results

3.3.1 Discrimination Accuracy

Table 3.6 summarises the results of accuracy according to given environments. The results show that native speakers of Japanese performed very well in discriminating given pairs, regardless of phonological environment: preceding context 88-98%; vowel 80-97%. It should be noted that the *different* trials have lower accuracy than the *same* trials.

Table 3.6

Percent Correct Discrimination across Environments

C1	Accuracy		Vowel	Accuracy		Pair		Accuracy	
	mean	SD		mean	SD			mean	SD
[b]	93%	25%	[a]	96%	19%	Same	aCVCa_aCVCa	96%	19%
[d]	98%	14%	[e]	97%	17%		aCCa_aCCa	95%	21%
[g]	96%	21%	[i]	91%	28%	Different	aCVCa_aCCa	87%	33%
[dʑ]	95%	22%	[o]	97%	18%		aCCa_aCVCa	93%	25%
[p]	88%	32%	[ɯ]	80%	40%				
[t]	94%	25%	no vowel	95%	21%				
[k]	92%	27%							
[tɕ]	89%	31%							

The results of accuracy by subject in Table 3.7 show that all 20 subjects successfully discriminated between pairs with more than 90% accuracy, with the highest at 97%.

Table 3.7

Percent Correct Discrimination across Subjects

Subject	Accuracy		Subject	Accuracy	
	mean	SD		mean	SD
1	94%	23%	12	93%	25%
2	92%	27%	13	95%	22%
3	94%	23%	14	92%	28%
4	91%	29%	15	94%	24%
5	96%	21%	16	93%	26%
6	94%	24%	17	94%	25%
7	92%	28%	18	91%	29%
8	94%	24%	19	97%	17%
9	90%	30%	20	92%	27%
10	92%	28%	21	94%	23%

When we look at the accuracy of distinguishing in *different* trials as shown in Figure 3.6 (voiced context, two graphs) and Figure 3.7 (voiceless context, two graphs), we can observe a tendency of the participants to poorly discriminate those pairs which contrast an expected vowel with no vowel. It is these vowels which we might expect participants to perceive as illusory vowels in aCCa-stimuli. Recall that the expected vowels are [ɯ] after labial and velar consonants, [o] after alveolars, and [i] after palatals. The darker bars in the figures indicate the expected epenthetic vowel according to the preceding consonant. That is, the participants should find it difficult to discriminate between the pairs. The box plots enable us to observe the distributional response patterns of the group (the dark line marks the median (middle of dataset) of the dataset, i.e. 50% of the data is greater than this value).

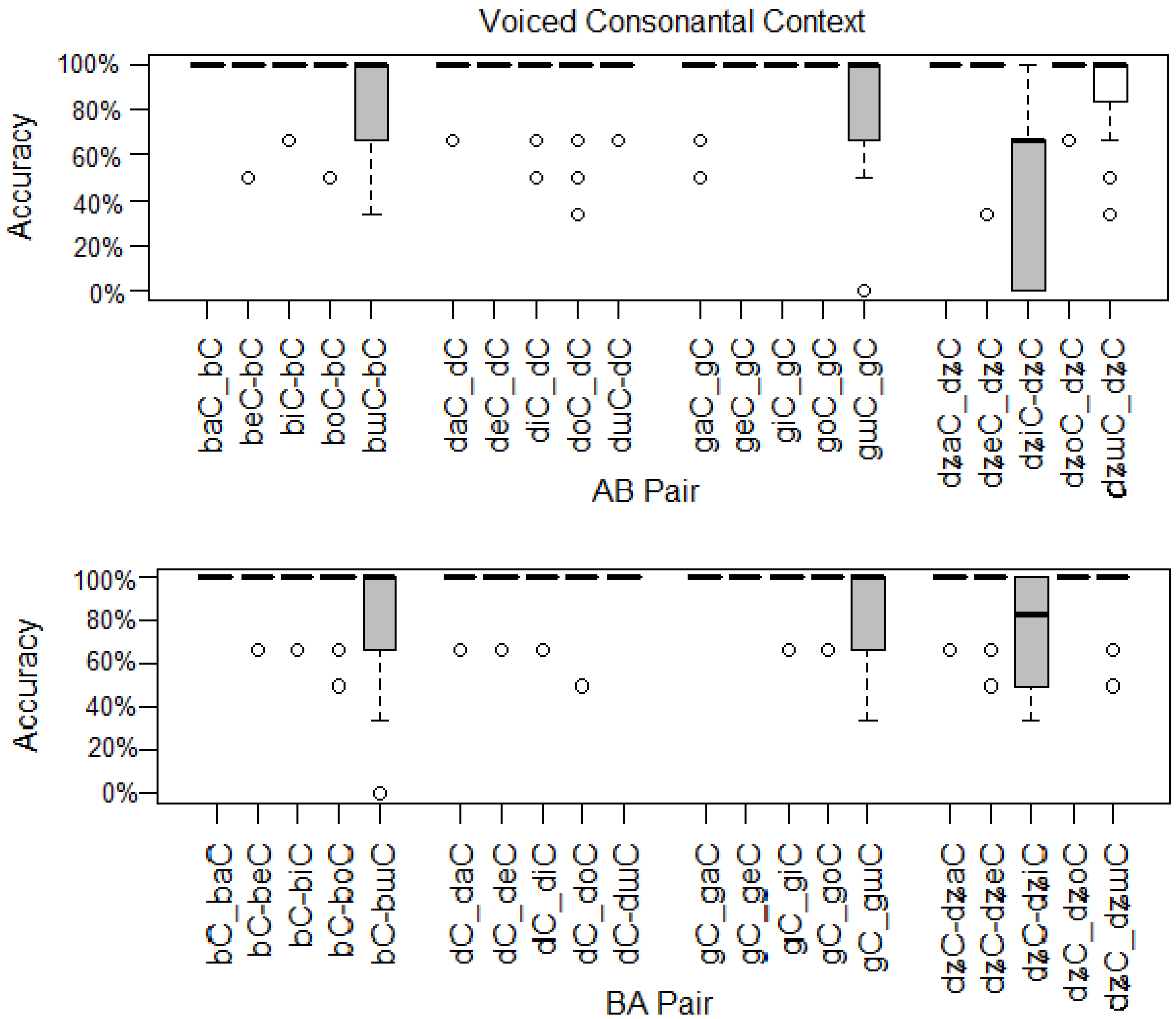

Figure 3.6. Box plots of percent discrimination accuracy for pairs with the voiced consonantal context in different trials. The darker bars in the figures indicate the expected epenthetic vowel according to the preceding consonant.

From Figure 3.6 above, in the [b] context, contrasting pairs were correctly discriminated at least 95% of the time, except when the medial vowels were [ɯ] in [aC_1VC_2a]. As predicted, when the contrast pair is [bɯC vs. bC], discrimination accuracy is lower than with any other vowel (~80%) for both ordered pairs <AB> and <BA>. Similarly, in the [g] context, each vowel was correctly discriminated at least 95% of the time, except for [ɯ], which was correct only 82% for <BA> and 89% for <AB>. In the alveolar contexts, listeners discriminated all contrasting pairs with at least 91% accuracy with [o] having the lowest score (91%). The listeners thus performed very well at discriminating alveolars regardless of vowel. In the preceding [dʑ] context, Japanese listeners showed difficulty in discriminating between [dʑiC - dʑC], being accurate 53% of the time for <AB> and 77 % of the time for <BA>. However, [dʑɯC - dʑC] cases are also a bit lower at 89% for <AB> which was unexpected since [i] is the expected illusory vowel in this context.

Considering all preceding contexts, when [a], [e], and [o] were the medial vowels in the [aC_1VC_2a] stimuli, listeners showed high accuracy in discriminating contrasting pairs. These results support the hypothesis that perceptual epenthesis in Japanese is constrained by native phonotactics, [i] after palatal affricates and [ɯ] elsewhere in [aC_1C_2a].

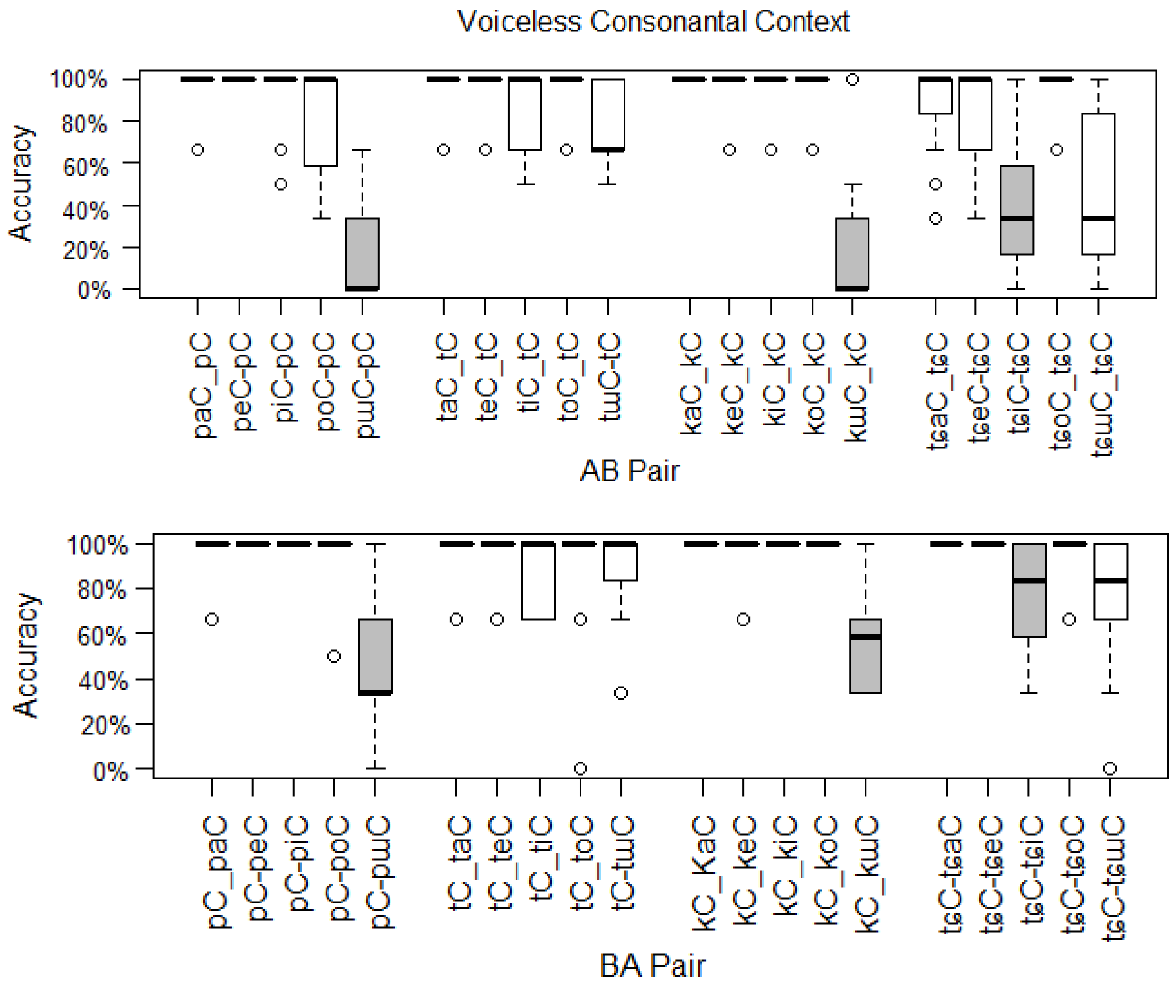

*Figure 3.7.*Box plots of discrimination accuracy for pairs with the voiceless consonantal context in different trials. The darker bars in the figures indicate the expected epenthetic vowel according to the preceding consonant.

Figure 3.7 reveals that, with the exception of [t], when preceding consonants were voiceless, Japanese listeners show greater difficulty in discriminating predicted pairs than when preceding consonants were voiced. Especially when the pairs had the order <AB> ([aCVCa] vs. [aCCa], as shown in the top graph of 3.7), listeners discriminated poorly between [pɯC] and [pC] (18% correct), [kɯC] and [kC] (20%), and [tɕiC] and [tɕC] (40%). In the [p] context, contrasting pairs were correctly discriminated at least 81% of the time, except when the medial vowels were [ɯ] in [aC_1VC_2a]. As predicted, when the corresponding pair is [pɯC - pC], discrimination accuracy is lower than with any other vowel (~43%) for both ordered pairs <AB> and <BA>. Similarly, in the [k] context, each vowel was correctly discriminated at least 95% of the time, except for [ɯ], which was correct only 20% for <AB> and 56% for <BA>. In the alveolar contexts, listeners discriminated all contrasting pairs at least 81% accuracy with [ɯ] having the lowest score (81%) while the expected vowel [o] was discriminated at least 94 % of the time. In the preceding [tɕ] context, Japanese listeners showed difficulty in discriminating between [tɕiC - tɕC] being accurate 40% of the time for <AB> and 76% of the time for <BA>, as expected. Contrasting pairs with [a], [e] and [o] were correctly discriminated at least 86% of the time. However, the listeners also unexpectedly showed difficulty in discriminating between [tɕɯC] and [tɕC] especially for <AB> at 47% and <BA> at 76% accuracy. Similar to the voiced context, considering all preceding contexts, when [a], [e], and [o] were the medial vowels in the [aC_1VC_2a] stimuli, listeners showed high accuracy in discriminating contrasting pairs.

Overall, these voiced and voiceless results are partially consistent with Shoji and Shoji (2014), where [ɯ] is considered the default epenthetic vowel and [i] is the context-dependent epenthetic vowel. The accuracy results showed that Japanese listeners are poor at discriminating between [aCVCa] with certain vowels and [aCCa]. Since Japanese does not permit the consonant sequences used in the current study, these results are interpreted as suggesting that the listeners are perceiving a vowel between the consonants, and that this illusory vowel is most confusable with [ɯ] in the labial and velar contexts, and with [i] in the palatal context. However, the current discrimination results do not support the claim that [o] is the context-dependent epenthetic vowel that is constrained by the preceding consonants [d] and [t]. On the other hand, to some extent, the findings of the present study support the findings of Monahan et al. (2009) which showed that Japanese listeners did not perceive an illusory epenthetic [ɯ] nor the contextually predicted vowel [o] after alveolar consonants. In the voiceless alveolar context in the current results, discrimination accuracy was slightly lower when the medial vowel was [ɯ] or [i] than when the medial vowel was [o] in the data. With regards to a preceding palatal consonant, a striking result to emerge from the data is that in the preceding [tɕ] context, the listeners also show difficulty in discriminating between [tɕɯC] and [tɕC] especially for <AB> with 47% accuracy. The results also reveal that the Japanese listeners show more difficulty discriminating the pairs predicted to be most confusable in voiceless consonantal contexts rather than in voiced contexts. However, it seems that devoicing contexts do not always influence speech perception to the same extent. While the devoiced high vowel [ɯ] seems to have an effect on alveolar and palatal contexts, the devoiced high vowel [i] in labial and velar contexts did not affect discrimination accuracy rates.

Interestingly, the order of pairs seems to have an influence on discriminating given pairs, at least in the voiceless context. The <AB> order consistently led to less accurate discrimination than the <BA> order. For [pɯC]-[pC], the <AB> order had an 18% accuracy rate, while the <BA> order had a 43% accuracy rate, and other pairs showed a similar pattern: 20% vs. 56% for [kɯC]-[kC], 40% vs. 76% for [tɕiC]-[tɕC], and 47% vs. 76% for [tɕɯC]-[tɕC]. Thus, when the first word was [aC$_1$VC$_2$a] and the second word was [aC$_1$C$_2$a], the accuracy rate was lower than when the stimuli were in the reverse order. This is consistent with the findings of Davidson (2011) where the order of presentation had an effect on perceptual epenthesis. This will be discussed this in the general discussion (Chapter 5).

Binominal logistic regression analyses were conducted to determine the extent to which certain variables predict the accuracy results. There are five predictors (explanatory variables): ‘voicing type’, ‘preceding consonant’, ‘vowel’, ‘stimulus order’ and ‘trial number’ with accuracy (1 = correct, 0 = wrong) as the response variable for each stimulus, as shown in Table 3.8. Trial number was rescaled in order to fit a regression model by converting the variable to a z-score (trial ranges from -0.87 to 0.86).

Table 3.8

Model Predictors: Independent Factors and Levels Coded for Analyses

Fixed Effect Factor (Predictor)	Levels
Voicing Type	Voiced/Voiceless
Place of Articulation	Labial/Alveolar/Palatal/Velar
Vowel	[a],[e],[i],[o],[ɯ], no vowel
Stimulus order	<AA>, <BB>,<AB>,<BA>
Trial Number	zTrial

Three-way interactions among voicing type, place of articulation and vowel, along with stimulus order and trial number were included as fixed effect factors, with subject as a random effect. Table 3.9 shows the output of a binominal logistic regression model of discrimination accuracy. This model includes reaction times for <AA>, <BB> pairs for *same*, and <AB> <BA> pairs for *different* trials. In the model shown here, the intercept is voiced alveolar [a] in <AA> order.

Table 3.9

Effect Estimates and P-values on Predictors for Accuracy

Coefficients	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	5.02378	0.59062	8.506	< 2e-16	***
Stimulus order <AB>	-1.60512	0.13883	-11.562	< 2e-16	***
Stimulus order <BA>	-0.77065	0.14802	-5.206	1.93E-07	***
Stimulus order <BB>	-1.25379	0.70082	-1.789	0.07361	.
zTrial	0.47316	0.09256	5.112	3.19E-07	***
Voicing.type Voiceless	0.44424	0.9031	0.492	0.62279	
Place of articulation Labial	-0.43964	0.73349	-0.599	0.54892	
Place of articulation Palatal	0.42739	0.91211	0.469	0.63937	
P Place of articulation Velar	-0.21317	0.76689	-0.278	0.78104	
Vowel [e]	0.51527	0.91276	0.565	0.5724	
Vowel [i]	0.06227	0.81917	0.076	0.9394	
Vowel [o]	-0.67855	0.71243	-0.952	0.34087	
Vowel [ɯ]	-0.04242	0.81869	-0.052	0.95867	
Voiceless*Labial	-2.1563	1.03473	-2.084	0.03717	*
Voiceless*Palatal	-1.70716	1.20749	-1.414	0.15742	
Voiceless*Velar	-0.50224	1.14965	-0.437	0.66221	
Voiceless*Vowel [e]	-1.66599	1.2183	-1.367	0.17148	
Voiceless*Vowel [i]	-2.61724	1.09852	-2.383	0.01719	*
Voiceless*Vowel no vowel	-1.20578	1.01657	-1.186	0.23557	
Voiceless*Vowel [o]	-0.46779	1.07809	-0.434	0.66436	
Voiceless*Vowel [ɯ]	-2.57537	1.09745	-2.347	0.01894	*

Labial*Vowel [e]	-0.78728	1.10101	-0.715	0.47458	
Palatal*Vowel [e]	-1.15811	1.25973	-0.919	0.35792	
Velar*Vowel [e]	-0.24954	1.19355	-0.209	0.83439	
Labial*Vowel [i]	0.17147	1.06406	0.161	0.87198	
Palatal*Vowel [i]	-3.38295	1.09964	-3.076	0.0021	**
Velar*Vowel [i]	-0.32346	1.06454	-0.304	0.76124	
Labial*Vowel no vowel	-1.29215	0.84707	-1.525	0.12715	
Palatal*Vowel no vowel	0.12863	1.10871	0.116	0.90764	
Velar*Vowel no vowel	-0.77797	0.89015	-0.874	0.38213	
Labial*Vowel [o]	0.86927	0.98444	0.883	0.37723	
Palatal*Vowel [o]	1.4479	1.41876	1.021	0.30747	
Velar*Vowel [o]	0.35029	0.98494	0.356	0.7221	
Labial*Vowel [ɯ]	-1.78354	0.96271	-1.853	0.06394	.
Palatal*Vowel [ɯ]	-1.72973	1.12534	-1.537	0.12427	
Velar*Vowel [ɯ]	-1.60302	0.99667	-1.608	0.10775	
Voiceless*Labial*Vowel [e]	18.62652	52.36149	0.356	0.72204	
Voiceless*Palatal*Vowel [e]	1.61945	1.57024	1.031	0.30238	
Voiceless*Velar*Vowel [e]	0.85473	1.573	0.543	0.58687	
Voiceless*Labial*Vowel [i]	4.58283	1.43391	3.196	0.00139	**
Voiceless*Palatal*Vowel [i]	3.54689	1.38629	2.559	0.01051	*
Voiceless*Velar*Vowel [i]	4.26241	1.70784	2.496	0.01257	*
Voiceless*Labial*Vowel no vowel	3.56149	1.1742	3.033	0.00242	**
Voiceless*Palatal*Vowel no vowel	2.48108	1.47633	1.681	0.09285	.
Voiceless*Velar*Vowel no vowel	2.581	1.34204	1.923	0.05446	.
Voiceless*Labial*Vowel [o]	0.91721	1.32588	0.692	0.48908	
Voiceless*Palatal*Vowel [o]	-0.58079	1.71123	-0.339	0.73431	
Voiceless*Velar*Vowel [o]	0.85897	1.4606	0.588	0.55647	
Voiceless*Labial*Vowel [ɯ]	2.5132	1.24026	2.026	0.04273	*
Voiceless*Palatal*Vowel [ɯ]	2.07226	1.40611	1.474	0.14055	
Voiceless*Velar*Vowel [ɯ]	0.63855	1.34396	0.475	0.6347	

(Significance codes: ***$p < .001$, ** $p < .01$, * $p < .05$)

Trial number appears to have a positive significant effect on the response ($p < .001$). This indicates that accuracy increases over the course of the experiment. There was also a significant effect of stimulus order for <AB> and <BA> ($p < .001$). The effect of voicing type, preceding consonant or vowel was not significant when each was tested. However, the interaction between the variables ‘voicing type’, ‘preceding consonant’ and ‘vowel’ proved significant for some combinations. That is, the model predicts that voiceless labial [ɯ] ($p < .05$) and voiceless palatal [i] ($p < .05$) are more difficult to discriminate for listeners than the individual factors predict.

3.3.2 Discrimination Reaction Time

In this section, I present the results of the reaction time and statistical analyses first for the full set of data, and then according to the quality of the preceding consonant. I examine how *different* pairs (e.g., [abaga vs.abga], [abɯga vs. abga]) influence reaction times for discriminating pairs of stimuli with reaction time being the dependent variable. Independent variables were ‘voicing type’, ‘place of articulation’, ‘vowel’, ‘stimulus order’ and ‘trial number’, as shown in Table 3.10.

Table 3.10

Model Predictors: Independent Factors and Levels Coded for Analyses

Fixed Effect Factor (Predictor)	Levels
Voicing Type	Voiced/Voiceless
Place of Articulation	Labial/Alveolar/Palatal/Velar
Vowel	[a],[e],[i],[o],[ɯ]
Stimulus order	<AB>/<BA>
Trial Number	zTrial

A linear mixed-effect regression analysis in R was conducted to analyse the effect of the independent factors on reaction time. Three-way interactions among voicing type, place of articulation and vowel, along with stimulus order and trial number were included as fixed effect factors, with a random slope of trial number by subject. Trial number was rescaled in order to fit a linear regression model by converting the variable to a z-score (trial ranges from -0.87 to 0.86). Adjustment to the random slope was under the assumption that all subjects get better after each trial, however, some subjects may improve more quickly than others. Within the random slope model, subjects are allowed to have individually varying intercepts and slopes. Since the mixed-effect regression model does not show p-values, the package ‘lmerTest’ (Kuznetsova, Brockhoff, & Christensen, 2015) in R was used to present the results with p-values.

Table 3.11 shows the output of a mixed-effect regression model of reaction time. This model includes reaction times for both <AB> and <BA> pairs for *different* trials. In the model shown here, the intercept is voiced alveolar [a] in <AB> order, which has an estimated reaction time of 725.31 ms. Non-significant interaction effects have been removed.

Table 3.11

Effect Estimates and P-values on Predictors for Reaction Time

Coefficient	Estimate	Std.Error	t value	Pr(>|t|)	
(Intercept)	725.305	23.521	30.836	< 2e-16	
Voicing.type Voiceless	8.1	16.898	0.479	0.63172	
Place of articulation Labial	18.307	16.603	1.103	0.27027	
Place of articulation Palatal	25.009	16.819	1.487	0.13712	
Place of articulation Velar	-4.417	16.749	-0.264	0.79202	
Vowel [e]	-9.084	16.603	-0.547	0.58431	
Vowel [i]	36.993	16.708	2.214	0.02688	*
Vowel [o]	2.604	17.025	0.153	0.87843	
Vowel [ɯ]	29.469	17.018	1.732	0.08342	.
Stimulus order <BA>	-43.34	3.952	-10.967	< 2e-16	***
Trial numbers	-65.229	15.362	-4.246	0.00044	***
Labial* Vowel [ɯ]	55.298	24.752	2.234	0.02553	*
Voiceless* Labial* Vowel [e]	-74.056	33.521	-2.209	0.02721	*

(Significance codes: ****p* < .001, ** *p* < .01, * *p* < .05)

There were no significant effects of place of articulation or voicing type by themselves. The model does not show an effect of vowel except for [i] (p <0.05), which is predicted to have a significantly longer RT than [a] in this context. There was a significant effect of trial number, with a negative estimate, indicating significantly shorter reaction times as trial number increases. The trial z-score ranges from -0.87 to 0.86, therefore this predicts 56.749 ms (i.e. –65.229 *-0.87) longer on the first trial and 56.749 ms shorter on the last trial. There was also a significant difference between predicted reaction times of <AB> and <BA> pairs, with <BA> pairs predicted as being responded to more quickly. In terms of interactions, the interaction of labial + [ɯ] indicates that when the preceding consonant is labial and the vowel is [ɯ], this combination makes discrimination even more difficult for listeners than the individual factors predict (*t* = 2.234, *p* < .05). Also, there is a significant effect of the interaction between voiceless labial + [e] (*t* = -2.209, *p* < .05). This interaction indicates that when these factors occur together, individual predicted effects are mitigated.

In order to take a closer look at the effect of each preceding context, separate models for each of the preceding contexts were examined. Each model includes reaction times for both <AB> and <BA> pairs for *different* trials. Mixed effects regression models allow us to compare any given level of a given factor against the intercept for that factor. By releveling the factor and setting different values as the intercept, we can test for significant differences between any two levels, or values. The reason that releveling is necessary is because when a given value (e.g., [i]) is set as the intercept, we are able to determine whether each of the other values is significantly different from [i]. However, when [i] is the intercept, we are not able to directly compare, for example, [e] to [a]. For that reason, we set each value as the intercept in turn, which then allows us to directly compare each value to each other value.

Labial Context

We begin by looking at reaction time differences in the labial context. In this context, the expected epenthetic vowel is the high vowel [ɯ]. Table 3.12 shows the output of a mixed-effect regression model of reaction time as a function of stimulus order, trial number and a two-way interaction between voicing type and vowel, with a random slope of trial number by subject. This model includes reaction times for both <AB> and <BA> pairs for *different* trials. The intercept is [a], in the voiced context, in AB pairs, which has a predicted reaction time of 734.2 ms. The results did not show a significant effect of vowel except for [ɯ], for which reaction time is predicted to be significantly longer ($t = 4.791$, $p < .001$). There was also a significant effect of stimulus order ($t = -3.394$, $p < .001$). This implies that for <BA> pairs, [abCa - abVCa], listeners are predicted to take significantly less time to make a decision than for <AB> pairs, [abVCa - abCa]. As for the effect of trial number, the results show significantly shorter reaction times as trial number increases. As with the full model, this model predicts that reaction time will decrease as trail number increases. In terms of interactions, the interaction of voiceless [p] + [e] indicates that when the preceding labial is voiceless, the vowel [e] is predicted to be discriminated significantly faster than the individual factors predict ($t = -2.331$, $p < .05$).

Table 3.12

Results of the Mixed-Effect Regression for Reaction Time in the Labial Context

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	734.219	23.172	31.686	< 2e-16	
Voicing.type Voiceless [p]	-10.027	16.248	-0.617	0.53731	
Vowel [e]	27.014	16.239	1.663	0.09654	.
Vowel [i]	6.773	16.085	0.421	0.67381	
Vowel [o]	-14.098	16.196	-0.87	0.38427	
Vowel [ɯ]	84.225	17.58	4.791	1.9E-06	***
Stimulus Order <BA>	-26.4	7.778	-3.394	0.00072	***
Trial Number	-79.438	16.588	-4.789	0.00013	***
Voiceless [p]*Vowel [e]	-53.81	23.086	-2.331	0.01997	*
Voiceless [p]*Vowel [i]	-44.218	23.017	-1.921	0.05502	.
Voiceless [p]*Vowel [o]	25.891	23.428	1.105	0.26938	
Voiceless [p]*Vowel [ɯ]	59.027	30.216	1.954	0.05105	.

(Significance codes: ***$p < .001$, ** $p < .01$, * $p < .05$)

Figure 3.8 shows estimated reaction times in the labial context based on the interaction of voicing and vowel, from the model in Table 3.12. In order to take a closer look at the effect of vowel in this context, this section walks through changing the intercept to test the significance of various differences between vowels.

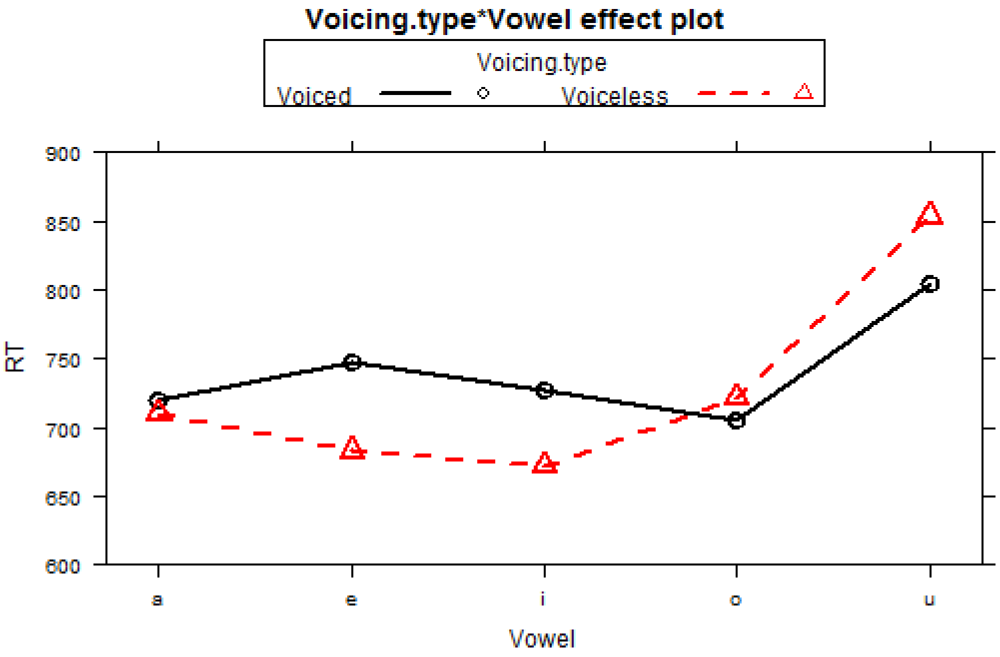

Figure 3.8. Model prediction for reaction times in the labial context based on the interaction of voicing type and vowel (from model in Table 3.12).

When [ɯ] was preceded by a voiced labial ([abɯCa-abCa]), pairs had an estimated reaction time of 818.44 ms, whereas when [ɯ] was preceded by a voiceless labial ([apɯCa-apCa]), the estimated reaction time was longer (867.445 ms). This difference approaches significance (*t* = -1.921, *p* = .055). For the predictor vowel, there was a significant effect on reaction time between the high vowel [ɯ] and each of the other vowels in both the voiced context ([a]: *t* = -4.791, *p* < .001); [e]: *t* = -3.212, *p* < .01; [i]: *t* =- 4.383, *p* < .001; [o]: *t* = -5.533, *p* < .001) and the voiceless context ([a]: *t* =- 5.807, *p* < .001; [e]: *t* = -6.909, *p* < .001; [i]: *t* = -7.332, *p* < .001; [o]: *t* = -5.262, *p* < .001). These results show that in the preceding labial context, listeners take a significantly longer time to discriminate pairs with the medial vowel [ɯ], which was predicted by the hypothesis.

Velar Context

In the velar contexts, the expected epenthetic is also the high vowel [ɯ]. Table 3.13 shows the output of a mixed-effect regression model of reaction time as a function of stimulus order, trial number, and two-way interactions between voicing type and vowel, with a random slope of trial number by subject. In the table, the intercept is [a], in the voiced context, in an <AB> pair, which has an estimated reaction time of 719.52 ms. There was no significant effect of voicing type of the preceding consonant on reaction time (*t* = .64, *p* = .522). The vowel factor was only significant for [ɯ] (*t* = 3.213, *p* < .01): the [agɯCa - agCa] pair is an average 59.14 ms slower than [agaCa - agCa]. There was a significant effect of stimulus order (*t* = -4.794, *p* < .001). This means that while [agɯCa - agCa] should take listeners an average of 778.66 ms to make a decision, [agCa - agɯCa] will be faster at 739.24 ms. As for the effect of trial number, the results show significantly shorter reaction times as trial number increases: this model predicts that reaction time will decrease as trail number increases. In terms of interactions, the combination of [k] (voiceless) and [ɯ] in the <AB> pair (*t* = 2.414, *p* < .05) is significantly slower than the individual factors predict, which is predicted to be 862.7 ms to judge for listeners.

Table 3.13

Results of the Mixed-Effect Regression for Reaction Time in the Velar Context

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	719.516	25.492	28.225	< 2e-16	
Voicing.type Voiceless [k]	11.179	17.462	0.64	0.52222	
Vowel [e]	11.132	17.301	0.643	0.52009	
Vowel [i]	10.989	17.544	0.626	0.53123	
Vowel [o]	3.283	17.538	0.187	0.85153	
Vowel [ɯ]	59.144	18.407	3.213	0.00135	**
Stimulus Order <BA>	-39.425	8.224	-4.794	1.89E-06	***
Trial Number	-48.124	16.15	-2.98	0.00791	**
Voiceless [k]*Vowel [e]	-31.387	24.662	-1.273	0.20342	
Voiceless [k]*Vowel [i]	6.033	24.847	0.243	0.80821	
Voiceless [k]*Vowel [o]	-25.75	24.724	-1.042	0.2979	
Voiceless [k]*Vowel [ɯ]	72.952	30.225	2.414	0.01598	*

(Significance codes: ****p* < .001, ** *p* < .01, * *p* < .05)

Figure 3.9 shows estimated reaction times in the velar context based on the interaction of voicing and vowel, from the model in Table 3.13.

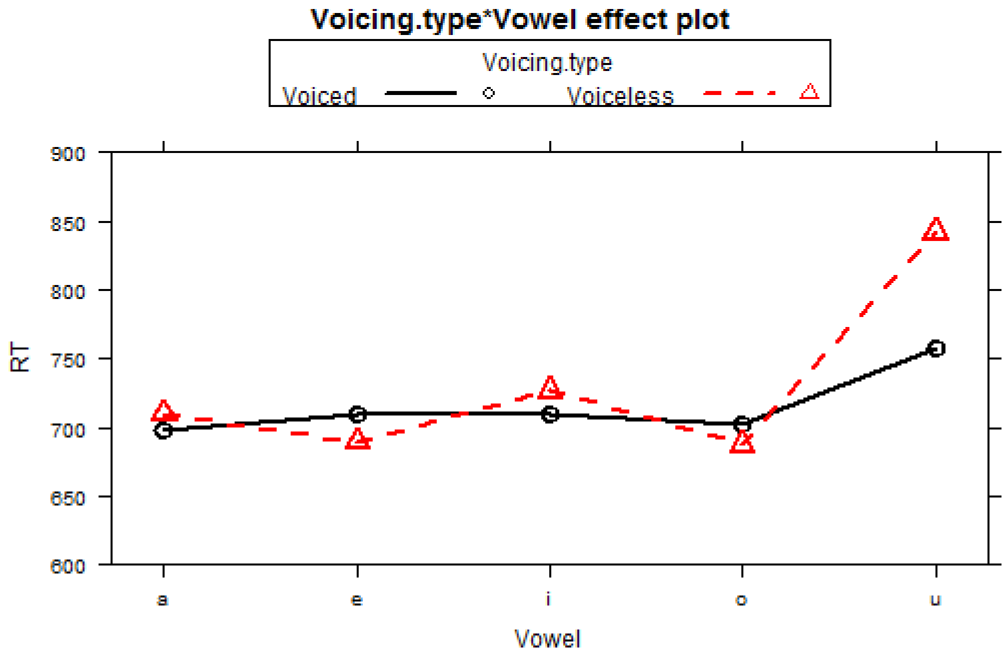

Figure 3.9. Model prediction for reaction times in the velar context based on the interaction of voicing type and vowel (from model in Table 3.13).

When [ɯ] was preceded by a voiced velar, the [agɯCa - agCa] pair had an estimated reaction time of 778.66 ms, whereas when preceded by a voiceless velar, the [akɯCa - akCa] pair had a slower estimated reaction time of 862.79 ms. In the velar context, this effect of preceding voicing on reaction time was significant (*t* = 3.41, *p* < .001). There were also significantly different reaction times between the high vowel [ɯ] and each of the other vowels in the voiced context ([a]: *t* = -3.213, *p* < .01: [e]: *t* = -2.625, *p* < .01; [i]: *t* = -2.598, *p* < .01; [o]: *t* = -3.017, *p* < .01) as well as in the voiceless context ([a]: *t* = -5.520, *p* < .001; [e]: *t* = -6.349, *p* < .001; [i]: *t* = -4.793, *p* < .001; [o]: *t* = -6.470, *p* < .001). These results show that listeners are predicted to take a significantly longer time to discriminate pairs in the velar context with the medial vowel [ɯ], which was predicted by the hypothesis.

Alveolar Context

Next, I present the statistical results for the alveolar context, in which the predicted epenthetic vowel is the mid vowel [o]. Table 3.14 shows the output of a mixed-effect regression model of reaction time as a function of stimulus order, trial number and a two-way interaction between voicing type and vowel, with a random slope of trial number by subject. In the model shown in 3.13, the intercept is [a], in the voiced context, in the AB pair, which has an estimated reaction time of 732.8 ms. In the alveolar context, there was no significant difference in reaction time between voiced and voiceless preceding consonants (*t* = .496, *p* = .61). There was a significant effect of vowel for [i], for which reaction time is predicted to be significantly longer (*t* = 2.209, *p* < .05). Similarly to the two previous contexts mentioned above, there was a significant effect of stimulus order (*t* = -7.591, *p* < .001). This indicates that while [adaCa - adCa] should take listeners 732.81 ms to make a decision, [adCa - adaCa] will be faster at 675.77 ms. There was also an effect of trial number; this model predicts that reaction time will decrease as trail number increases. In terms of interactions, there was a significant effect of the combination of voiceless [t] + [ɯ], for which reaction time is significantly longer (*t* = 2.079, *p* < .05).

Table 3.14

Results of the Mixed-Effect Regression for Reaction Time in the Alveolar Context

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	732.812	23.666	30.964	< 2e-16	
Voicing.type Voiceless [t]	8.21	16.539	0.496	0.61971	
Vowel [e]	-9.828	16.244	-0.605	0.54532	
Vowel [i]	36.105	16.346	2.209	0.02741	*
Vowel [o]	2.18	16.681	0.131	0.89604	
Vowel [ɯ]	28.455	16.661	1.708	0.08797	.
Stimulus Order <BA>	-57.045	7.515	-7.591	7.1E-14	***
Trial Number	-75.567	16.731	-4.516	0.00023	***
Voiceless [t]*Vowel [e]	21.654	23.23	0.932	0.35146	
Voiceless [t]*Vowel [i]	5.208	23.58	0.221	0.82526	
Voiceless [t]*Vowel [o]	6.602	23.705	0.279	0.78067	
Voiceless [t]*Vowel [ɯ]	49.822	23.967	2.079	0.03788	*

(Significance codes: ****p* < .001, ** *p* < .01, * *p* < .05)

Figure 3.10 shows estimated reaction times in the alveolar context based on the interaction of voicing and vowel, from the model in Table 3.14. In the model, when [o] was preceded by a voiced alveolar, the [adoCa - adCa] pair had an estimated reaction time of 734.99 ms, whereas when [o] was preceded by a voiceless alveolar, the [atoCa_atCa] pair had a slightly slower average reaction time of 749.81 ms. However, this difference was not significant (*t* = .873, *p* = .382).

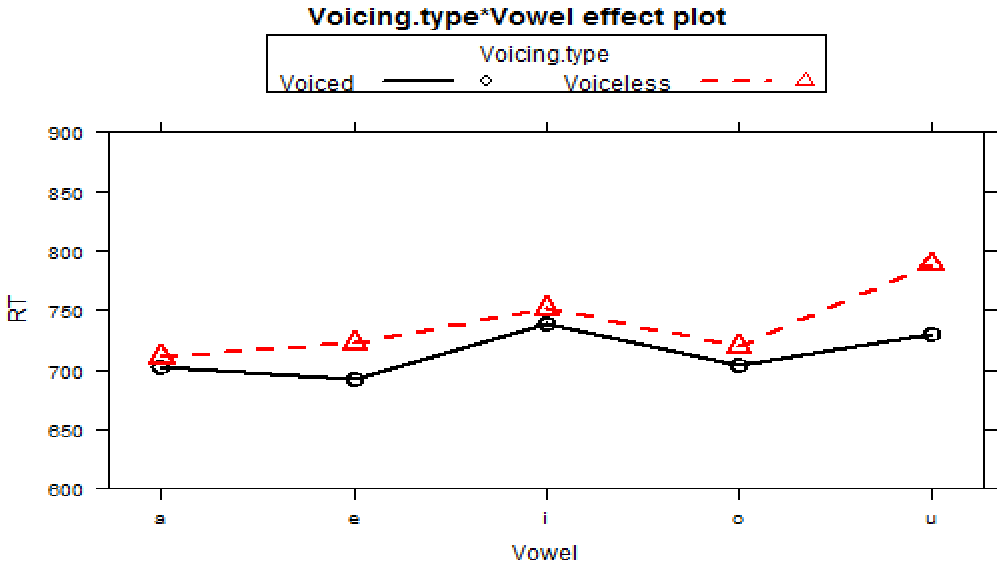

Figure 3.10. Model prediction for reaction times in the alveolar context based on the interaction of voicing type and vowel (from model in Table 3.14).

In the voiced context, there was a small but significant effect of vowel on reaction time between the mid back vowel [o] and [i] ($t = 2.049$, $p < .05$). That is, estimated reaction times for [adiCa-adCa] pairs were 33.93 ms slower than [adoCa-adCa] pairs (contra the prediction that [o] pairs would be slower than other vowel pairs). However, there was no significant difference between [o] and [ɯ], or between any other vowels. For the voiceless context, there was a significant difference between [o] and [ɯ] ($t = 3.991$, $p < .001$). That is, the [atɯCa-atCa] pairs yielded significantly slower reaction times than the [atoCa-atCa] pairs, contrary to what was expected. There was also a significant difference, in the voiceless context, between [ɯ] and all of the other vowels ([a]: $t = -4.547$, $p < .001$; [e]: $t = -3.861$, $p < .001$; [i]: $t = -2.013$, $p < .05$), but there was no significant difference between [atiCa-atCa] and [atoCa-atCa]. These results suggest that Japanese-speaking listeners did not show significantly slower discrimination on the stimuli in which [o] should be perceived as the epenthetic vowel in both voiced and voiceless contexts. This result was not expected by the hypothesis. Contrary to the predictions, listeners took a significantly longer time to discriminate pairs with the medial vowel [i] in the voiced alveolar context and with the medial vowel [ɯ] in the voiceless alveolar context. Japanese listeners were faster at discriminating alveolar followed by [o].

Palatal Context

Finally, the results for the palatal context will be reported; the predicted epenthetic vowel is the high vowel [i]. In the model shown in 3.13, the intercept is [a], in the voiced context, in the AB pair, which has an estimated reaction time of 753.44 ms. As is the case with other contexts, there was no significant difference in reaction time between voiced and voiceless preceding consonants ($t = -.553$, $p =.58$). The vowel factor was only significant for [i] ($t = 4.162$, $p < .001$): the [adʑiCa - adʑCa] pair is estimated to be 80.539 ms slower than [adʑaCa - adʑCa]. These results suggest that Japanese-speaking listeners responded significantly slower to the pair in which [i] was expected to be perceived as the epenthetic vowel. This finding is consistent with our hypothesis. Similar to other contexts, there was a significant effect of stimulus order ($t = -6.1$, $p < .001$). This indicates that [adʑiCa - adʑCa] takes listeners an average 833.98 ms to make a decision, whereas [adʑCa - adʑiCa] takes 784.62 ms. There was also an effect of trial number; the results show significantly shorter reaction times as trial number increases. As with the full model, this model predicts that reaction time will decrease as trail number increases. In terms of the interaction, there was no significant effect of the combination of voiceless consonant + vowel.

Table 3.15

Results of the Mixed-Effect Regression for Reaction Time in the Palatal Context

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	753.44	22.578	33.371	< 2e-16	
Voicing.type Voiceless [tɕ]	-9.29	16.796	-0.553	0.58033	
Vowel [e]	31.024	16.634	1.865	0.06248	.
Vowel [i]	80.539	19.351	4.162	3.45E-05	***
Vowel [o]	-15.182	16.475	-0.922	0.35701	
Vowel [ɯ]	22.177	16.907	1.312	0.18996	
Stimulus Order <BA>	-49.36	8.091	-6.1	1.56E-09	***
Trial Number	-56.093	17.439	-3.217	0.00462	**
Voiceless [tɕ]*Vowel [e]	-16.027	23.804	-0.673	0.50093	
Voiceless [tɕ]*Vowel [i]	-13.308	27.813	-0.478	0.63242	
Voiceless [tɕ]*Vowel [o]	-11.397	23.555	-0.484	0.62861	
Voiceless [tɕ]*Vowel [ɯ]	34.351	25.766	1.333	0.1828	

(Significance codes: ****p*< .001, ** *p*< .01, * *p*< .05)

Figure 3.11 shows estimated reaction times in the palatal context based on the interaction of voicing and vowel, from the model in Table 3.15.

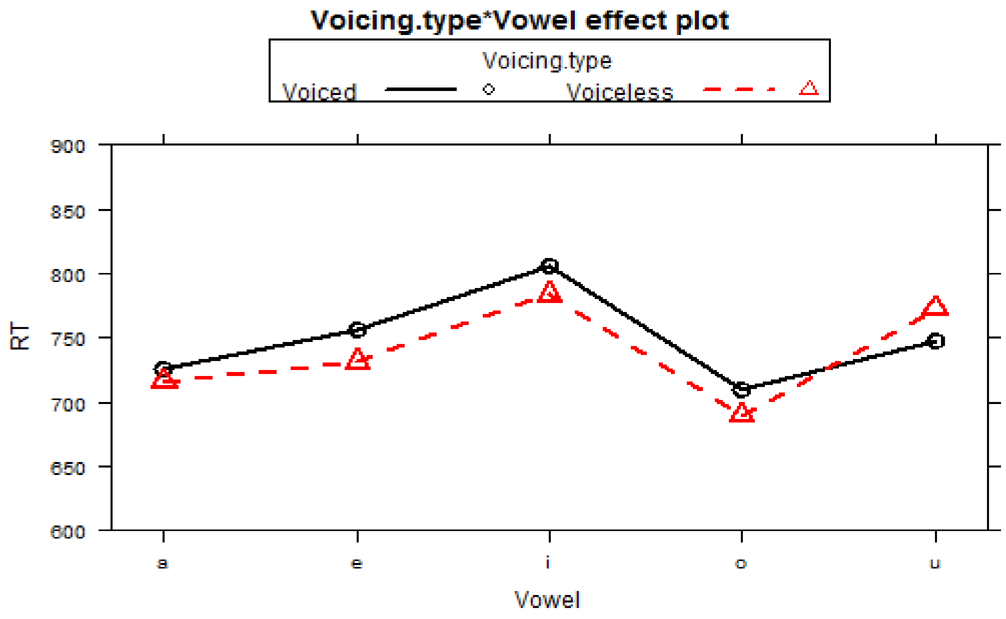

Figure 3.11. Model prediction for reaction times in the palatal context based on the interaction of voicing type and vowel (from model in Table 3.15).

There was no significant effect on reaction time between voiced and voiceless preceding consonants for [i]. In the voiced context, there was a significant difference in reaction time between the high vowel [i] and each other vowel: [a] ($t = -4.162$, $p < .001$), [e] ($t = -2.554$, $p < .01$), [o] ($t = -4.974$, $p < .001$), and [ɯ] ($t = -2.977$, $p < .01$). These results suggest that in the voiced context Japanese-speaking listeners responded significantly slower to the pair in which [i] was the expected epenthetic vowel, consistent with our hypothesis. For the voiceless context, there was also a significant effect of vowel on reaction time between [i] and other vowels, except [ɯ]: [a] ($t = -3.34$, $p < .001$), [e] ($t = -2.589$. $p < .01$), [o] ($t = -4.68$, $p < .001$), [ɯ] ($t = -.483$, $p = .629$). The analysis suggests that the effect of vowel on voiceless palatals with the medial vowel [i] is similar to [ɯ] in that they are more likely to take longer to discriminate than other vowels.

3.4 Discussion

This study investigated the extent to which perceptual epenthesis in Japanese is influenced by the preceding consonantal context and constrained by native phonotactic patterns. As previously mentioned, in the previous literature (e.g., Irvin, 2011; Shoji & Shoji, 2014) it is argued that the preceding consonantal context impacts the selection of epenthetic vowels. That is, a language can have more than one epenthetic vowel, depending on the preceding consonants given the phonotactic patterns of Japanese. Based on these studies, the high back vowel [ɯ] was predicted to occur in the most different contexts. The palato-alveolar affricates [dʑ] and [tɕ] attract the high front vowel [i]. In addition, since [ɯ] and [i] do not occur after the alveolar stops [d] and [t] in Japanese, the mid back vowel [o] was predicted after the alveolar stops.

The findings of the present study are partially consistent with previous findings, but also show important differences. Specifically, while there do seem to be language-specific perceptual effects consistent with Japanese phonotactics, these were found for only the labial, palatal, and velar contexts, but not for the alveolar context. When the preceding consonants were labial or velar, the listeners showed poorer discrimination between [aC_1ɯC_2a] and [aC_1C_2a], compared to other vowels. This effect was enhanced when the preceding consonants were voiceless. In terms of reaction time, even when the listeners discriminated contrasting pairs correctly, the results show, regardless of voicing types, significantly slower reaction times with the predicted epenthetic context for [ɯ], in comparison to other vowels.

Consistent with the previous literature, [i] was predominantly perceived as the epenthetic vowel after the palato-alveolar affricates [dʑ] and [tɕ]. Reaction times for [i] in the palatal contexts are significantly slower than for other vowels, except for [ɯ] in the voiceless context. On the other hand, contrary to our predictions based on native Japanese phonotactics, we found that the mid back vowel [o] was not perceived after the alveolar stops [d] and [t], as evidenced by the fact that it did not yield increased reaction times. However, while [ɯ] was the illusory vowel perceived in voiceless alveolar contexts, this was not the case in the voiced alveolar context. This finding differs in some ways from the findings of Mattingley et al. (2015), which used an identification task rather than discrimination, and was limited to the voiced context. Similar to the current study, the illusory vowel was identified as [ɯ] (43%) but also identified as ‘no vowel’ (37%). The vowel [o] was perceived after the voiced alveolar stop [d] only 10% of the time. To some extent, the results are also consistent with the findings of Monahan et al. (2009) who found that Japanese listeners did not illusorily epenthesize [o] after alveolar stops.

In addition to the hypothesis regarding Japanese phonotactics, a goal of this study was to determine the extent to which Japanese listeners would show difficulty discriminating contrasting pairs, irrespective of the preceding environment, when [ɯ] was in the stimuli [aC$_1$VC$_2$a]. Consistent with Dupoux et al. (2011), we did find that to some extent Japanese listeners have a bias toward perceiving [ɯ]. As mentioned before, in the labial and velar contexts, the listeners showed poorer discrimination between [aC$_1$ɯC$_2$a] and [aC$_1$C$_2$a] than with other vowels. One of the most exciting findings in the current research is that [ɯ] was perceived even in some contexts where it was not expected based on the preceding consonant. That is, after the palatal [tɕ], the listeners showed poor discrimination between [tɕɯ] and [tɕC], indicating that the illusory vowel they perceive in [aC$_1$C$_2$a] tokens is similar to [ɯ]. This contrasting pair showed similar reaction times to the [tɕi - tɕC] pair, both of which were significantly slower than for other vowel contexts. For the alveolar context, in voiceless stimuli, when the vowel [ɯ] was the medial vowel, perception accuracy was lower than when the medial vowel was [o], even though its accuracy was high. Reaction times were also significantly slower when the vowel [ɯ] occurred in stimulus pairs [atVCa] compared to [o]. This finding suggests that the domain of the default vowel [ɯ] is generalising to beyond what would be predicted by phonotactics.

A remaining question is why Japanese listeners have slower reaction times for [ɯ] in contexts where the vowel [o] might be expected from a native Japanese perspective. The results suggest that devoicing influences discrimination of contrasting pairs, because in the corresponding voiced context, the reaction times for [ɯ] were not significantly different from those other vowels. The perception study of spoken Japanese words by Cutler, Otake and McQueen (2009) claims that the vowel devoicing context makes speech segmentation and word recognition more difficult than when devoicing is not allowed. For other potential explanations of why [ɯ] is perceived and thus led to longer reaction times, this study considers the phonetic characteristics of the Japanese vowel [ɯ]. One potential explanation relates to the weak phonetic nature of [ɯ]. For example, Sagisaka and Tokuhara (1984, as cited in Irwin, 2011, p. 106) have pointed to the high back vowel as “the Japanese vowel most subject to weakening and deletion, as well as being the shortest phonetically”. Yoshida (2006) also states that the high vowel [ɯ] is the shortest vowel among the five vowels and it is less likely to be accented, at least in Tokyo Japanese. In fact, the duration of [ɯ] in our dataset was the shortest in length among the five vowels in both voiced and voiceless contexts (see § 3.2.2). In addition, the vowel [ɯ] is the least sonorant vowel. In terms of the quality of the epenthetic vowel, Dupoux et al. (2011) argue that the phonetically minimal vowel of the language would be a candidate for being the epenthetic vowel. It may also be the case that the language’s statistical patterns such as predicting certain phonological processes and changes support a strong perceptual bias toward [ɯ]. As such, in contexts with weak perceptual cues, subjects are biased to perceive the most expected vowel in the language (Hume & Bromberg, 2005). When taken together, the quality of epenthetic vowels seems to be not only influenced by first language phonology but also the phonetic properties of vowels.

The reason why I conducted an AX discrimination task in the current study was to observe whether the inconsistency in the findings of Mattingley et al. (2015) and Monahan et al. (2009) stemmed from different tasks (Mattingley et al. 2015 using an identification task, while Monahan et al. 2009 used an AX discrimination task). Recall that the mid back vowel [o] was not perceived much after the voiced alveolar stop [d] in either study, contra the expectation based on loanword studies. However, listeners in Mattingley et al. (2015) were strongly biased to perceive [ɯ] after [d] even though *[dɯ] is an illicit phonotactic sequence in native Japanese. This result differs from Monahan et al. (2009) who investigated the relationship between perceptual epenthesis and native language phonology using an AX discrimination task. In the present AX discrimination study, as is the case with Monahan et al.

(2009), Japanese listeners did not perceive the contextually predicted vowel [o] after alveolars, nor did they perceive an illusory epenthetic [ɯ] in the voiced context. It may be that the reason the results of the two studies differ has to do with the different tasks. Werker & Logan (1985) argued that, depending on the task, listeners can access different levels of information (e.g., acoustic, phonological). Auditory discrimination tasks such as AX discrimination require the ability to perceive differences in two words but it is not required to identify the differences. As for the previous identification task, participants needed to classify a sound into one of six given vowel categories. To enable listeners to classify a sound into given vowel categories, listeners have to be aware of phonemic details (Gerrit & Schouten, 1998). In the identification task by Mattingley et al. (2015), listeners seem to be using phonological knowledge while in the AX task listeners were using more low level acoustic information. Ideally, both identification and AX discrimination tasks will be employed in future research to determine whether this might be the reason for the discrepancy in results.

Chapter 4
Production Experiment

4.1 Introduction

This chapter presents a production experiment which explored whether the quality of epenthetic vowel differs across phonological environments. Previous studies of vowel epenthesis in speech production (Yazawa et al, 2015) found that preceding context had a significant effect on the quality of epenthetic vowel in real-word speech production. The question addressed in this chapter is whether the quality of the preceding consonant influences the quality of the epenthetic vowel that Japanese speakers produce between an unfamiliar sequence of consonants in the pseudo-word stimuli. Based on previous studies on Japanese loanword phonology, if a vowel is inserted, all else being equal, we would expect to observe [ɯ] after [b] and [g], [o] after [d], and [i] after [dʑ], respectively. The goal of this Chapter is also to investigate any potential differences regarding the influence of preceding consonant on epenthesis in production and perception.

4.2 Methodology

4.2.1 Stimuli

In order to investigate any potential differences regarding the influence of preceding consonant on epenthesis in production and perception, the pseudo-word stimuli for the production study had the same structure as in the perceptual experiments. The structure of the pseudo-words for the control condition is [aC_1VC_2a] where V was one of the five Japanese vowels {a, e, i, o, ɯ}. The structure of the experimental condition was [aC_1C_2a], and the consonants were selected from either the set of voiced obstruents {b, d, g, d͡ʑ} or their voiceless counterparts {p, t, k, t͡ɕ}, and $C1 \neq C2$ (e.g.[bd],[bg],[bd͡ʑ], [db],[dg][dd͡ʑ], [gb],[gd],[gd͡ʑ], [d͡ʑb],[d͡ʑd],[d͡ʑg]). There were 60 items in the form of [aC_1VC_2a] (12 consonant combinations * 5 vowels) and 12 items in the form of [aC_1C_2a] (12 consonant combinations), for a total of 72 items for each voicing type. Those items were exactly the same stimuli used for the identification and AX discrimination tasks. Additionally, 24 pseudo-word fillers were created, for a total of 96 items for each voicing type. A full list of production stimuli is given in Appendix D. Two counter-balanced lists of 228 trials were created, one for each session in

Table 4.1. The control stimuli were repeated two times while the target stimuli and fillers were repeated for a total of three times during the two sessions.

Table 4.1

Example Session Schedule in Production Experiment

			Items			
Session	Block	Name of lists	Control	Target	Fillers	# of trials
Session A	Block 1	List 1_Voiced	60	18	36	114
	Block 2	Voiceless	60	18	36	114
Session B	Block 1	List 2_Voiceless	60	18	36	114
	Block 2	Voiced	60	18	36	114
Grand Total			240	72	144	456

4.2.2 Participants

The participants in the production experiment were exactly the same as those in the perception experiment (see § 3.2.4). The production task was conducted immediately after the perception experiment.

4.2.3 Procedure

In the production experiment, speakers were orthographically presented with the same word-medial consonant sequences used in the perception study. Each pseudo-word was written in the Roman alphabet (Hepburn system). Thus, [dʑ] was spelled with *j*. For example, [dʑa] and [dʑi] were spelled with *ja* and *ji*, respectively. This is a system that all the participants are familiar with. All of the participants, however, had listened to all of the control and target stimuli in the AX discrimination task since the production task was conducted immediately after the perception task. All stimuli maintained the division between voiced and voiceless consonants, with one block of voiceless and one block of voiced stimuli. These blocks were presented in a different random order for each participant using E-prime software. A Tascam HD-P2 audio recorder with 44,100 samples/s, 16 bit/s and Beyerdynamic head-mounted microphone were used for recording, and speakers were recorded individually in a sound-proof room at the University of Canterbury.

The participants were informed about the procedure in Japanese. After seeing the stimulus on a computer screen, they were asked to pronounce aloud a stimulus as if the stimulus was a Japanese word. The stimuli were produced in the carrier sentence in Japanese characters, for example, *Kore mo* abada *desu* “This is abada, too.” Then the participants pressed any key on the keyboard to display the next stimulus. If participants realised that they had misread a stimulus, they were allowed to pronounce it one more time. They had the opportunity to take a break after the first block of stimuli. Each participant produced a randomised list of 228 words during each session. Half of the participants produced the voiced block followed by the voiceless block in the first session, with voiceless followed by voiced in the second session. The other half of the participants were given the stimuli in reverse order.

4.2.4 Analysis

In the present study, acoustic analyses were done on the control and epenthetic vowels in the test words for eight speakers (4 male, 4 female). Among the five male participants, one was excluded because he frequently misread test words (see (1)). The four female participants were chosen because, compared to the other female participants, they made fewer reading errors. Only epenthetic vowels in the voiced context were analysed in this thesis, due to vowel devoicing in voiceless consonant contexts. The demographic information of the participants is in Appendix E.

The total possible number of epenthetic vowels for each speaker was 36 (12 target stimuli*3 repetitions). For the vowels in the control stimuli, {a, e, i, o, ɯ}, there was a maximum of 24 instances for each (12 voiced consonantal environments * 2 repetitions). Thus there were 156 possible tokens for each speaker. Table 4.2 shows the total possible number of epenthetic and vowels in control words.

Table 4.2

Possible Number of Epenthetic Vowels and Vowels in Control Words

	Target		**[a]**		[e]		[i]		[o]		[ɯ]		Grand Total
	aCCa		**aCaCa**		**aCeCa**		**aCiCa**		**aCoCa**		**aCɯCa**		
C1 =bilabial	abda	3	abada	2	abeda	2	abida	2	aboda	2	abɯda	2	13
[b]	abga	3	abaga	2	abega	2	abiga	2	aboga	2	abɯga	2	13
	abdʑa	3	abadʑa	2	abedʑa	2	abidʑa	2	abodʑa	2	abɯdʑa	2	13
C1 =coronal	adba	3	adaba	2	adeba	2	adiba	2	adoba	2	adɯba	2	13
[d]	adga	3	adaga	2	adega	2	adiga	2	adoga	2	adɯga	2	13
	addʑa	3	adadʑa	2	adedʑa	2	adidʑa	2	adodʑa	2	adɯdʑa	2	13
C1=velar	agba	3	agaba	2	ageba	2	agiba	2	agoba	2	agɯba	2	13
[g]	agda	3	agada	2	ageda	2	agida	2	agoda	2	agɯda	2	13
	agdʑa	3	agadʑa	2	agedʑa	2	agidʑa	2	agodʑa	2	agɯdʑa	2	13
C1=alveo-palatal	adʑba	3	adʑaba	2	adʑeba	2	adʑiba	2	adʑoba	2	adʑɯba	2	13
[dʑ]	adʑda	3	adʑada	2	adʑeda	2	adʑida	2	adʑoda	2	adʑɯda	2	13
	adʑga	3	adʑaga	2	adʑega	2	adʑiga	2	adʑoga	2	adʑɯga	2	13
Grand Total		36		24		24		24		24		24	156

After data collection, tokens with production errors were excluded. The most common errors that were made by speakers are listed in (1).

(1) Most common production errors

(a) C1 was read incorrectly (e.g., /adid͡ʑa/ →/ad͡ʑid͡ʑa/, /abida/→/adida/)

(b) C2 was read incorrectly (e.g., /agida/→/agiba/, /aboga/→/aboba/)

(c) C1 and C2 were produced in reverse order (e.g., /abda/ →/adba/)

(d) The middle vowel was read incorrectly (e.g., /aboda/→/abuda/, /adaga/→/adoga/)

(e) The initial vowel was pronounced incorrectly (e.g., /aboga/ →/oboga/)

In some instances, the manner of articulation or phonation type of the preceding consonant was not produced as the intended consonant. For example, the /g/ in *agida* became devoiced, and the /b/ in *abaga* became devoiced. Since this study is concerned with the place of articulation of the preceding consonant, if there was agreement in the place of articulation in the intended and actually produced consonants, the produced sounds were treated as allophones of the target consonant and not as errors, and were thus not excluded. A total of 14 tokens were devoiced versions of the intended stimulus. Once errors were removed, there was a total of 1106 recorded tokens from eight speakers to analyse. Table 4.3 shows the total number of vowel tokens analysed for each speaker according to the preceding consonants. V indicates an epenthetic vowel.

Table 4.3

Number of Vowel and Epenthetic Vowel Tokens by Preceding Consonant

Speaker	Preceding Consonant	Vowels + Epenthetic Vowel (n=1106)						
		[a]	[e]	[i]	[o]	[ɯ]	V	Total
F4	[b]	6	5	5	5	6	9	36
F5		6	6	6	5	5	6	34
F7		6	6	5	6	5	7	35
F9		5	5	5	6	5	6	32
M1		6	6	6	5	6	8	37
M2		6	6	6	5	6	9	38
M3		6	6	6	4	5	7	34
M4		4	5	6	5	4	9	33
Subtotal		45	45	45	41	42	61	279
F4	[d]	5	6	5	6	4	7	33
F5		6	6	6	6	4	9	37
F7		5	3	5	5	4	7	29
F9		6	4	5	4	6	8	33
M1		4	5	6	5	5	9	34
M2		6	5	6	6	5	9	37
M3		5	5	4	6	5	9	34
M4		6	6	5	6	6	9	38
Subtotal		43	40	42	44	39	67	275
F4	[dʑ]	6	6	4	6	6	9	37
F5		6	6	6	6	6	9	39
F7		5	6	5	5	6	8	35
F9		6	6	6	5	6	8	37
M1		5	4	4	5	6	8	32
M2		6	6	6	5	6	9	38
M3		5	4	6	6	5	7	33
M4		5	6	4	6	6	9	36
Subtotal		44	44	41	44	47	67	287
F4	[g]	6	6	4	6	6	8	36
F5		6	6	5	6	6	9	38
F7		6	0	2	6	6	8	28
F9		6	3	4	6	5	8	32
M1		5	5	4	6	6	9	35
M2		6	3	1	6	6	9	31
M3		5	4	4	6	6	9	34
M4		5	1	5	6	5	9	31
Subtotal		45	28	29	48	46	69	265
Grand Total		177	157	157	177	174	264	1106

The duration of each target vowel (i.e., all of the 1106 vowels listed above) was measured and the values for F1, F2 and F3 for the stimulus vowels were extracted using a Praat script. All formant measurements were taken at the midpoint of the relevant vowel. All extracted formants were checked to ensure the validity of the values and the formant values from a few vowels were corrected.

4.3 Results

The task was designed to yield the production of epenthetic vowels and control vowels /a, e, i, o, ɯ/ after /b, d, ɡ, d͡ʑ/. These will be represented in normalized F1/F2 plots by V and the phonetic symbols /a, e, i, o, u/, respectively. In this section, the symbol 'u' in all plots stands for /ɯ/, and V stands for epenthetic vowels. All values are in Hz. The formant values were normalized and plotted by speaker using NORM (Thomas & Kendall, 2007). This is because the different physical sizes of speakers can cause different resonances. In order to compare the vowel realization of different speakers, it is important to eliminate differences between individuals’ acoustic realizations.

Recall that based on previous studies on Japanese loanword phonology, if a vowel is inserted, all else being equal, we would expect to observe [ɯ] after [b] and [ɡ], [o] after [d], and [i] after [dʑ], respectively. Participants inserted epenthetic vowels between two consonants most of the time (265 out of 288 tokens). The findings are relatively consistent with the expectations with regard to [b] and [ɡ]. However, for [d] and [dʑ], there is variability across speakers that will be discussed later.

Since the dataset is small, the results will be described in terms of observable trends in the data and not analysed statistically. I focus on two factors, formant frequency and vowel duration. I start by presenting the overall acoustic characteristics of the production tokens from the eight speakers. Then, the results of acoustic analysis in the formant plots according to the preceding consonants will be discussed followed by the vowel duration plots together. For only the palatal context will the results be discussed individually by speaker. The summarized non-normalised mean F1, F2 and F3 frequencies and duration data for each vowel for each speaker is in Appendix F.

4.3.1 Overall Acoustic Characteristics of the Production Tokens

Figure 4.1 shows vowel duration differences for all control vowels produced by all speakers. The number of tokens of each vowel are: [a] = 177, [e] = 157 [i] = 157, [o] = 177 and [ɯ] = 174. Vowel category ranked from longest to shortest is [a], [e], [o], [i] and [ɯ], consistent with vowel duration studies by Han (1962, cited in Shoji & Shoji, 2014). The lexical high back vowel [ɯ] is the shortest vowel (mean = 73.37 ms, median = 71 ms) and slightly shorter than [i] (mean = 75.50 ms, median = 74 ms). According to McGill et al. (1978), if the notches of any two box plots do not overlap, the two medians tend to be significantly different with 95% confidence level. An ANOVA showed an effect of vowel [*F* (4, 837) = 34.9, *p* < .001] and a Tukey post-hoc test showed that there was a significant effect of vowel on duration between [ɯ] and other vowels, except [i]: [a] (*p* < .001), [e] (*p* < .001), [i] (*p* = .906), [o] (*p* < .001). The differences between [i] and [o], [i] and [e] are also significant (*p* < .001), respectively, but not between [a] and [e] (*p* =.134) or [e] and [o] (*p* = .269).

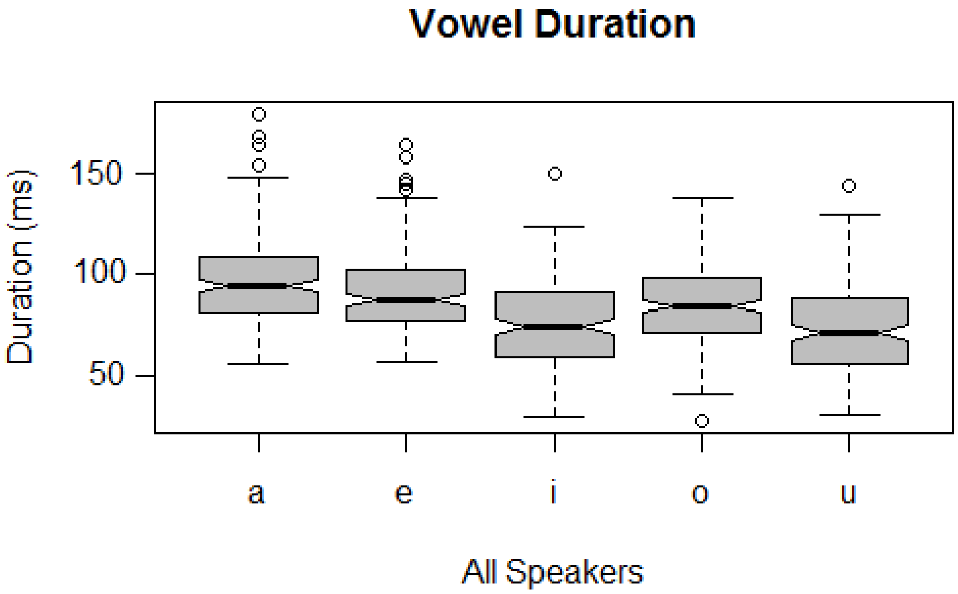

Figure 4.1. Boxplots of durations of the lexical vowels from all eight speakers.

Normalised ellipse plots in Figure 4.2 show the overall F1/F2 spaces with mean values and 2.0 standard deviations for each control vowel across all of the eight speakers. It can be seen that the [i] vowel, slightly overlaps with [e] and [ɯ], and [o] also overlaps with [ɯ]. However, these vowel spaces are consistent with the Japanese vowel space presented in Vance (2008) (see Chapter 2). 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 almost equal in backness.

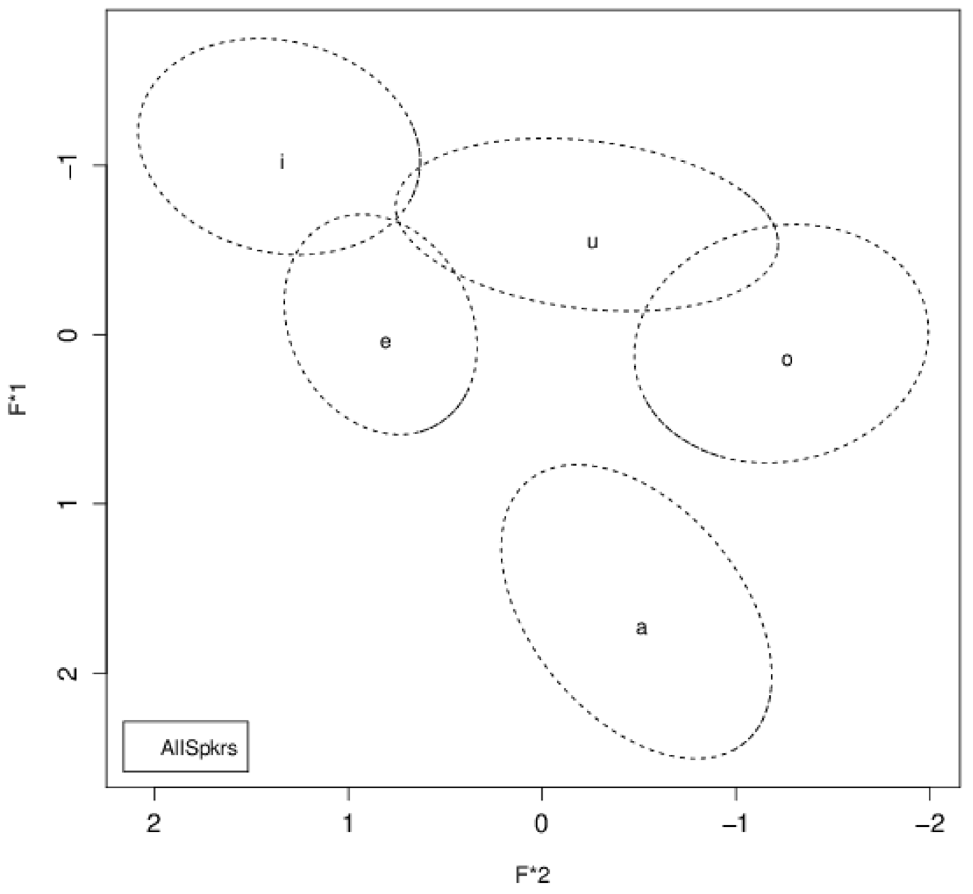

Figure 4.2. Normalised mean F1 and F2 values for each lexical vowel from all eight speakers.

4.3.2 Epenthetic Vowels in the Labial Context

We begin by looking at the quality of the epenthetic vowel in the labial context. We would expect the quality of epenthetic vowel in this context to be similar to the lexical vowel [ɯ]. When the results of vowel production for each speaker are compared, a pattern which is similar to loanword adaptation in Japanese is observed across speakers, i.e. speakers utilise the [ɯ] vowel. However, for two speakers, some epenthetic vowels fall within an unpredicted area.

Let us first consider the speakers whose quality of epenthetic vowel in the labial context has a quality close to the vowel [ɯ]. Figure 3.3 presents a vowel space plot of both lexical vowels (in [bVC]-forms) and epenthetic vowels (V in [bC]-forms) for six speakers

(F4, F5, F7, M1, M2, M4). The normalised ellipses show the overall F1/F2 spaces with mean values and 2.0 standard deviations for each vowel. For these six speakers, the epenthetic vowel /V/ (=smaller oval in comparison to [ɯ]) overlaps with their lexical vowel /ɯ/ (=bigger oval). The number of tokens of each vowel are: [a] =34, [e] =34, [i] =34, [o] = 31, [ɯ] =32 and V = 48.

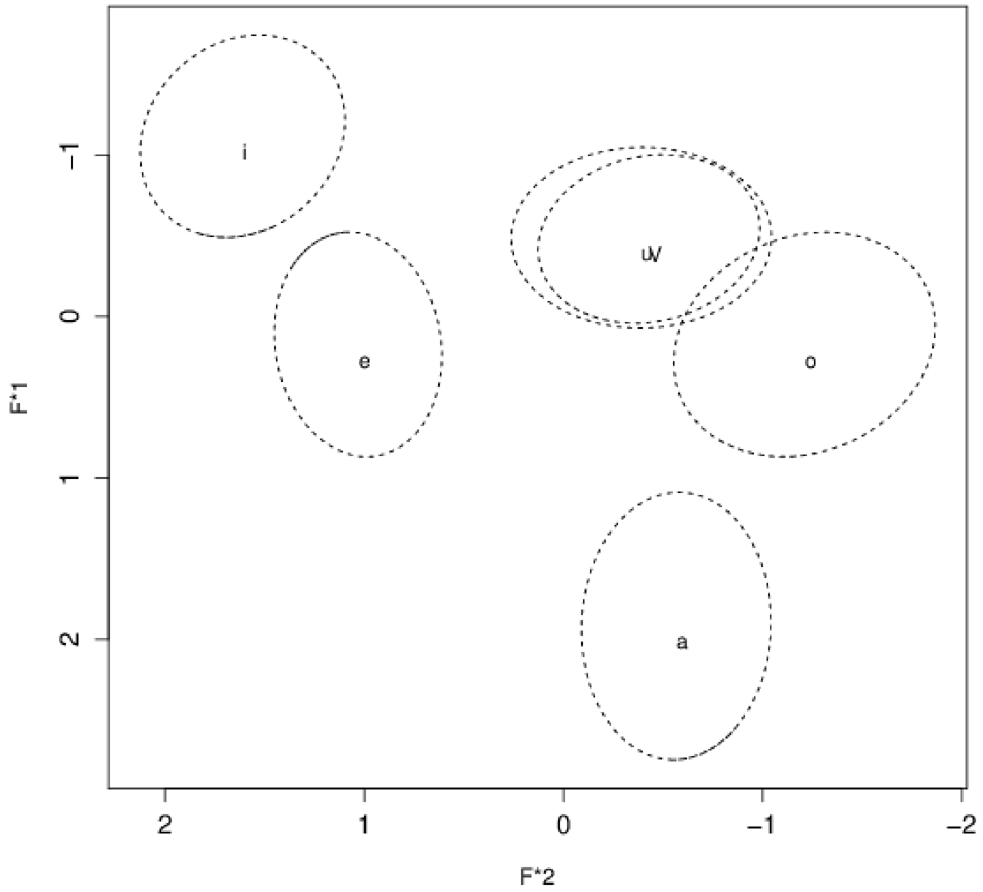

Figure 4.3. Normalised mean F1 and F2 values for each lexical vowel and the epenthetic vowel from six speakers (F4, F5, F7, M1, M2, and M4) in the labial context.

Note that in Figure 4.3, the [ɯ] vowel overlaps slightly with the mid back [o]. This overlap in the combined vowel chart is most likely due to the productions of one speaker, F4, whose [ɯ] vowels are typically more back than those of other speakers, and whose [o] vowels are higher than those of other speakers. Although this causes overlap in the combined plot, each speaker's individual ellipses are distinct.

For two speakers, F9 and M3, the produced epenthetic vowels overlap with not only the lexical [ɯ] but also with other vowels, as shown in Figures 4.4 and 4.5, respectively. For speaker F9, the epenthetic vowels /V/ overlap with [ɯ] and [i]. Note that the width of the ellipse is due to two tokens of epenthetic vowels that seem to be clear examples of [i], while there are three epenthetic vowel tokens that seem to be clear examples of [ɯ]. It may be that the insertion of unexpected [i] vowels was due to the influence of a similar vowel occurring in a preceding word. For one token ‘abja’, it is possible that the previous word ‘abiga’ influenced the epenthetic vowel, making it more like [i], however, two other target words with [i] were not preceded by a stimulus with [i].

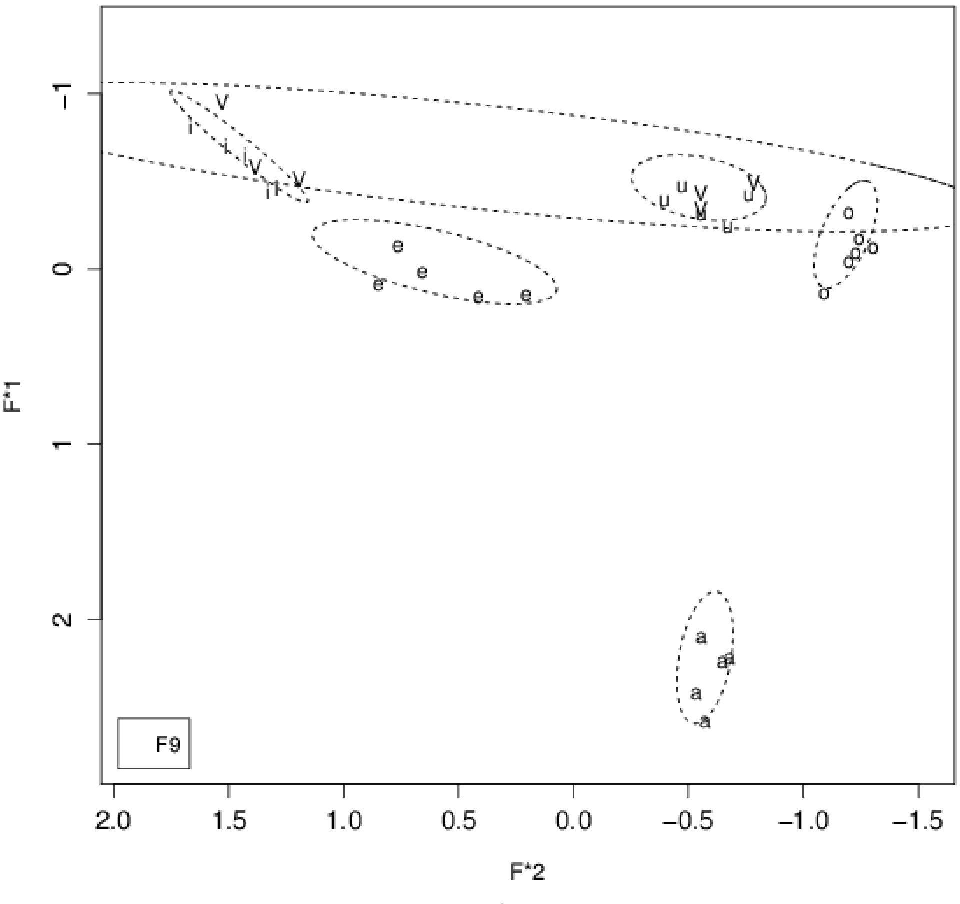

Figure 4.4. Normalized individual vowel plot for speaker F9 in the labial context.

For speaker M3, one epenthetic vowel token /V/ is in the lexical [a] area, otherwise the epenthetic vowels are close to the lexical [ɯ] vowels.

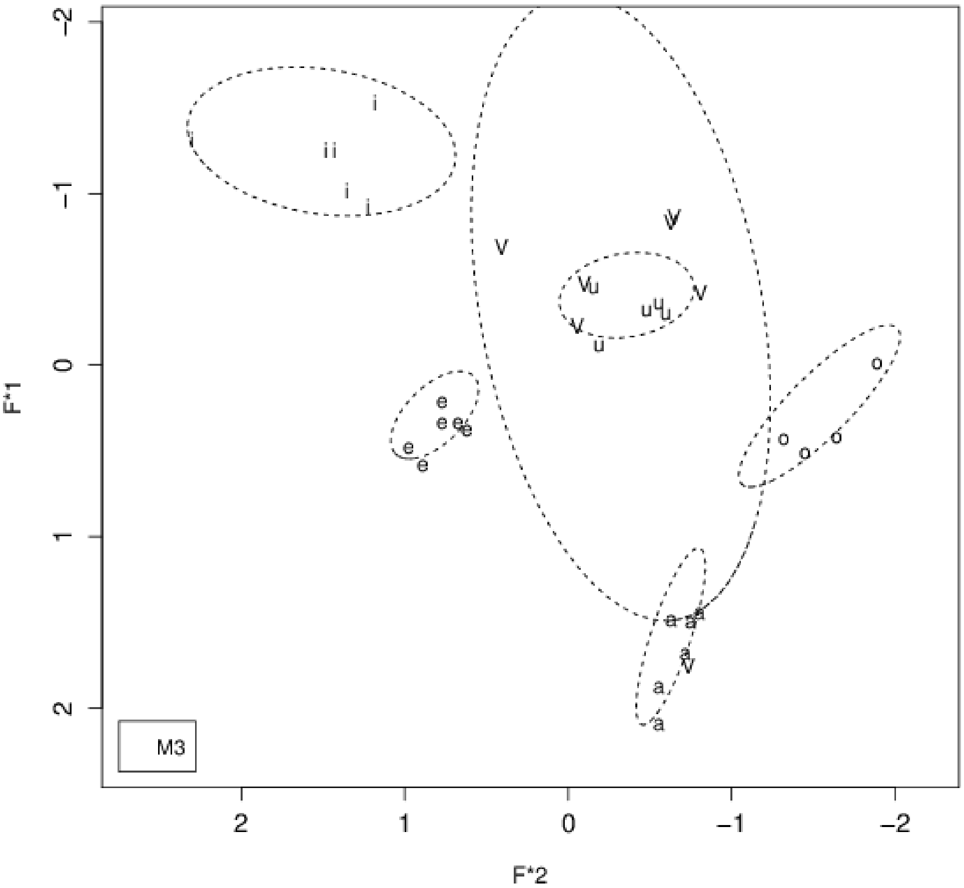

Figure 4.5. Normalized individual vowel plot for speaker M3 in the labial context.

In addition to looking at vowel quality, it is worth investigating whether the epenthetic vowel corresponds to the vowel with the shortest duration among Japanese's five vowel qualities.

Figure 4.6 presents a set of box plots for the duration (ms) of lexical vowels (in [bVC]-forms) and epenthetic vowels (/V/ in [bC]-forms) for seven speakers (F4, F5, F7, M1, M2, M3 and M4). The mean duration of the epenthetic vowel is 70.12 ms (median = 71 ms) which is closest in duration to the lexical vowel [ɯ], whose mean duration is 68.67 ms (median = 69 ms). An ANOVA showed an effect of vowel [*F* (5, 241) = 15.55, *p* < .001] and a Tukey post-hoc test showed that there was a significant effect of vowel on duration between [ɯ] and other vowels, except [i] and V: [a] (*p* < .001), [e] (*p* < .001), [i] (*p* = .062), [o] (*p* <

.01), V (*p* = .999). For these seven speakers, it can be seen that the vowel inserted after a labial consonant is most similar in duration to the shortest vowel which is [ɯ].

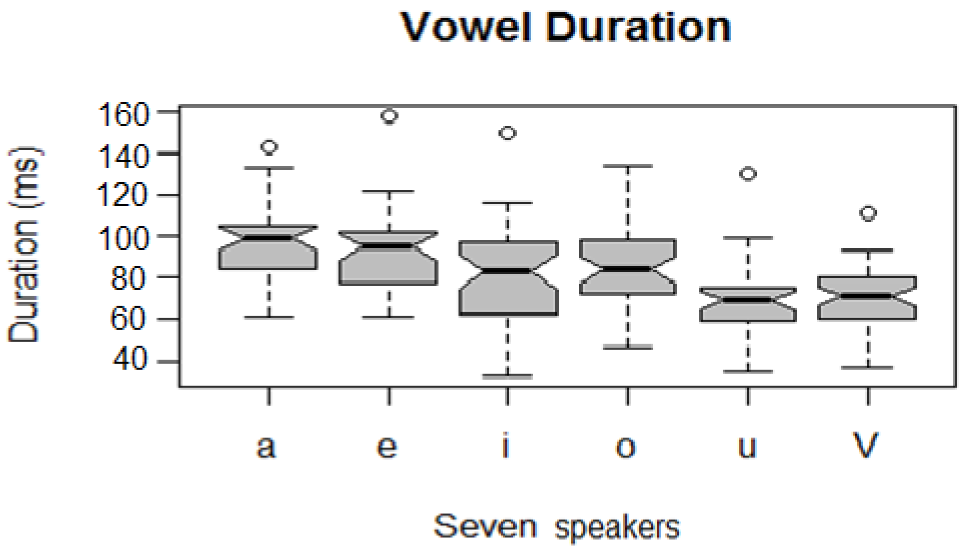

Figure 4.6. Boxplots of six vowels across seven speakers in the labial context.

In terms of vowel duration, speaker F9 behaves differently from the others. First, the mean duration of the lexical vowel [i] is 102.6 ms (median = 105 ms) which is the shortest vowel for her. Second, F9 produced the epenthetic vowel with longer duration (mean = 131.5, median = 129.5) than any other vowel as shown in Figure 4.7. Since sample size is less than 10 tokens for each vowel, a statistical analysis was not conducted for speaker F9.

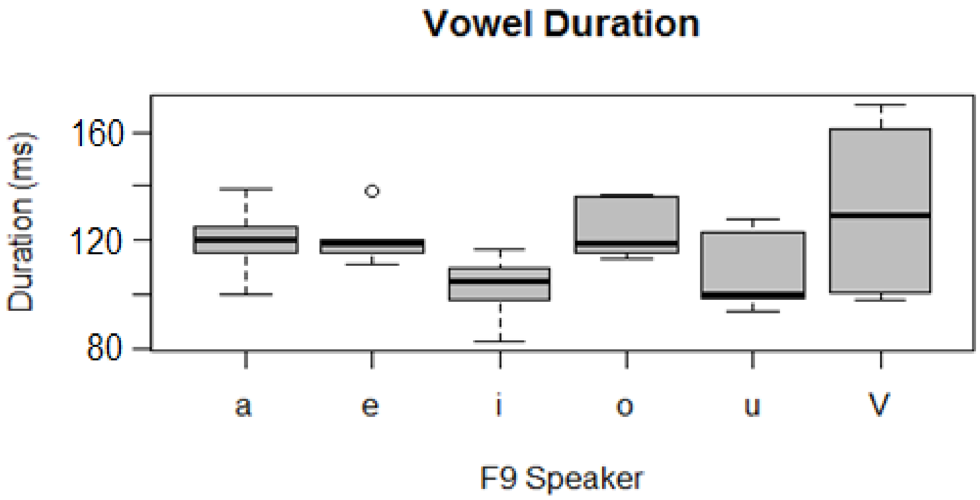

Figure 4.7. Boxplots of six vowels for speaker F9 in the labial context.

4.3.3 Epenthetic Vowels in the Velar Context

Next, the quality of the epenthetic vowel in the velar context [g] will be presented. As is the case in the labial context, we would expect that the quality of epenthetic vowel to be similar to the lexical vowel [ɯ]. Figure 4.8 shows the overall F1/F2 spaces with mean values and 2.0 standard deviations for both lexical vowels (in [gVC]-forms) and epenthetic vowels (in [gC]-form) for all eight speakers. The number of tokens of each vowel are: [a] = 45, [e] = 28, [i] = 29, [o] = 48, [u] = 45 and V = 69.

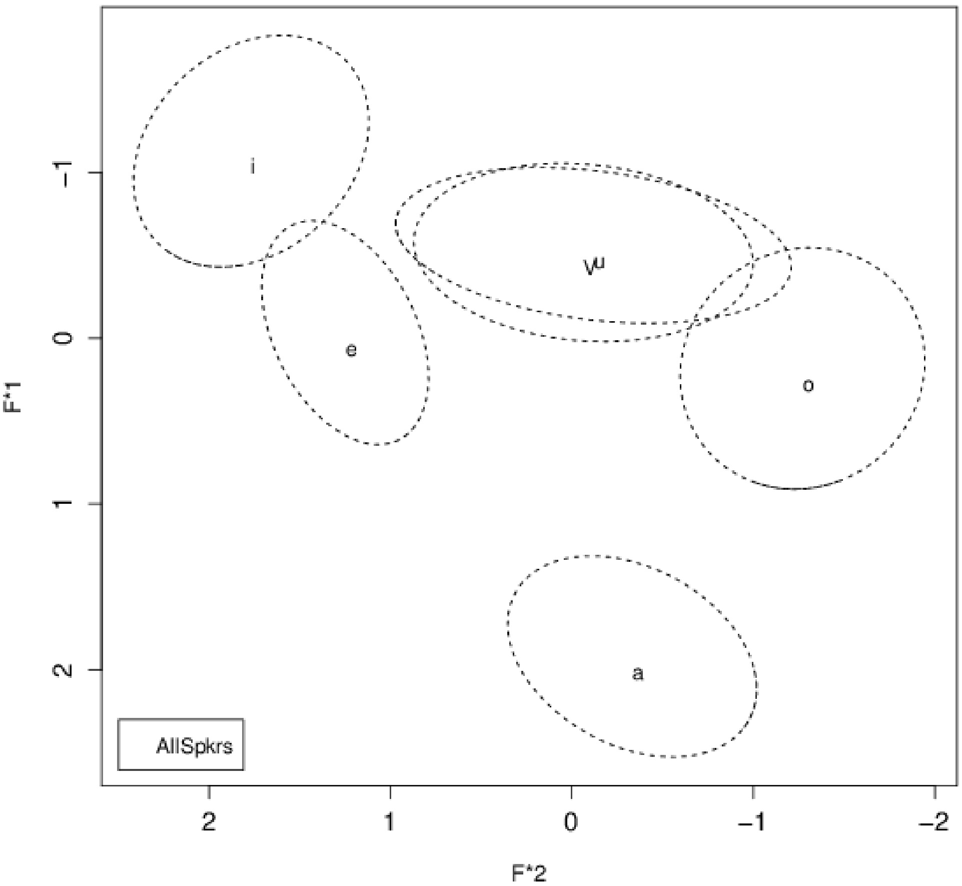

Figure 4.8. Normalised mean F1 and F2 values for each lexical vowel and the epenthetic vowel from all eight speakers in the velar context.

As can be seen in Figure 4.8, the epenthetic vowel space V (vertically longer circle in comparison to [ɯ]) is almost entirely overlapping with the vowel space of [ɯ] (longer horizontal circle).

Let us now consider vowel duration after the velar consonant. Figure 4.9 presents a set of box plots for the duration of lexical vowels (in [gVC]-forms) and epenthetic vowels (V in [gC]-forms) for all eight speakers.

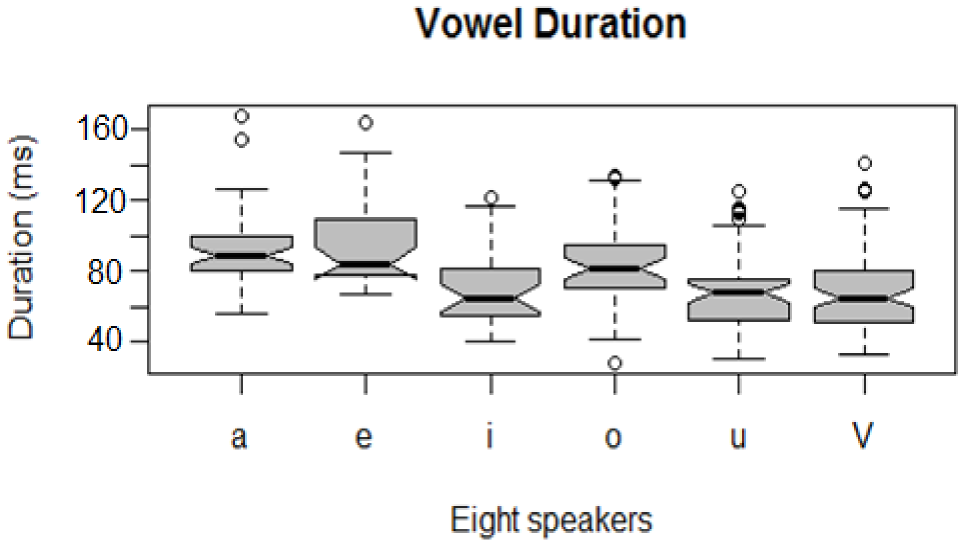

Figure 4.9. Boxplot of six vowels across eight speakers in the velar context.

The mean duration of the epenthetic vowel is 69.40 ms (median = 65 ms). This is similar to that of [ɯ] (69.73 ms, median = 67.5ms), the predicted quality of the epenthetic vowel and the high front vowel [i] (69.89 ms, median = 65 ms). An ANOVA showed an effect of vowel [*F* (5, 259) = 10.12, *p* < .001]. However a Tukey post-hoc test showed that there were no significant differences between [ɯ], [i] and V (*p* = 1.00). 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].

4.3.4 Epenthetic Vowels in the Alveolar Context

In this section, the epenthetic vowel in the alveolar context is examined. Normalised ellipse plots in Figure 4.10 show the overall F1/F2 spaces with mean values and 2.0 standard deviations for lexical and epenthetic vowels from all of the eight speakers. Recall that based on previous studies, the epenthetic vowel is predicted to be most similar to [o]. The number of tokens of each vowel are: [a] = 43, [e] = 40, [i] = 42, [o] = 44, [u] = 39 and V = 67.

As can be seen in the figure, the vowel space of the epenthetic vowel overlaps with the high vowel [ɯ] and the mid vowel [o]. The plot indicates that there is variability across the speakers as to which vowels they seem to epenthesize. When the results for each speaker are compared, they can be classified into three general patterns. Three speakers (F7, M1, M3) use [o] as the epenthetic vowel which is what we would expect to observe. Three other speakers (F4, F9, M2) use [ɯ], and two speakers (F5, M4) fall somewhere in between.

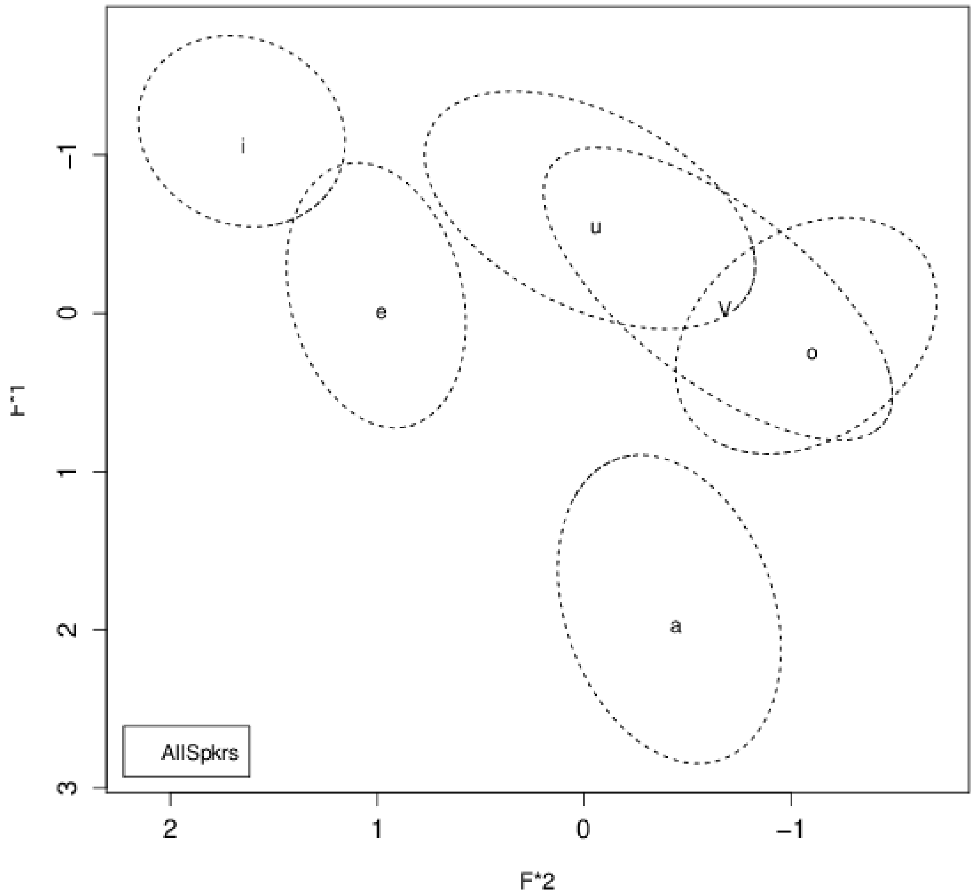

Figure 4.10. Normalized mean F1/F2 frequencies for lexical vowels and the epenthetic vowel from all speakers in the alveolar context.

To begin, the current study considers the quality of the epenthetic vowels which exhibit a pattern similar to loanword adaptation in Japanese (i.e. speakers utilise [o]). Figure 4.11 shows the vowel space of three speakers, one female (F7) and two males (M1 and M3). Their epenthetic vowels are very similar to the acoustic space of the mid back [o]. Note the smaller oval circle is for the mid back [o] whereas the bigger oval circle is for the epenthetic vowel V. The number of tokens of each vowel are: [a] = 14, [e] = 13, [i] = 15, [o] = 16, [ɯ] = 14, V = 25. Although the space of the epenthetic vowel is larger than that of the mid back [o], it overlaps only with the acoustic space of [o].

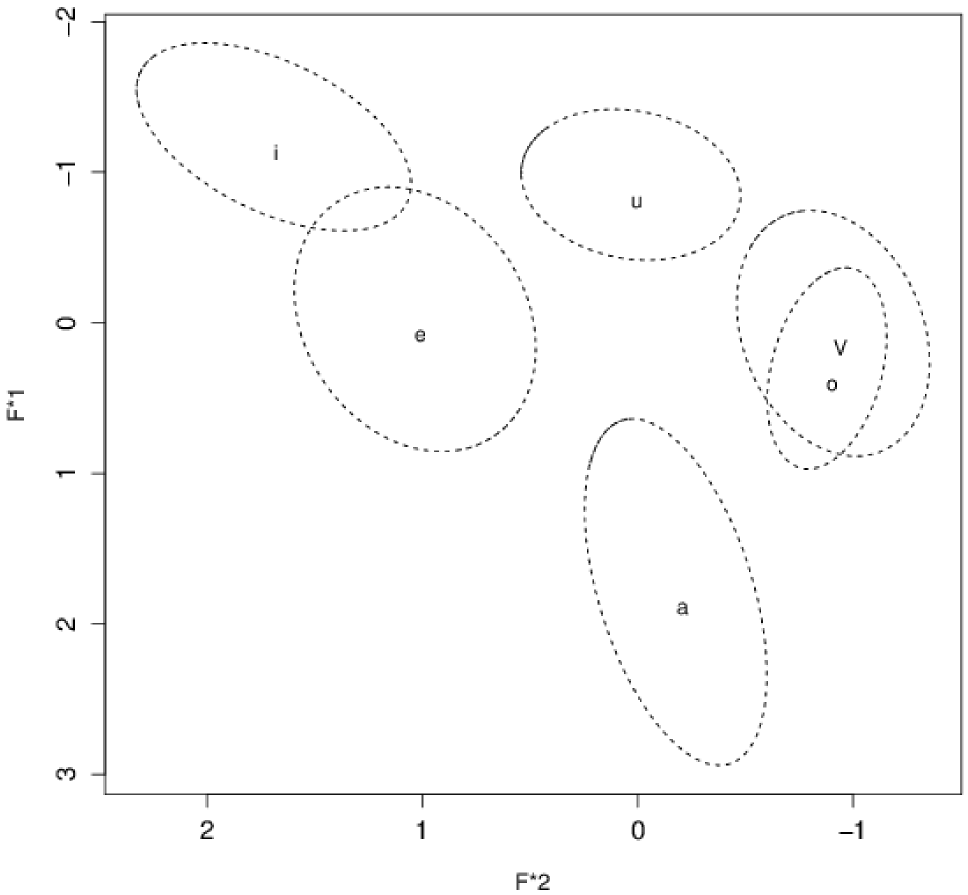

Figure 4.11. Normalized mean F1/F2 frequencies for vowels for three speakers (F7, M1 and M3) in the alveolar context.

Figure 4.12 shows a formant plot for three speakers, 2 female (F4 and F9) and 1 male (M2), who seem to epenthesize the high vowel [ɯ] in this context. The number of tokens for each vowel is: [a] = 17, [e] = 15, [i] = 16, [o] = 16, [ɯ] = 15 and V = 24. The epenthetic vowel and the high vowel [ɯ] completely overlap, indicating that the quality of the epenthetic vowel is similar to the acoustic space of the [ɯ] vowel for these three speakers.

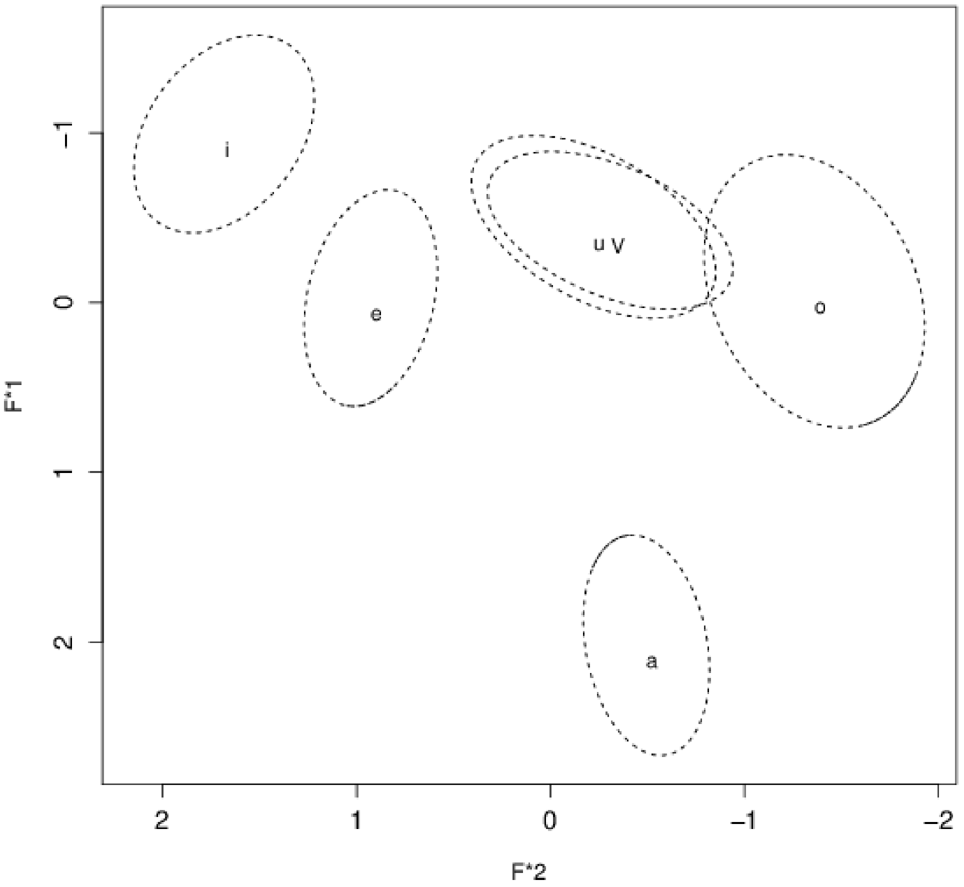

Figure 4.12. Normalised mean F1 and F2 values for each lexical vowel and the epenthetic vowel from three speakers (F4, F9, and M2) in the alveolar context.

Note that the ellipse of the epenthetic vowel V in Figure 4.12 overlaps slightly with the [o]. This may indicate that the epenthetic vowel was produced as [o] by some speakers. In order to verify whether all V tokens were the high vowel [ɯ] or whether some were actually [o], two native speakers of Japanese were asked to listen to and identify the quality of all epenthetic tokens in this context. All tokens were judged to be [ɯ]. For speaker M2, two [o] tokens were very close to the vowel space of [ɯ] and both were produced in the word ‘adoja’. The following palatal consonant [dʑ] may have resulted in a higher F2 and thus more fronted vowel. The same holds true for the F4 speaker.

Finally we look at the plots for two speakers (F5 and M4) that show the epenthetic vowel in the alveolar context to be between the vowels [o] and [ɯ]. Figure 4.13 shows that epenthetic vowels for female speaker F5 occur from the high central area to the mid back area in the chart (six token of [o], two of [ɯ], and one outlier).

.

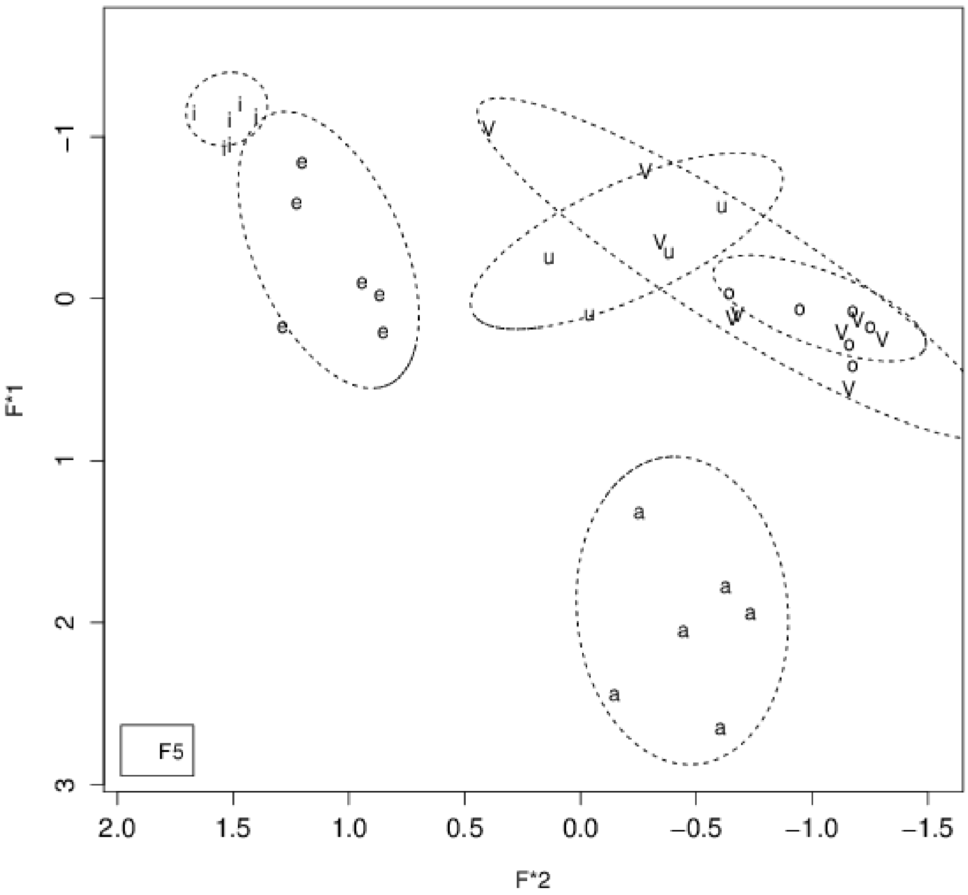

Figure 4.13. Normalized individual vowel plot for speaker F5 in the alveolar context.

Figure 4.14 shows that the epenthetic vowels for male speaker M4 occur around the high back and mid back area in the chart. The F2 space of the [ɯ] is considerably broader than in previous cases which may be due to the fact that the two [ɯ] tokens with high F2 frequency were produced with an alveo-palatal as the second consonant (i.e., [adɯdʑa]).

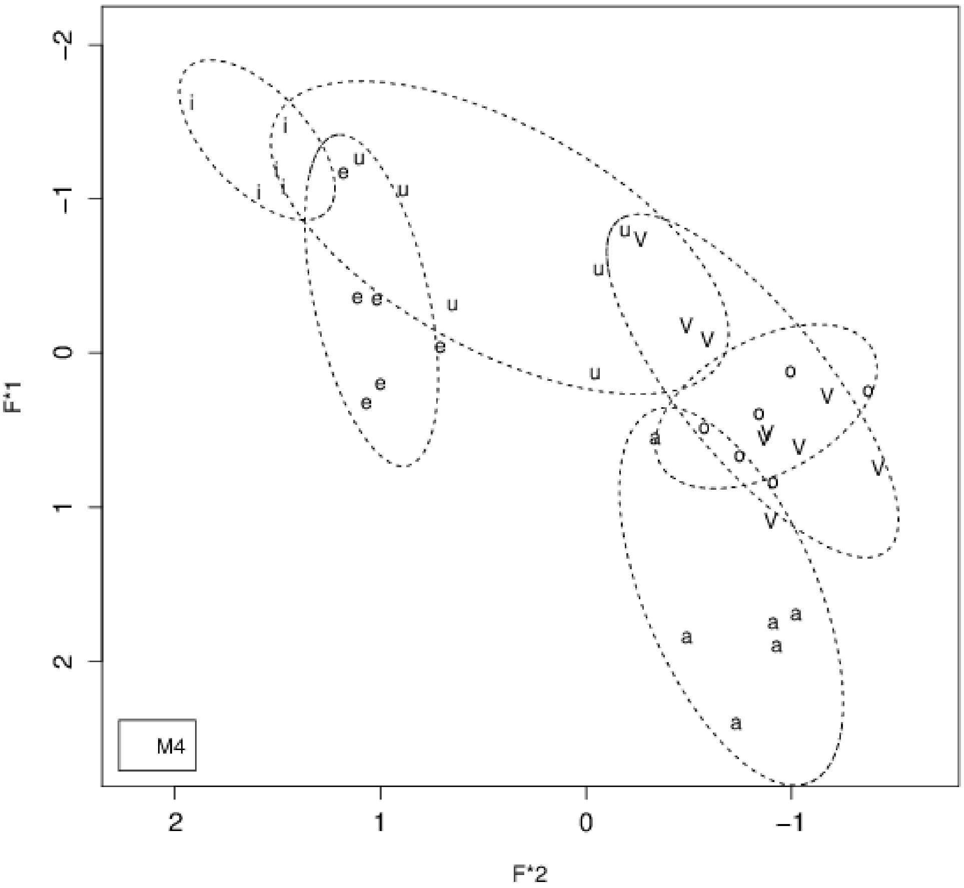

Figure 4.14. Normalized individual vowel plot for speaker M4 in the alveolar context.

Let us now consider the duration of vowels in this context. Figures 4.15 through 4.17 present the duration of lexical vowels (in [dVC]-forms) and epenthetic vowels (V in [dC-forms] for three groups that correspond to the groupings above. The first group is for those speakers who use the contextually appropriate epenthetic vowel [o] in Figure 4.15.

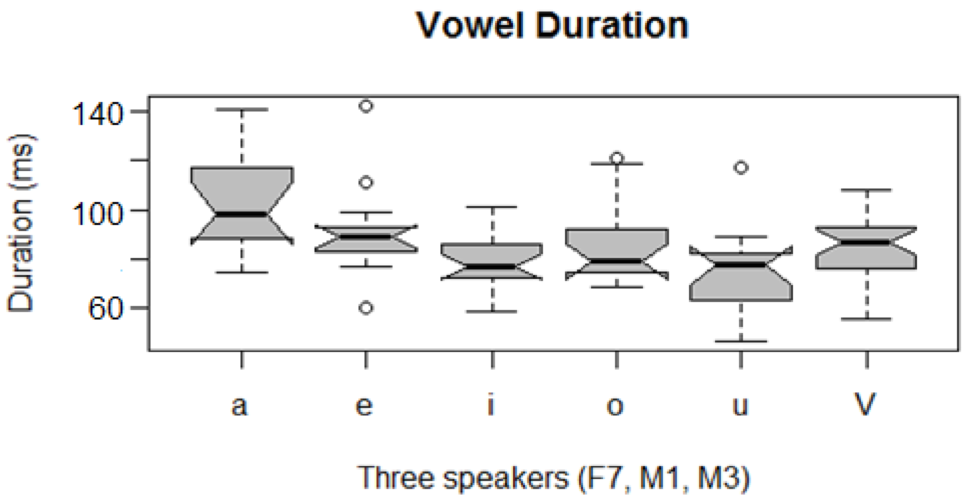

Figure 4.15. Boxplots of six vowels across three speakers who use [o] in the alveolar context.

According to Figure 4.15, the contextually appropriate epenthetic vowel is predicted to be similar to the lexical vowel [o], whose mean duration is 85.31 ms (median = 79 ms); the mean duration of the epenthetic vowel is 85ms (median = 87 ms). The mean duration of the lexical vowel [ɯ] is the shortest for this group with 75.57ms (median = 77.5 ms) while the second shortest vowel in mean duration is the lexical vowel [i] (78.20 ms, median = 77 ms). An ANOVA showed an effect of vowel [*F* (5, 91) = 4.998, *p* < .001]. A Tukey post-hoc test showed that there was a significant effect of vowel on duration only between [a] and [i], [ɯ], V, respectively: [i] (*p* < .01), [ɯ] (*p* < .001), V (*p* < .05) and the duration of the epenthetic vowel in this context is most similar in duration to [o] (*p* = 1.0). That is, the epenthetic vowel is not matching the duration of the shortest vowel in this context though the difference beween them is not statistically significant.

For the second group which used the vowel [ɯ] as epenthetic after the alveolar [d] in Figure 4.16, the mean duration of [ɯ] is 86.13 ms (median = 79) while the high front vowel [i] is slightly shorter (85.62 ms, median = 85 ms). The mean duration of their epenthetic vowel is 84.25 ms (median = 72 ms). The mean duration of [o] is 89.87 ms (median = 85 ms). An ANOVA did not show an effect of vowel [*F* (5, 97) = 1.395, *p* = .233].

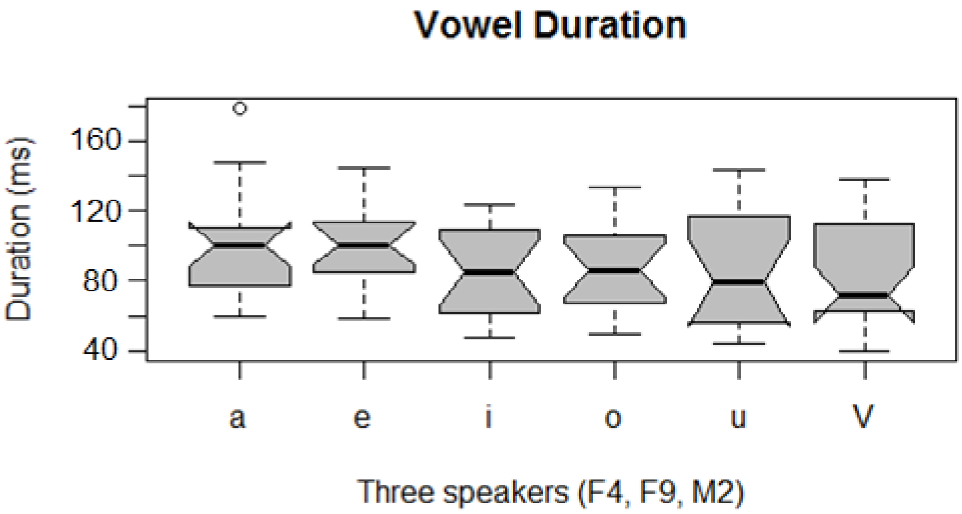

Figure 4.16. Boxplots of six vowels across two speakers who use [ɯ] in the alveolar context.

Figure 4.17 shows boxplots for the third group whose epenthetic vowels fall around the [ɯ] and [o] areas after the alveolar [d]. The mean duration of [ɯ] is 84.30 ms (median = 88 ms) while that of the mid back vowel [o] is 80.08 ms (median = 83 ms). The shortest mean duration is 79.66 ms (median = 81ms) for the mid front vowel [e]. The epenthetic vowel [V] is shorter than any other lexical vowels (mean = 74.88 ms, median = 74 ms). An ANOVA showed an effect of vowel [*F* (5, 69) = 3.039, *p* < .05]. However, a Tukey post-hoc test showed there was a significant difference only between [a] and V (*p* < .01).

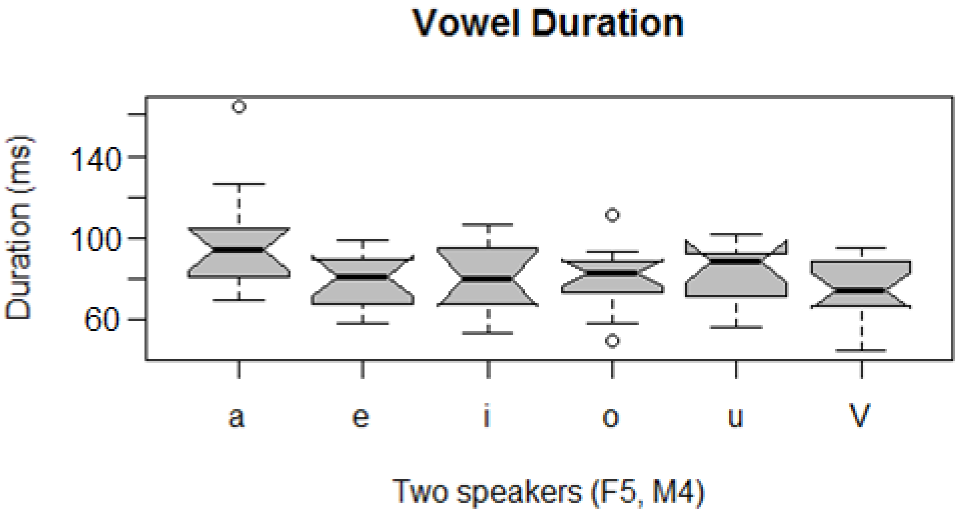

Figure 4.17. Boxplots of six vowels across two speakers who use [ɯ] and [o] in the alveolar context.

According to the duration figures, the epenthetic vowel does not consistently correspond to the shortest vowel in this context. This suggests that the duration of lexical vowels is not a crucial factor in predicting which vowel will be used as epenthetic.

4.3.5 Epenthetic Vowels in the Palatal Context

In the palatal [dʑ] context, we would expect that the quality of epenthetic vowel to be similar to the lexical vowel [i]. However, this is not true for the majority of speakers, and epenthetic vowels that are not predicted by loanword phonology are observed. Only one speaker, F7, seems to utilise the contextually appropriate vowel [i], while one male speaker, M2, produces epenthetic vowels that are similar to the vowel [ɯ]. For the other speakers, there is great deal of variability. As a result of this variability, the vowel spaces of each single speaker will be presented.

Figures 4.18 through 4.25 (8 plots) show the results for individual lexical vowels (in [dʑVC]-forms) and epenthetic vowels (in [dʑC]-forms) by each single speaker. Let us start with the female speaker (F7) who consistently uses the high front [i] in this context, as shown in Figure 4.18, with one exception where [ɯ] is used.

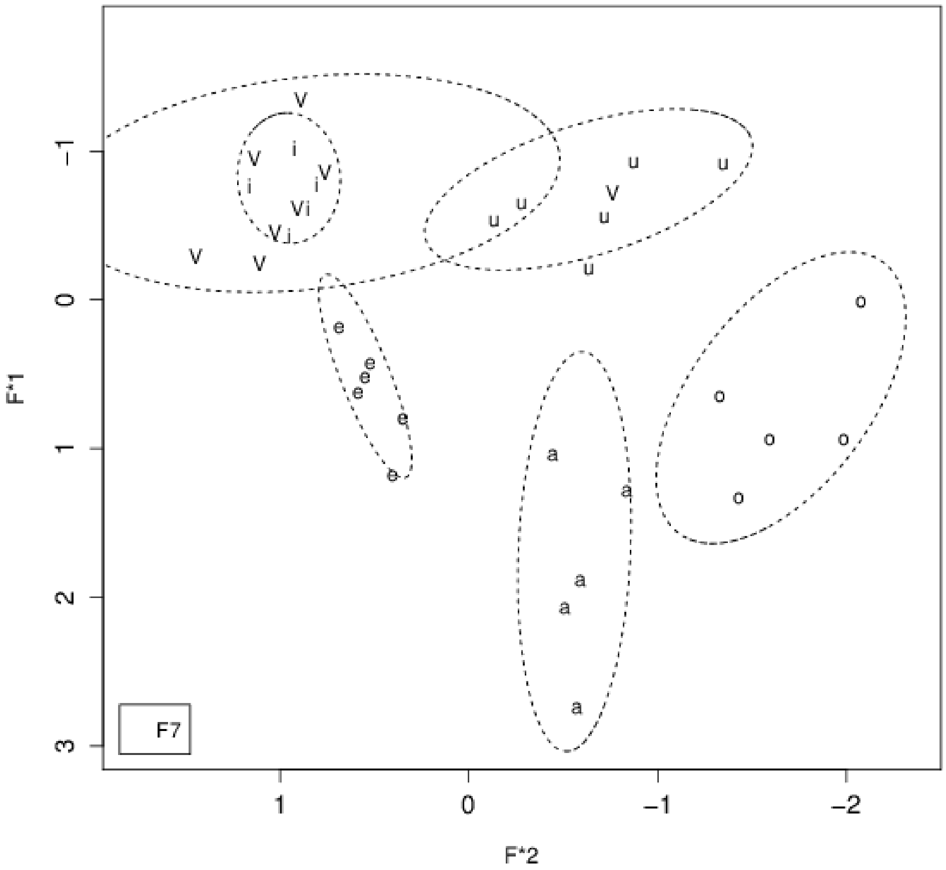

Figure 4.18. Normalized individual vowel plot for speaker F7 in the palatal context.

Next, the speaker who uses the epenthetic vowel [ɯ] in this context is considered. Figure 4.19 shows the vowel space of speaker M2.

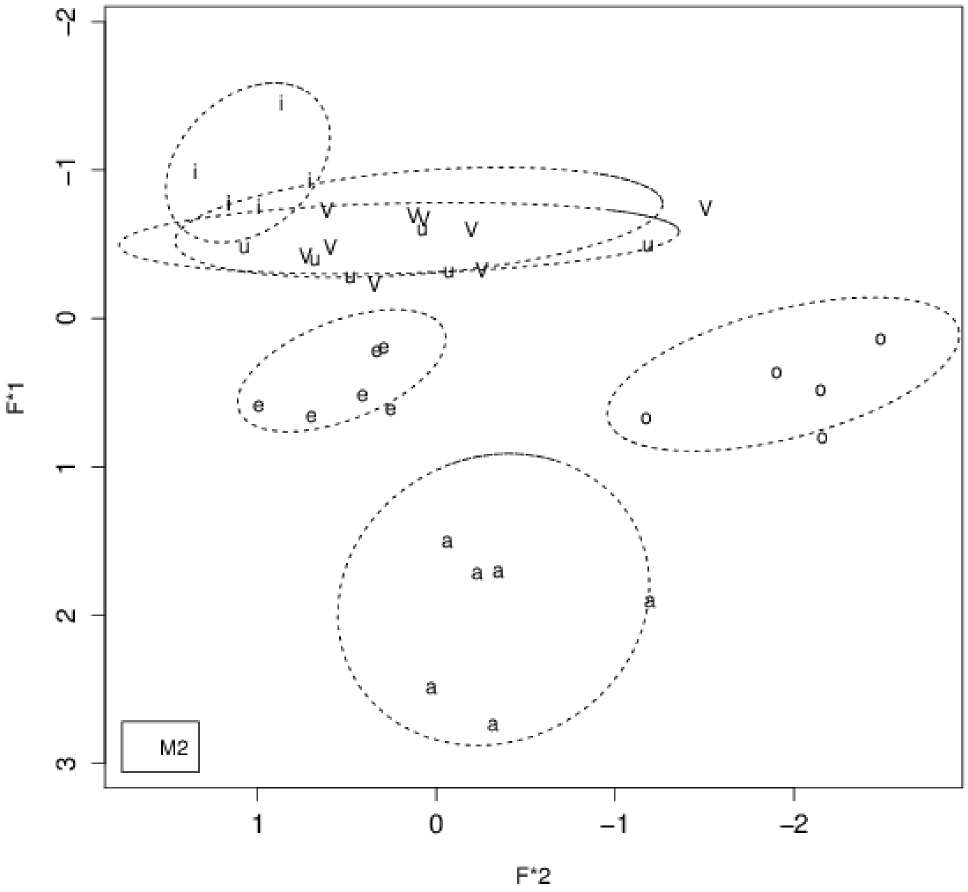

Figure 4.19. Normalized individual vowel plot for speaker M2 in the palatal context.

Figure 4.20 shows the vowel space for speaker F5. In contrast to speaker F7, she never inserted the contextually appropriate vowel [i] in the palatal context. Rather, she mainly used the vowels [ɯ] (three instances) and [o] (four instances), and [a] and [e] once each.

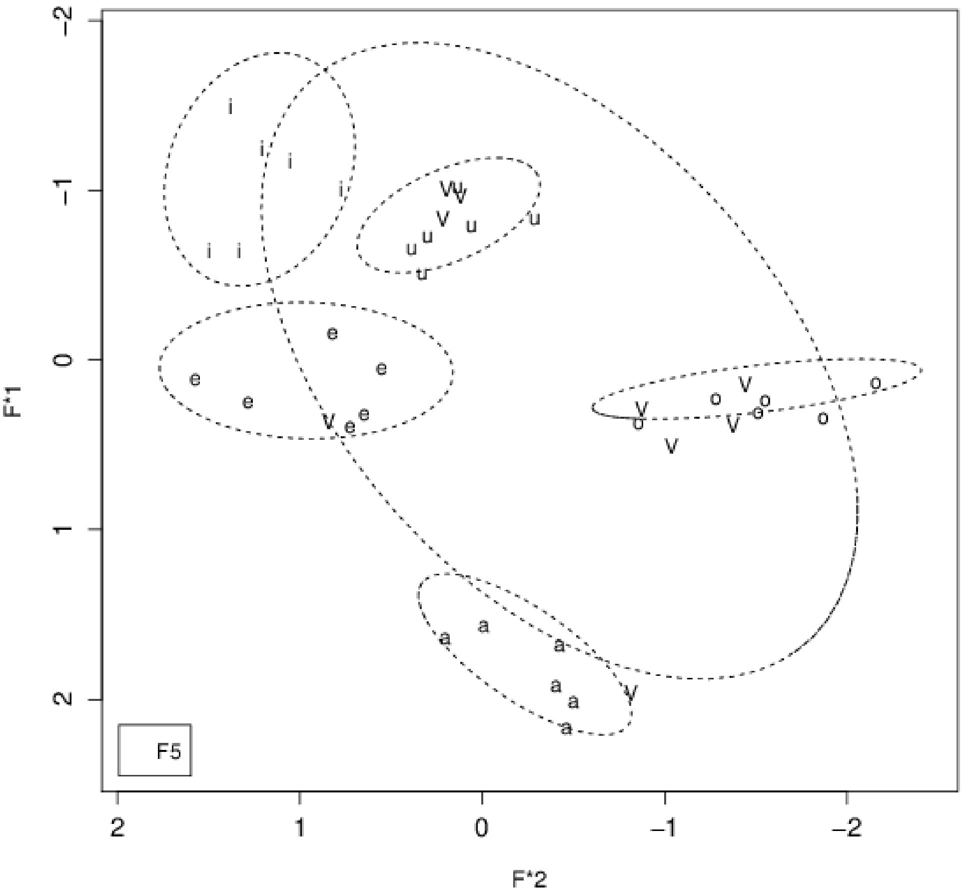

Figure 4.20. Normalized individual vowel plot for speaker F5 in the palatal context.

Speaker M1 never used [ɯ], as shown in Figure 4.21. However, all other vowel qualities were used as epenthetic (one token of [i], three of [e], one of [o], two of [a], and one intermediate between [i] and [e]).

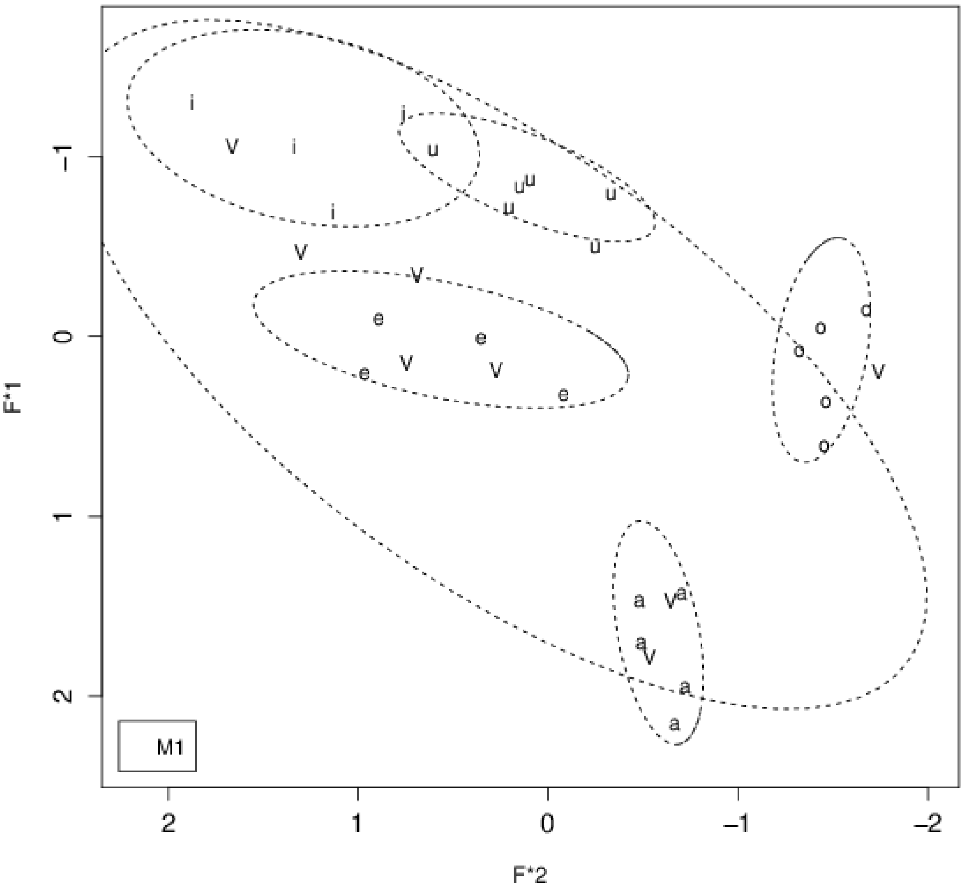

Figure 4.21. Normalized individual vowel plot for speaker M1 in the palatal context

Figure 4.22 shows the vowel space for F4 speaker. As can be seen, this speaker used only non-low unrounded vowels as epenthetic, covering the space of her lexical [i], [e], and [ɯ].

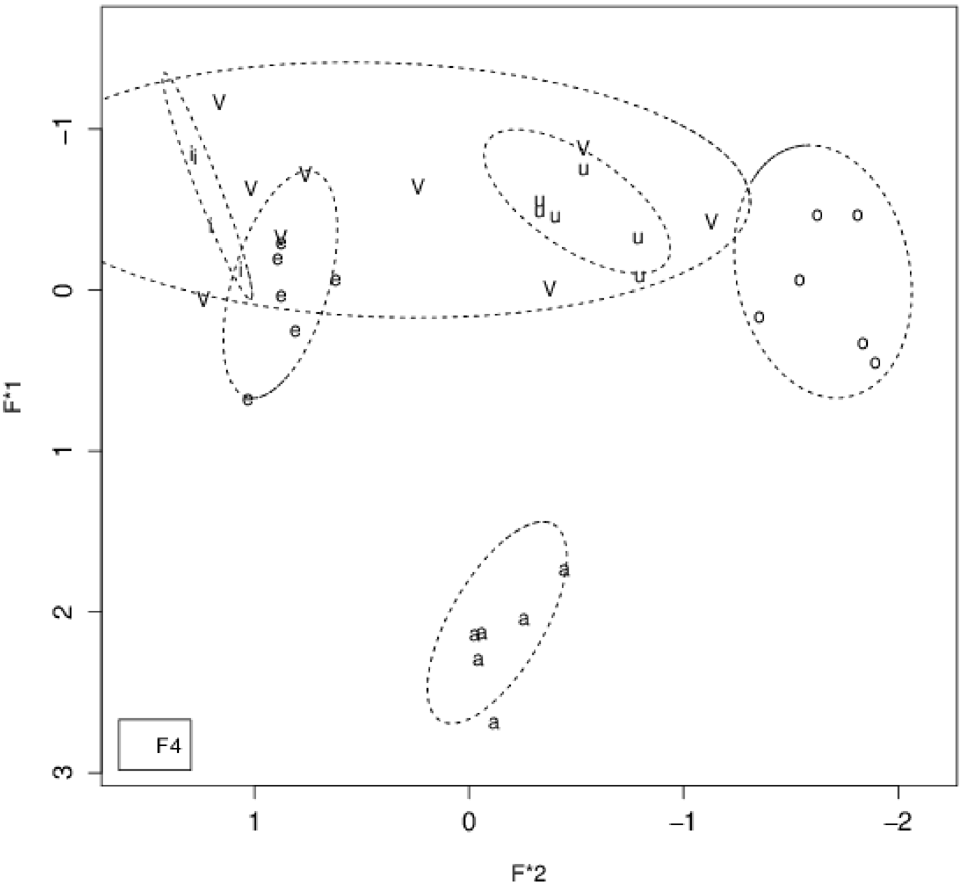

Figure 4.22. Normalized individual vowel plot for speaker F4 in the palatal context

For F9, she did not use the mid back [o], but used other all other vowels [a] (three tokens), [e] (one token), [i] (two tokens), and [ɯ] (two tokens), as shown in Figure 4.23.

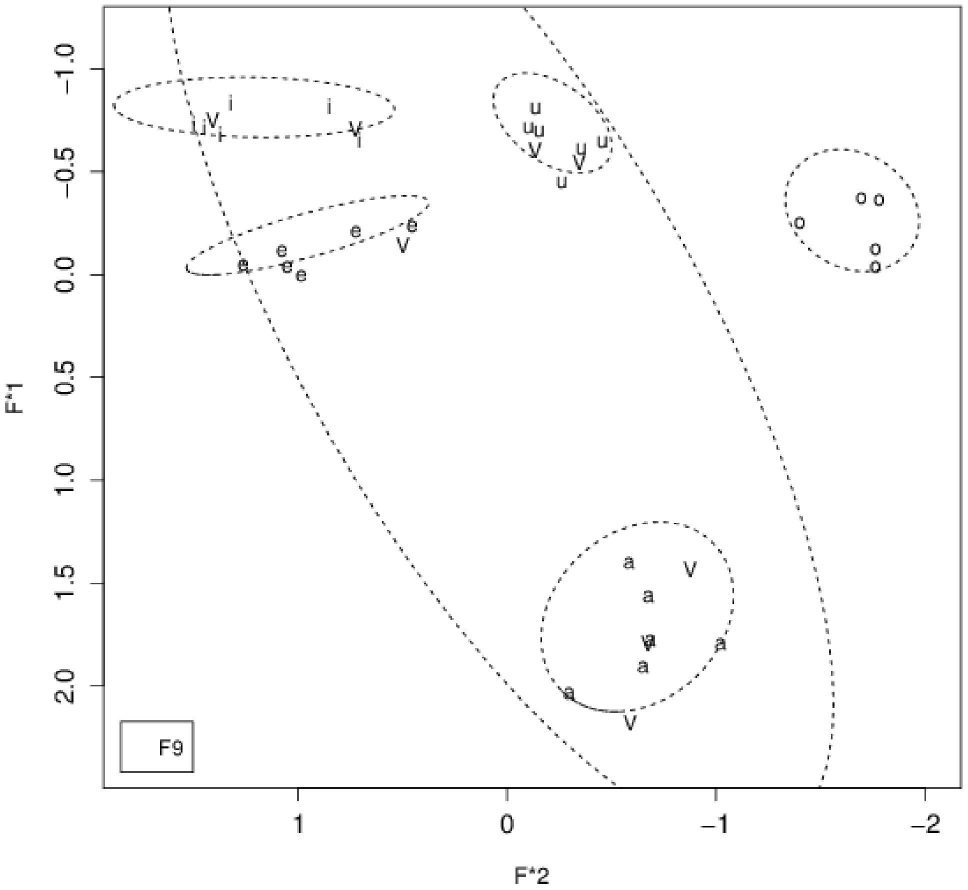

Figure 4.23. Normalized individual vowel plot for speaker F9 in the palatal context

Figure 4.24 shows the vowel space for M3 speaker who used [i] (one token), [e] (three tokens) and [a] (three tokens), but not [ɯ] or [o].

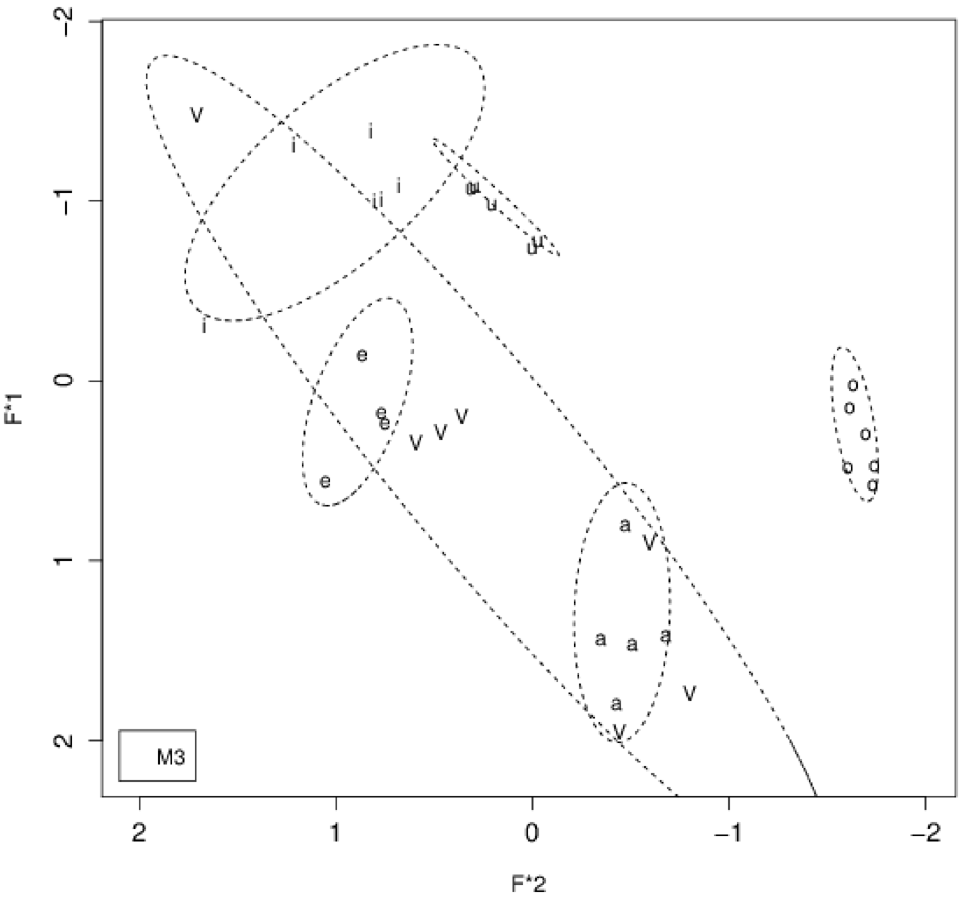

Figure 4.24. Normalized individual vowel plot for speaker M3 in the palatal context.

Figure 4.25 shows the vowel space for M4 who is more likely to use central sounds rather than the high vowel [i] (three of [e], two of [a], three intermediate between [ɯ] and [o], and one outlier).

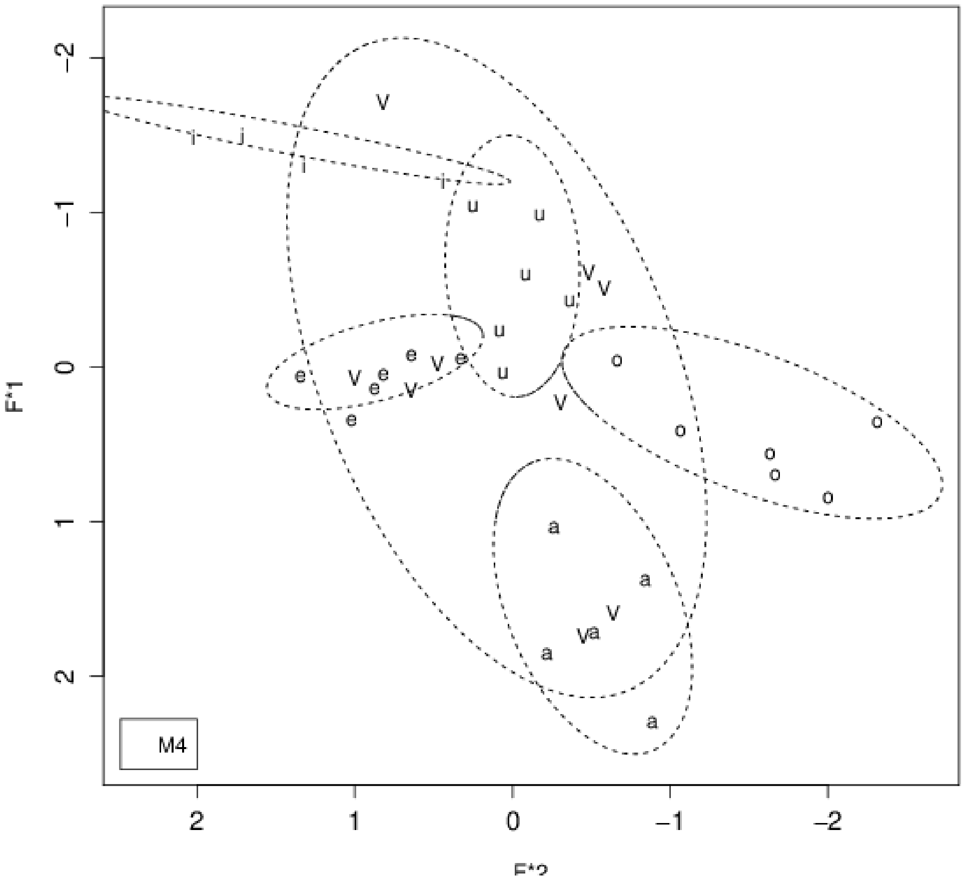

Figure 4.25. Normalized individual vowel plot for speaker M4 in the palatal context.

Turning now to vowel duration, Figure 4.26 presents a set of box plots for the duration of lexical vowels (in [dʑ VC]-forms) and epenthetic vowels (/V/ in [dʑC]-forms) for seven speakers (F4, F5, F7, M1, M2, M3, M4). Since the vowel duration of speaker F9 differs from other speakers, her data was kept separate. The number of tokens for each vowel is: [a] = 38, [e] = 38, [i] = 35, [o] = 39, [ɯ] = 41 and V = 59.

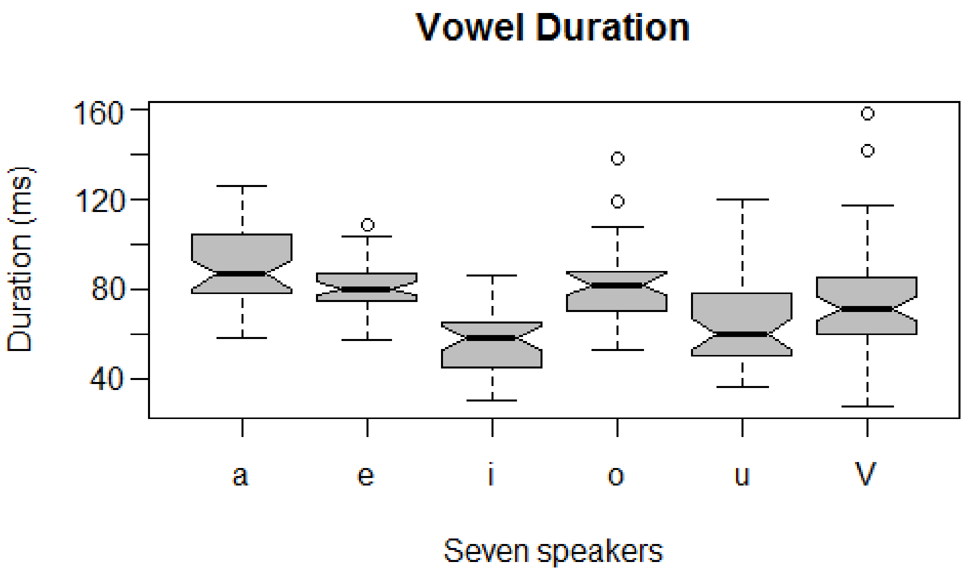

Figure 4.26. Boxplots of six vowels in the palatal context from seven speakers.

The contextually appropriate epenthetic vowel is the high front vowel [i] whose mean duration is 56.77 ms (median = 58 ms), which is the shortest vowel in this context. The second shortest vowel is the [ɯ] vowel whose mean duration is 64.17 ms (median = 60.0 ms), while the mean duration of the epenthetic vowel is 74.11 ms (median = 71 ms). An ANOVA showed an effect of vowel [*F* (5, 244) = 17.22, $p < .001$]. A Tukey post-hoc test showed that there was no significant difference between [ɯ] and [i] ($p = .497$). There was, however, a significant difference between [i] and V ($p < .001$), but there was no significant difference between [ɯ] and V ($p = .085$). For four speakers (F5, F7, M2, M4), the lexical high front vowel [i] is the shortest vowel among Japanese five vowels. For three other speakers (F4, M1, and M3), the [ɯ] vowel was the shortest vowel. However, of all the speakers only F7 produces an epenthetic vowel that corresponds in quality to her shortest vowel (i.e., [i]).

In terms of vowel duration, speaker F9 behaves differently from the others, as shown in Figure 4.27. The mean duration of the lexical vowel [i] is 110 ms (median = 113 ms), which is almost twice as long as the mean duration of the vowels of the other seven speakers. Her shortest vowel in this context is the lexical vowel [ɯ] (mean = 108.5 ms, median =110.5 ms). She produced the epenthetic vowel with a longer duration (mean = 121.25, median = 122.5) than any other speaker. Her epenthetic vowel is most similar in duration to her tokens of [a] and [o], though recall that the quality of her epenthetic vowels included all vowels except [o]. As with the other preceding contexts, since sample size is less than 10 tokens for each vowel, statistical analysis was not conducted for speaker F9.

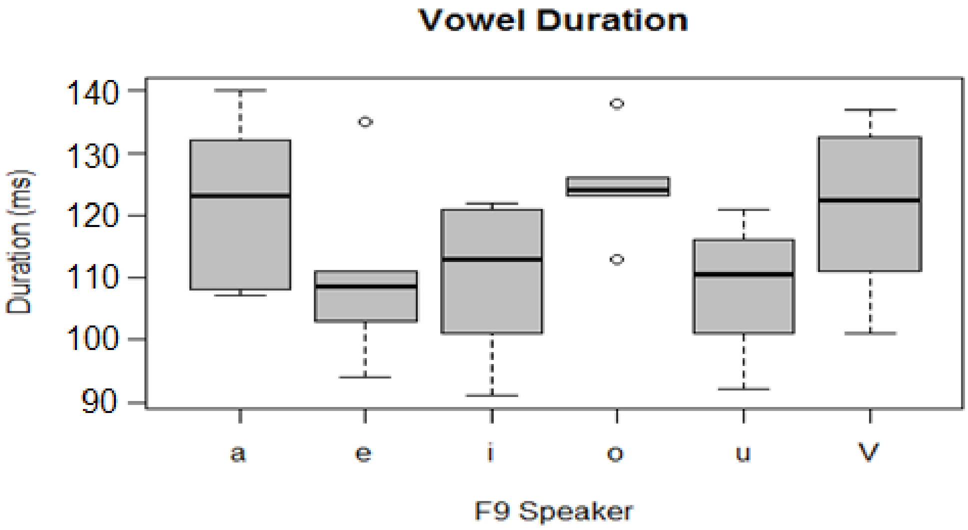

Figure 4.27. Boxplots of six vowels for speaker F9 in the palatal context.

4.3.6 Summary

In this chapter, the effect of the preceding consonant and vowel duration on the choice of epenthetic vowel produced was investigated. Consistent with loanword phonology, the epenthetic vowel [ɯ] was observed in both labial and velar contexts. However, in alveolar and palatal contexts, there is variability across speakers as to which vowels are epenthesized. While the predicted epenthetic vowel [o] in the alveolar context was used by some speakers, the epenthetic vowel [ɯ] in Japanese was used by other speakers. In the palatal context, the data show that only one speaker utilised the predicted epenthetic vowel [i]. Furthermore, the vowel duration study has shown that inserted vowels do not consistently correspond to the shortest vowel in a given context.

4.4 Discussion

The aim of the current study was to determine to what extent the production of an unfamiliar medial consonant cluster, [aCCa], is constrained by a speaker's native phonological patterns and especially, whether the preceding consonants influence the quality of epenthetic vowels. First, it was observed that the insertion of epenthetic vowels in consonant clusters is influenced by the speakers’ native phonological knowledge, as expected from previous research. The production of the consonant clusters consistently yielded epenthetic vowels when [aCCa] pseudo-word stimuli were presented (265 out of 288 tokens).

The results suggest that Japanese phonotactics influence the production of non-native CC clusters.

Second, the results showed that the quality of epenthetic vowels was influenced by preceding consonantal environment consistent, at least to some extent, with Japanese phonotactics. After labial and velar consonants, the present study found that the quality of the epenthetic vowel produced by native speakers of Japanese showed similar patterns to those found in studies of Japanese loanword adaptation. That is, speakers used the epenthetic vowel [ɯ] in these contexts. These results are consistent with the study of Yazawa et al. (2015) where it is claimed that, irrespective of a learner's proficiency level, the quality of epenthetic vowels is predicted to be similar to patterns found in loanword phonology. As expected, their study also shows that an epenthetic vowel has a quality close to [o] after [t, d], and [i] after [tʃ, dʒ]. However, the findings of the present study differ in these contexts. In the current study, we observed that some speakers utilise the default epenthetic vowel [ɯ] after the coronal stop [d] instead of the expected epenthetic vowel [o], although some speakers did use the expected [o]. Even more surprisingly, there was a great degree of variability among individuals in the type of vowel that was inserted after the palatal consonant.

As noted just above, three speakers utilised the epenthetic vowel [ɯ] in Japanese after the alveolar stop [d] instead of the expected vowel [o] (see § 4.3.4). It seems reasonable to utilise the vowel [ɯ] after [t] and [d] given the vowel's phonetic nature. Kobayashi (2000) states that [dɯ] and [tɯ] are perceptually similar to the original sound [t] and [d]. In addition, he points out that the vowel [ɯ] is the most neutral vowel sound in the Japanese vowel inventory since the tongue and lips move less than the other four vowels in the production of speech sounds. In fact, one male speaker (M2) seems to use the default epenthetic vowel [ɯ] in Japanese regardless of the preceding consonantal environment. Another potential explanation may be that since the stimuli were nonsense words, subjects considered them to be foreign and thus not subject to the phonotactic sequencing constraints of their own language. In fact, Hall (2009) found seven instances of loanwords with [dɯ] in the NTT lexicon of Japanese.

Let us now turn to the palatal context, where it was found that there was a great degree of variation among individuals in the present results (see § 4.3.5). Recall that the results from the current production experiment differed from Yazawa et al. (2015) where the expected vowel [i] followed the palatal consonant. One plausible reason for this discrepancy may be that Yazawa et al. (2015) used a text reading task in which participants read the Aesop fable “The North Wind and the Sun”, whereas pseudo-words were used in the current study. It may be that the participants in the reading task knew how the English words in the book should be pronounced. In addition, there were only two words “which” and “obliged” which appeared with palato-alveolar affricates [tʃ dʒ]. Nonetheless, the results from Yazawa et al. (2015) are consistent with loanword studies.

The individual variation among the speakers in the current study is more difficult to explain. Based on the argument from above that the high vowel [ɯ] was used after the coronal stop [d] on the basis of auditory or articulatory similarity, we might then expect the high front vowel [i] to be inserted after the alveolo-palatal [dʑ]. The high vowel [i] is not only similar in terms of the preceding consonant's place of articulation (Ladefoged & Maddieson, 1996), it is also the phonetically shortest in this context (see § 4.3.3), giving it low salience in this context. Despite the presence of this motivation for [i] epenthesis, only one speaker (F7) constantly produced the high vowel [i]. In contrast, another speaker (F5) never used the high front vowel [i] and another speaker (M2) used only [ɯ]. For the other five speakers, the choice of vowel appears to be random with the exception of the mid back vowel [o] which was never used after the palatal. Some potential reasons for the variability are considered.

One reason for the insertion of unexpected vowels may be due to the order in which words were presented by E-prime to speakers. It may be that choice of epenthetic vowel was influenced by the vowel in the preceding stimuli word. Eleven samples of 67 vowel productions across speakers had this pattern (i.e., [a] =3, [e] =2, [o] =3, [ɯ] =3). A typical example is speaker, F5. She produced vowels that were the same vowel in the preceding stimuli word 5 out of 9 times. For example, she inserted [a] and [e] once each in the palatal context; [e] occurred when the previous stimulus was [adʑeda] whereas the low vowel [a] was inserted when the previous stimulus was [adaba]. She inserted [o] four times in [adʑCa]; two of them occurred when the previous stimulus was [adʑoga] and [adʑoda], respectively. However, the other one occurred in a [ɯ] context.

Another potential explanation may relate to frequency since it has been argued that the epenthetic vowel in a language is likely to be a high frequency (low information) vowel (e.g., Eddington 2001; Hume & Bromberg 2005; Hume et al. 2013). Here, the frequency of

CV sequences in Japanese is considered in order to find out whether the frequency of the [dʑV] sequences affects on the choice of the epenthetic vowels in this context. Token and type frequency of [dʑa], [dʑe], [dʑi], [dʑo] and [dʑɯ] were examined. The data was extracted from the Corpus of Spontaneous Japanese (CSJ)[7]. The segment labelling used for the palatal affricates in the CSJ is ‘zya’, ‘zye’, ‘zj’, ‘zyo’, and ‘zyu’ and these vowels include not only short vowels but also long vowels. As can be seen from Table 4.4, if the frequency of CV plays a crucial role in the choice of epenthetic vowel in this context, the high vowel [i] would again be the best candidate, in line with the predictions of both phonotactics and acoustic similarity. The sequences with [ɯ] and [o] are the next most frequent, but this does not correspond to epenthesis patterns either; the former is used by some speakers while [o] never is. From this we can conclude that frequency involving the palatal consonant does not seem to play a role.

Table 4.4

[dʑV] Frequency from the CSJ

	[dʑa]/[dʑa:]	[dʑe]/[dʑe:]	[dʑi]/[dʑi:]	[dʑo]/[dʑo:]	[dʑɯ]/[dʑɯ:]
Token	793	146	6661	2740	2509
Type	73	44	944	321	329

Some speakers used the vowels [e] and/or [a] as epenthetic after [dʑ]. A potential explanation for insertion of the mid vowel [e] after [dʑ] may be due to the status of [dʑe] in Japanese. As can be seen in Table 4.5, the number of [dʑe, dʑe:] for both token frequency and type frequency is very low in comparison to the other vowels. This is because [dʑe] does not exist in traditional Japanese, and occurs only in loanwords (e.g. [dʑesɯtɕa] ‘gesture’, [dʑeɹato] ‘gelato’). Participants in this experiment knew the words were not Japanese and so may have been influenced by the presence of [e] after [dʑ] in other foreign words. On the other hand, insertion of [a] might be due to vowel harmony effects. Since the initial and final vowels of the pseudo-words were always [a], the choice of epenthetic vowel might have been influenced by neighbouring vowels. However, Uffmann (2006) states that the low vowel /a/ is least favoured to trigger vowel harmony in the three languages he investigated (Sranan, Shona, and Samoa).

---

[7] See Maekawa, Koiso, Furui & Isahara, 2000; Maekawa, 2015 for details of the CSJ.

The other possibility for why [e] might be inserted is that speakers are inserting a sound which is weakly noticeable phonetically. The perceptual identification study in Mattingley (2014) found that [e] was by far most frequently confused with [i]. This result shows that Japanese listeners have difficulty discriminating [e] from [i] in speech perception, suggesting perceptual similarity between these two vowels. Therefore, native speakers of Japanese have difficulty perceiving the contrast. The same pattern might be said to apply in varying degrees to epenthesis in production; the epenthetic sound is possibly close to [e] or [i] which varies across speakers. Speakers may be intending to insert [i], but it is not fully articulated (e.g., Lindblom, 1963). It can therefore be assumed that candidates for epenthetic vowels are sounds which are phonetically not noticeable according to context (cf. Fleischhacker, 2001; Steriade 2001), not necessarily the shortest vowel in the language (see Chapter 5 for related discussion). All of these potential explanations are speculative at this point and I leave them open for further testing and consideration.

Chapter 5
General Discussion and Conclusion

5.1 Summary of Findings

This thesis has presented the results of perception and production experiments to investigate the quality of epenthetic vowel that native speakers of Japanese tend to produce and perceive between unfamiliar sequences of consonants. The current study extended previous research on perceptual epenthesis in Japanese by taking into account a broader range of potential vowel percepts, [a, e, i, o, ɯ], as well as the preceding consonantal context, using two measures: accuracy and reaction time. In addition, this thesis investigated to what extent the preceding consonants influence the quality of epenthetic vowels in speech production. Most studies in epenthesis have carried out either a perception or production experiment, but have not looked at the phenomenon from both perspectives.

In this study, it was hypothesized that epenthesis in perception and production would be constrained by native phonology and, as a result, the quality of epenthetic vowels would be influenced by different preceding consonantal contexts. The findings of the present study are partially consistent with this hypothesis, but show important differences with previous studies. The following results were revealed in the current study. First, epenthetic vowels were observed in both speech perception and production. Second, the results show that to some extent the preceding consonant does influence the vowel perceived and produced yet, at the same time, there is a bias toward perceiving and producing [ɯ] in contexts not predicted by the language's phonotactic patterns. Third, there is cross-speaker variability as well as within-speaker variability in the palatal context in production.

In order to determine whether a consonant's place of articulation influences epenthesis in the same way in perception and production, the findings of this thesis are summarised in Table 5.1.

Table 5.1

Summary of Findings

Preceding Consonant	Vowel Predicted	Voicing Context	Findings		
			AX Accuracy	AX RT	Production
Labial	[ɯ]	Voiced	more errors with [ɯ]	significantly slower with [ɯ]	mostly [ɯ]
		Voiceless	more errors with [ɯ]	significantly slower with [ɯ]	N/A[8]
Velar	[ɯ]	Voiced	more errors with [ɯ]	significantly slower with [ɯ]	mostly [ɯ]
		Voiceless	more errors with [ɯ]	significantly slower with [ɯ]	N/A
Alveolar	[o]	Voiced	most errors with [o]	significantly slower with [i]	[o] and [ɯ]
		Voiceless	more errors with [i] and [ɯ]	significantly slower with [ɯ]	N/A
Palatal	[i]	Voiced	more errors with [i]	significantly slower with [i]	variability among speakers
		Voiceless	more errors with [i] and [ɯ]	significantly slower with [i] and [ɯ]	N/A

As can be seen in Table 5.1, for the labial and velar contexts, this study showed that the quality of epenthetic vowels is consistent with our hypothesis based on previous literature (e.g., Irwin, 2011; Shoji & Shoji; 2014): for both perception and production, the epenthetic vowel was predominantly the default vowel [ɯ] irrespective of the voicing of preceding consonants. The results also are consistent with the findings of Dupoux et al. (1999) and Dupoux et al. (2011) who show that native Japanese listeners tend to perceive an "illusory" vowel [ɯ] in an illicit word-medial consonant sequence. Additionally, reaction time data revealed that even though the listeners discriminate contrasting pairs correctly, significantly

---

[8] Note that in cells for voiceless production, there is an N/A due to the fact that only epenthetic vowels in the voiced context were analysed in this thesis.

slower reaction times were found with the appropriate epenthetic context for [ɯ], in comparison to other vowels. These findings suggest that [ɯ] is the phonetically minimal vowel and most closely resembles the absence of a vowel in these contexts.

While the results for the velar and labial contexts closely mirrored our expectations, the alveolar and palatal results were more complex. In the preceding alveolar context, table 5.1 shows that three different epenthetic vowels [o, i, ɯ] are observed across the three different experimental measurements. Japanese listeners showed difficulty discriminating contrasting pairs when [o] was in the stimuli [adVCa] compared to other vowels, although their accuracy score in this context was still very high (<AB>: 91%, <BA>: 98%). When the preceding alveolar was voiceless, discrimination accuracy was slightly lower when the medial vowel was [ɯ] (<AB>: 81%, <BA> 89%) or [i] (<AB>: 89%, <BA>: 88%) than when the medial vowel was [o] (<AB>: 95%, <BA>: 94%). These results suggest that Japanese listeners successfully discriminate contrasting pairs when the medial vowel is [o] in the voiceless context, consistent with the findings of Monahan et al. (2009). In terms of reaction time, in the voiced alveolar context, participants took significantly longer to discriminate pairs with [i] than those with [o] (i.e., the alveolar expected pair [adoCa_adCa]). For the voiceless context, pairs with [ɯ] took significantly longer to discriminate than contrasting pairs in which the medial vowel was [o]. These discrimination results suggest that [o] is not perceptually minimal in the alveolar context. For epenthesis in production, we observed that some speakers utilise the default epenthetic vowel [ɯ] after the coronal stop [d] instead of the contextually appropriate epenthetic vowel [o]. The results suggest that in both perception and production the vowel [ɯ] is expanding beyond what is predicted by the language’s phonotactic patterns.

In the palatal context, consistent with the previous literature, [i] was predominantly perceived as the epenthetic vowel. In the voiceless context, however, discrimination was poorer in pairs when the medial vowel was [ɯ] as well as [i], which is not predicted by the patterns of loanword adaptation, but again suggests expansion of [ɯ] as the epenthetic vowel. In addition, the reaction time for the [tɕɯC-tɕC] pair is not significantly different from that of the [tɕiC-tɕC] pair. Again, we see the vowel [ɯ] expanding into contexts not previously identified based on previous studies. In the speech production results, there is considerable variability across speakers as to which vowels was epenthesized, which suggests that the impact of the preceding consonantal context is weak and that other factors (e.g., phonetics) influence the choice of epenthetic vowels.

In addition, it was found that the order that the stimuli were presented to subjects influences epenthesis in perception. Japanese listeners were less accurate in identifying whether members of a pair were same-different with the [aCVCa-aCCa] order than with the [aCCa-aCVCa] order. This result is consistent with Davidson (2011) who claims that the order of presentation has an effect on perceptual epenthesis. Davidson found that Catalan and English listeners were less accurate on ‘native/non-native order’ than on ‘non-native/native order’ in an AX discrimination task. As Davidson speculates, native-sequences might have hindered listeners in perceiving non-native sequences because the non-native sequences are treated as a variant of the possible native word.

5.2 General Discussion

Results from the perception and production experiments suggest that the influence of the preceding consonant on the quality of epenthetic vowel is not uniform across contexts and experimental methodologies. Taking Japanese phonotactics into account, we would expect the quality of epenthetic vowel to be more constrained by a preceding alveolar than other preceding consonants. Since [t d] consonants are realised as different allophones before high vowels [ɯ] or [i] in native Japanese, the mid back vowel [o] is typically inserted after the alveolar stops. In the voiced alveolar context, even though the listeners discriminated the contrasting pairs correctly, most errors occurred with the predicted epenthetic [o], as expected. However, in the voiceless alveolar context, native speakers of Japanese exhibited greater difficulty in discriminating contrasting pairs in which the medial vowel was [ɯ] than those with [o]. Difficulty discriminating between [aCɯCa] and [aCCa] was also observed in the palatal context. This variability is not surprising perceptually given the phonetic nature of the high vowel [ɯ]; it is the phonetically weakest vowel in Japanese under the assumption that shorter duration, vowel devoicing and unaccentedness correlate with weaker salience. Listeners are thus associating the illusory vowel with the least salient vowel (Shoji & Shoji, 2014; Steriade, 2001). This account is consistent with the view that perceptual adaptations to non-native speech are phonetically minimal (Peperkamp & Dupoux, 2003; Peperkamp, 2005).

Another finding may support this view as well. When we focus on reaction time results, Table 5.1 shows that [ɯ] and [i] had longer reaction times than other vowels. However, it should be noted that while the devoiced [ɯ] seems to impact discrimination accuracy rates to some extent in the alveolar and palatal contexts as well as the labial and velar contexts, the devoiced [i] did not appear to be difficult to discriminate in the labial and velar contexts. That is, listeners correctly discriminated between licit [aCiCa] and illicit [aCCa] pairs in the voiceless consonantal contexts (i.e., labial: <AB> 95%, <BA>100%; velar <AB> 98%, <BA> 100%), despite [i] being weakly sonorous (Carr, 1999; Ladefoged & Keith, 2015). The implication is that devoiced vowels are not always illusorily epenthesized; the perceptually minimal vowel varies depending on the preceding consonant and order of presentation. This supports the view that an adequate account of perceptual epenthesis requires reference to the interaction of phonological, phonetic, as well as potentially other factors (Davidson & Shaw, 2012; Hume et al., 2013; Monahan et al., 2009).

In terms of production, the insertion of [ɯ] in unexpected contexts also makes sense from an articulatory perspective. Based on the assumption that speakers attempt to be faithful in the production of pseudo-word stimuli, speech sounds are articulated with minimal differences in tongue and lip movements from that in the stimuli. Kobayashi (2000) states that the vowel [ɯ] is the most neutral vowel sound in the Japanese vowel inventory since the tongue and lips move less than for the other four vowels. From an articulatory perspective, this segment minimizes the transition from the first consonant to the next segment, resulting in a minimal modification between the visual representation and phonetic production. Producing [ɯ] results in maximum auditory similarity between the input and output (e.g., Fleischhacker, 2001). That is, the input [adba], for example, is more similar to the output [adɯba] than [adoba]. Thus, epenthetic [ɯ] would be expected from both articulatory and perceptual perspectives of non-native sound adaptation, as it satisfies the condition that the epenthetic vowel should involve a phonetically minimal change between input and output (Steriade, 2001). It seems possible that what is minimal comes from both phonological and phonetic information used during speech production. We speculate that some participants are more likely to rely on native phonological information (phonotactics) to produce epenthetic vowels, and others rely more on phonetic detail. In the alveolar context, the former people utilised the mid back [o] as the epenthetic vowel, consistent with Japanese phonotactics. However, other participants (i.e., those who used [ɯ]) had a tendency to rely on phonetic information to produce epenthetic vowels since they used the least salient vowel more. Hence different available information may lead to different choices of epenthetic vowel.

However, it is still difficult to explain the results showing a huge discrepancy in the quality of epenthetic vowels between perception and production in the palatal context. Consistent with loanword studies in Japanese, in the perception experiments, [i] was predominantly perceived as the epenthetic vowel. On the other hand, in the speech production task, there is great variability across speakers as to which vowels they epenthesize. The high vowel [i] is similar in terms of the preceding consonant's place of articulation and is the phonetically shortest in this context. Despite the presence of this motivation for [i] epenthesis, only one person consistently used the contextually appropriate epenthetic vowel [i] in this condition. Thus, the variability found in the production study is difficult to explain from a phonetic perspective, since some people even used epenthetic [e] and [a]. Many studies on epenthesis in production found that several factors influence patterns of epenthesis (e.g., Carlisle, 1999; Eckaman & Iverson, 1993; Davidson, 2006, 2011; Lin, 2003). One potential explanation for the great variability is related to the location of the epenthetic vowel in the word. Epenthesis in a writing production study by Shoji & Shoji (2014) demonstrates that in word-initial consonant clusters, the expected vowel [i] was selected after the palatal context only 34.4 % of the time and all other vowels [a,e,o,ɯ] were chosen to be epenthetic vowels at least some of the time. However, in the word-final coda condition, [i] was epenthesized 85.6% of the time. This pattern is seen with the other preceding contexts as well. That is, there is greater variety in the selection of epenthetic vowels in word-initial consonant clusters than in word-final codas. Thus, the expected illusory vowel [i] might have been seen in the palatal context in the current study if the palatal had been the word-final coda. Finally, individual variation in speech production might be also attributed in part to extralinguistic factors such as age, dialect, and educational background, a topic to be explored in future studies.

The current series of studies revealed that the quality of epenthetic vowels was not merely influenced by the phonotactics of the native language in speech perception and production. Although three different vowels [i,o,ɯ] were inserted depending on the quality of the preceding consonants, the quality of epenthetic vowels is not always systematically derived by the preceding consonantal context in practice. This suggests that the native speakers in the current study did not rely solely on phonotactic knowledge when modifying non-native consonant clusters. Instead, other factors interact in a complex way during speech perception and production. It could be said that the quality of epenthetic vowels is intricately intertwined with several variables: the preceding consonant, its voicing type, the stimuli order and possibly other factors. In general, however, the quality of epenthetic vowel is most likely to be the perceptually minimal sound in a given context.

5.3 Limitations and Future Directions

The studies presented in this thesis have a number of limitations that have methodological implications for related future work. First, since all participants in the current study were recruited under the condition that they had been in New Zealand for less than two years, the number of participants was small. Therefore, the participants are not evenly distributed across social factors: gender, age, and educational background. Further research will be benefit from including sociolinguistic factors in order to determine to what extent they might be responsible for variation among individuals in the current study.

Second, some speakers showed difficulties reading pseudo-words, both control and target words, in the production task. This resulted in a limited number of tokens of the lexical vowels [i] and [e] after [ɡ], but also led to the need to discard some target words from analysis. It could be useful to obtain a set of real words for each control vowel or explore other methods of data collection.

Third, epenthetic vowels in voiceless consonant contexts underwent devoicing regardless of speakers’ dialect. Although vowel devoicing for Tokyo dialect speakers was expected, it was not expected that Japanese speakers from western Japan would produce devoiced high vowels. Therefore the second consonant in CC clusters is an important element to consider in future studies, in order to avoid devoicing contexts.

Lastly, this study did not compare individual perception and production behaviour. That is, it did not examine epenthesis patterns in speech perception and production for the same speakers. If the current study could investigate whether the quality of epenthetic vowels in production resembles epenthesis patterns in perception across individuals, it would help clarify the role of predictive factors. Future studies addressing individual variability between epenthesis in perception and production may provide further insight into predicting the quality of epenthetic vowels.

5.4 Conclusion

This study carried out perceptual and production experiments in an investigation of the contextual environments that contribute to predicting the quality of epenthetic vowels in Japanese. Consistent with the language's phonotactic patterns, the preceding consonant is shown to influence the perception and production of an epenthetic vowel, though not in all cases. Contrary to native phonotactics, an arguably low-salience vowel is perceived as epenthetic in an otherwise illicit consonant sequence. Production experiments suggest that the quality of epenthetic vowels is not predicted by only phonological and phonetic influences. Rather, there are several factors that can influence the quality of epenthetic vowels during adaptation of unfamiliar consonant sequences.

Appendices

Appendix A: List of Dialects

Location	Region	Dialect	Spoken area (prefecture, city)	Number of speakers *
North	Tohoku Region	Tohoku dialect	Iwate	1
	Hokuriku Region	Niigata dialect	Niigata	1
↑	Kanto region	Tokyo dialect (Standard Japaese)	Tokyo, Chiba, Kanagawa	13
		Tochigi dialect	Tochigi	1
Mid	Tokai Region	Nagoya dialect	Nagoya	3
	Kansai Region	Kansai dialect	Osaka	4
		Sensyu dialect	South-West Osaka	1
		Ako dialect	Hyogo	1
↓	Kyusyu Region	Fukuoka dialect	Fukuoka	1
		Saga dialect	Saga	1
South	Okinawa Region	Okinawa dialect	Okinawa	1

*multiple answers allowed

Appendix B: Session Schedule

Session A		Session B	
Section 1	AX Discrimination Task	Section 1	AX Discrimination Task
	Instruction		Instruction
	Practice		Practice
	Question & Answer Session		Question & Answer Session
	Block 1: Voiced Consonantal Stimuli List 1		Block 1: Voiceless Consonantal Stimuli List 2
	Break 1		Break 1
	Block 2: Voiceless Consonantal Stimuli List 1		Block 2: Voiced Consonantal Stimuli List 2
	Break 2		Break 2
Section 2	Production Task	Section 2	Production Task
	Instruction		Instruction
	Practice		Practice
	Question & Answer Session		Question & Answer Session
	Block 1: Voiced Consonantal Stimuli List 1		Block 1: Voiceless Consonantal Stimuli List 2
	Break 3		Break 3
	Block 2: Voiceless Consonantal Stimuli List 1		Block 2: Voiced Consonantal Stimuli List 2

Appendix C: Background Questionnaire (original)

Post-Experiment Questionnaire　質問表

A. General Information 一般事項

1. Date 日付 ________________________________

2. Date of Birth 年齢 ________________________________

3. Gender 性別 ________________________________

4. Dominant hand 利き手 ________________________________

5. Do you have any speech or hearing disorders? (If yes, please describe)
発話・聴覚に支障がありますか？ある場合は記載してください。

B. Known Languages and Uses　知っている言語と使用について

1. Where did you grow up? (E.g. Japan: Osaka)
どこの国、または地域で育ちましたか？（例：日本：大阪）

2. Do you speak any dialect of Japanese? (E.g. Tokyo, Kawachi, Hakata dialect, etc.)
日本語のどちらの方言を話すか教えてください。（例：東京、河内、博多弁等）

3. What other languages do you speak, and how many years have you spoken the languages, if any?日本語以外で話すことができる言語があれば, その言語を教えてください。 何年ぐらいその言語を使用していますか？

Language:	*How many years have you spoken the language?*	*Do you speak the language fluently? Yes/ No*

4. How long have you lived in an English speaking country?
英語圏に住んだことがある方は、滞在期間を教えてください。

Appendix D: List of Items for Production Experiment

	Control					Target	Fillers	
	aCaCa	**aCeCa**	**aCiCa**	**aCoCa**	**aCuCa**	**aCCa**	**eCCa** **iCCa** **oCCa** **uCCa**	
Voiced (Control=60, Target=12, Fillers=24)	abada	abeda	abida	aboda	abuda	abda	ebga	edga
	abaga	abega	abiga	aboga	abuga	abga	edma	egja
	abaja	abeja	abija	aboja	abuja	abja	ejba	ejna
	adaba	adeba	adiba	adoba	aduba	adba	ibda	ibja
	adaga	adega	adiga	adoga	aduga	adga	idba	igba
	adaja	adeja	adija	adoja	aduja	adja	igma	ijna
	agaba	ageba	agiba	agoba	aguba	agba	ubma	udna
	agada	ageda	agida	agoda	aguda	agda	udga	ugda
	agaja	ageja	agija	agoja	aguja	agja	ujga	ujma
	ajaba	ajeba	ajiba	ajoba	ajuba	ajba	obna	obga
	ajada	ajeda	ajida	ajoda	ajuda	ajda	odba	ogja
	ajaga	ajega	ajiga	ajoga	ajuga	ajga	ogma	ojda
Voiceless (Control=60, Target=12, Fillers=24)	apata	apeta	apita	apota	aputa	apta	epka	etma
	apaka	apeka	apika	apoka	apuka	apka	ekma	ekta
	apacha	apecha	apicha	apocha	apucha	apcha	echna	echpa
	atapa	atepa	atipa	atopa	atupa	atpa	ipta	ipcha
	ataka	ateka	atika	atoka	atuka	atka	itpa	ikpa
	atacha	atecha	aticha	atocha	atucha	atcha	ichna	ichta
	akapa	akepa	akipa	akopa	akupa	akpa	upka	utna
	akata	aketa	akita	akota	akuta	akta	utpa	ukta
	akacha	akecha	akicha	akocha	akucha	akcha	ukcha	uchka
	achapa	achepa	achipa	achopa	achupa	achpa	opna	opka
	achata	acheta	achita	achota	achuta	achta	okcha	otma
	achaka	acheka	achika	achoka	achuka	achka	otpa	ochta

Appendix E: Background Information of Speakers in Production

Speaker	Age	Region Speaker is From	Dialect	Overseas Experiences	Self-Assessment English Fluency 'yes' or 'no'
F4	22	Aichi	Standard Japanese Nagoya dialect Osaka dialect	3 months (NZ)	yes
F5	21	Niigata	Niigata dialect Standard Japanese	2 months (NZ)	no
F7	46	Fukuoka	Hakata dialect Standard Japanese	1 month (NZ)	no
F9	30	Osaka	Senshyu dialect	3 months (NZ)	yes
M1	22	Saga	Saga dialect	6 months (NZ)	yes
M2	40	Tokyo	Standard Japanese	10 months (NZ) 2 months (US)	yes
M3	24	Tochigi	Tochigi dialect Okinawa dialect	2 weeks (NZ)	no
M4	28	Tokyo	Standard Japanese	6 months (NZ)	yes

Appendix F: Mean Acoustic Measurements across the Eight Speakers

				Duration	F1		F2		F3	
PreC	Speaker	# of tokens	Vowel	mean	mean	sd	mean	sd	mean	sd
[b]	F4	6	a	111	747	37	1314	66	3345	98
		5	e	101	516	58	2467	42	3296	49
		5	i	91	348	75	2919	80	3535	227
		5	o	109	440	18	878	71	3115	327
		6	u	83	414	37	1271	185	3055	323
		9	V	77	414	30	1311	177	3200	161
	F5	6	a	106	740	53	1322	93	2945	120
		6	e	83	490	27	2159	68	2946	202
		6	i	70	332	26	2340	18	3315	85
		5	o	81	499	14	991	104	2966	64
		5	u	63	427	25	1599	126	2820	157
		6	V	78	448	14	1485	114	2713	90
	F7	6	a	88	777	75	1761	113	3164	149
		6	e	100	577	58	2538	51	3296	43
		5	i	78	370	34	2906	64	3634	82
		6	o	84	553	54	1192	175	3349	48
		5	u	66	391	23	1561	238	3180	78
		7	V	68	403	42	1611	83	3146	75
	F9	5	a	120	783	24	1383	50	3237	89
		5	e	121	496	15	2336	215	3081	154
		5	i	103	412	20	3025	121	3363	130
		6	o	123	476	19	888	56	3526	110
		5	u	109	442	12	1405	119	3161	84
		6	V	132	420	27	2170	895	3223	136
	M1	6	a	102	581	46	1067	76	2760	80
		6	e	90	420	37	1845	173	2552	67
		6	i	84	316	17	2271	244	2624	104
		5	o	90	421	35	840	70	3034	105
		6	u	83	341	9	1257	112	2602	189
		8	V	71	346	11	1286	116	2558	67
	M2	6	a	73	727	39	1264	66	2628	35
		6	e	72	475	43	1951	64	2581	125
		6	i	53	283	16	2131	29	3234	76
		5	o	62	517	13	1000	266	2643	210
		6	u	51	427	21	1402	191	2468	69
		9	V	52	428	22	1415	146	2473	68
	M3	6	a	108	670	34	1169	44	2100	112
		6	e	107	501	17	1765	55	2509	220
		6	i	90	293	28	2050	169	3065	118
		4	o	103	495	31	794	102	2675	358
		5	u	68	410	16	1280	85	2254	62
		7	V	77	419	119	1295	189	2497	476
	M4	4	a	90	622	61	1226	47	2351	62
		5	e	101	434	26	2049	108	2797	180
		6	i	102	307	24	2324	89	3256	239
		5	o	75	464	37	979	98	2371	152
		4	u	66	412	21	1292	84	2359	93
		9	V	72	380	33	1186	137	2459	81

					F1		F2		F3	
PreC	Speaker	# of tokens	Vowel	Duration	mean	sd	mean	sd	mean	sd
[g]	F4	6	a	87	767	43	1411	68	3380	150
		6	e	92	465	56	2608	60	3291	97
		4	i	68	361	64	3015	51	3559	164
		6	o	80	499	70	902	116	3268	182
		6	u	64	441	27	1390	204	3260	73
		8	V	58	421	45	1393	191	3260	125
	F5	6	a	89	741	33	1411	124	2783	129
		6	e	91	479	16	2261	97	2892	148
		5	i	54	326	18	2440	64	3317	173
		6	o	72	523	31	1009	153	2856	69
		6	u	58	413	31	1528	157	2706	104
		9	V	67	417	28	1548	109	2651	94
	F7	6	a	86	748	72	1810	106	3167	48
		0	e							
		2	i	64	316	106	2686	165	3487	169
		6	o	72	548	82	1143	197	3296	138
		6	u	79	372	36	1763	375	3133	63
		8	V	69	421	32	1743	204	3162	66
	F9	6	a	125	817	43	1480	53	3067	83
		3	e	134	451	26	2450	98	2969	103
		4	i	102	409	7	2993	63	3654	69
		6	o	121	451	13	892	70	3373	80
		5	u	112	423	10	1429	83	3116	90
		8	V	116	428	18	1681	500	3267	162
	M1	5	a	101	564	11	1130	46	2692	124
		5	e	96	397	33	2113	163	2560	75
		4	i	61	313	9	2401	192	2655	80
		6	o	92	417	15	903	80	2770	54
		6	u	78	344	14	1333	124	2490	96
		9	V	65	337	22	1382	157	2457	88
	M2	6	a	70	737	28	1369	138	2588	100
		3	e	73	475	37	1984	107	2565	31
		1	i	93	274	NA	2113	NA	3283	NA
		6	o	60	529	10	982	191	2765	216
		6	u	51	399	29	1513	280	2427	66
		9	V	50	408	27	1490	182	2418	50
	M3	5	a	86	654	13	1312	79	2100	78
		4	e	86	477	12	1894	29	2410	146
		4	i	67	290	5	1975	143	2931	170
		6	o	79	482	22	931	98	2625	358
		6	u	61	385	15	1614	171	2199	146
		9	V	65	383	13	1565	176	2198	107
	M4	5	a	90	657	25	1371	89	2333	104
		1	e	74	471	NA	2098	NA	2782	NA
		5	i	69	292	12	2340	54	3348	131
		6	o	83	469	33	996	129	2360	114
		5	u	63	406	31	1383	154	2478	93
		9	V	68	418	47	1492	177	2451	139

PreC	Speaker	# of tokens	Vowel	Duration	F1 mean	F1 sd	F2 mean	F2 sd	F3 mean	F3 sd
[d]	F4	5	a	97	766	57	1560	45	3480	106
		6	e	95	474	61	2488	73	3323	71
		5	i	81	355	40	2903	85	3460	124
		6	o	94	451	61	1047	141	3267	262
		4	u	67	431	42	1788	313	3270	37
		7	V	76	434	28	1688	91	3235	65
	F5	6	a	98	738	58	1453	109	2847	453
		6	e	74	467	52	2182	93	2987	136
		6	i	72	361	14	2391	42	3208	120
		6	o	84	510	20	1171	110	2983	98
		4	u	80	460	33	1568	161	2836	91
		9	V	74	480	64	1342	267	2902	171
	F7	5	a	96	776	143	1878	80	3180	173
		3	e	80	512	57	2450	134	3223	32
		5	i	79	332	36	2906	82	3425	42
		5	o	94	598	49	1351	130	3342	94
		4	u	81	371	37	1782	179	3116	40
		7	V	80	512	98	1356	236	3244	115
	F9	6	a	135	810	43	1529	65	3144	90
		4	e	123	502	27	2398	81	3034	315
		5	i	115	402	12	3026	175	3385	186
		4	o	120	468	17	1046	46	3614	293
		6	u	122	427	16	1721	30	3119	69
		8	V	120	434	15	1637	239	3225	189
	M1	4	a	114	573	25	1277	33	2965	173
		5	e	96	412	24	1909	29	2640	106
		6	i	81	324	17	2208	223	2615	97
		5	o	88	434	34	1086	44	2969	128
		5	u	79	341	9	1469	77	2616	99
		9	V	93	418	28	1031	68	2952	53
	M2	6	a	77	683	34	1514	78	2637	80
		5	e	86	457	31	1959	85	2554	49
		6	i	65	303	17	2162	50	3021	103
		6	o	66	483	35	1197	140	2676	25
		5	u	58	410	44	1568	128	2527	25
		9	V	59	405	39	1572	117	2532	66
	M3	5	a	98	632	19	1341	59	2476	676
		5	e	93	468	58	1776	135	2426	61
		4	i	74	337	36	1985	54	2855	284
		6	o	76	487	20	1092	31	3018	281
		5	u	68	388	35	1465	53	2350	115
		9	V	81	485	25	1099	32	2689	273
	M4	6	a	100	598	59	1380	114	2547	121
		6	e	85	415	51	2110	68	2959	84
		5	i	91	315	25	2342	79	3323	117
		6	o	77	480	25	1309	112	2558	86
		6	u	87	376	49	1850	233	2578	97
		9	V	76	467	53	1337	148	2640	204

					F1		F2		F3	
PreC	Speaker	# of tokens	Vowel	Duration	mean	sd	mean	sd	mean	sd
[dʑ]	F4	6	a	95	682	32	1799	83	3292	64
		6	e	85	464	44	2325	80	3348	86
		4	i	77	408	36	2484	53	3373	65
		6	o	102	462	40	1027	105	3081	162
		7	u	66	414	23	1671	311	2954	192
		8	V	74	413	42	2041	456	3110	297
	F5	6	a	92	678	25	1631	99	2806	273
		6	e	81	499	22	2039	137	2823	207
		6	i	56	372	37	2125	89	2881	142
		6	o	80	511	10	1199	153	2674	175
		6	u	78	401	18	1773	85	2766	126
		9	V	79	493	101	1566	278	2680	220
	F7	5	a	85	719	86	1905	52	3116	147
		6	e	85	569	44	2289	43	3168	46
		5	i	50	399	28	2429	47	3196	93
		5	o	80	587	62	1525	115	3267	108
		6	u	67	409	35	1879	151	2929	324
		8	V	56	403	47	2395	232	3155	133
	F9	6	a	122	772	32	1602	107	3058	138
		6	e	110	501	27	2272	175	3104	155
		6	i	110	423	10	2452	156	3221	167
		5	o	125	493	21	1126	74	3530	236
		6	u	109	434	17	1790	66	3251	103
		8	V	121	574	176	1908	371	3058	171
	M1	5	a	104	544	25	1422	40	2858	186
		4	e	78	413	15	1812	168	2699	81
		4	i	57	320	24	2055	159	2843	60
		5	o	82	418	25	1131	44	2824	168
		6	u	53	345	12	1750	168	2682	170
		8	V	91	424	77	1708	383	2716	272
	M2	6	a	83	615	47	1717	111	2599	145
		6	e	68	469	20	1932	74	2626	74
		6	i	54	336	26	2065	58	2803	29
		5	o	73	471	25	1306	125	2452	73
		6	u	60	384	11	1851	199	2515	119
		9	V	54	374	18	1823	173	2502	127
	M3	5	a	81	584	32	1467	25	2302	311
		4	e	91	480	26	1746	29	2650	251
		6	i	55	374	33	1923	209	3005	377
		6	o	77	491	19	1222	12	2488	347
		5	u	54	383	24	1626	45	2782	505
		7	V	88	512	102	1607	181	2534	317
	M4	5	a	93	565	41	1669	91	2666	221
		6	e	81	431	13	2064	105	3069	69
		4	i	54	309	13	2197	233	2984	269
		6	o	81	464	26	1377	174	2367	108
		6	u	75	378	36	1816	62	2447	115
		9	V	78	434	91	1844	192	2727	285

References

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

Babel, M., & Johnson, K. (2010). Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, *1*(1), 179-205.

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. In W.Strange (Ed.), *Speech perception and linguistic experience*, 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.

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

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (2008). The impact of allophony versus contrast on speech perception. *Contrasts in phonology: Theory, perception, acquisition*, 145-171.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Second Language Research*, *7*(1), 35-59.

Carr, P. (1999). *English phonetics and phonology: an introduction*. Malden, Mass: Blackwell Publishers.

Carlisle, R. S. (1991). The influence of environment on vowel epenthesis in Spanish/English interphonology. *Applied Linguistics*, *12*(1), 76-95.

Cutler, A., Otake, T., & McQueen, J. M. (2009). Vowel devoicing and the perception of spoken Japanese words. *The Journal of the Acoustical Society of America*, *125*(3), 1693-1703. doi: 10.1121/1.3075556

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

Davidson, L. (2011). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 270. doi: 10.1037/a0020988

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of Phonetics*, *40*(2), 234-248. doi: 10.1016/j.wocn.2011.11.005

Donselaar, W. van., Kuijpers, C., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of memory and language*, *41*(1), 59-77.

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.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes, 16*(5), 491-505. doi: 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.

Eckman, F., & Iverson, G. (1994). Pronunciation difficulties in ESL: Coda consonants in English interlanguage. In Yavaş, M. S. (Ed.), *First and second language phonology* (pp. 251-265). San Diego, CA: Singular.

Eddington, D. (2001). Spanish epenthesis: Formal and performance perspectives. *Studies in the Linguistic Science*s, *31*(2), 33-53.

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

Gerrits, E., & Schouten, B. (1998). Categorical perception of vowels. In *Proceedings of the 5th International Conference on Spoken Language Processing in Sydney, 6,* 2279-2283.

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

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.

Hallé, P. A., Segui, J., Frauenfelder, U., & 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.

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

Hardison, D. M., & Saigo, M. M. (2010). Development of perception of second language Japanese geminates: Role of duration, sonority, and segmentation strategy. *Applied Psycholinguistics*, *31*(01), 81-99.

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

Hume, E (2016). Phonological Markedness and its Relation to the Uncertainty of Word. *Phonological Studies 19*, 107-116.

Hume, E., & Bromberg, I. (2005). Predicting epenthesis: An Information-theoretic account. In *Proceeding 7th Annual Meeting of the French Network of Phonology*.

Hume, E., Hall, K. C., Wedel, A., Ussishkin, A., Adda-Dekker, M., & Gendrot, C. (2013). Anti-markedness patterns in French epenthesis: An information-theoretic approach. In *Proceedings of the 37 th Annual Meeting of the Berkeley Linguistics Society, 37*(1), 104-123.

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

Itô, J. (1989). A prosodic theory of epenthesis. *Natural Language & Linguistic Theory, 7*(2), 217-259.

Itô, J., & Mester, A. (2003). Japanese phonology. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp.817- 838). Oxford: Blackwell.

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.

Kaneko, E. (2006). Vowel selection in Japanese loanwords from English. In *Proceeding*s *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: 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.

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

Kawahara, S., Erickson, E., & Suemitsu, A. (submitted). A quantitative study of jaw opening: An EMA study of Japanese vowels.

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

Kobayashi, Y. (2000). A pronunciation of English loanwords in the Japanese language. In Hiroshima Associated Repository Portal, *9*, 59-87. Retrieved from http//harp.lib.hiroshima-u.ac.jp/handle/harp/7985

Kondo, M. (2005). Syllable structre and its acoustic effect on vowesl in devoicing. In J. van de Weijer, K. Nanjo, & T. Nshihara (Eds.), *Voicing in Japanese* (pp 229-245). Berlin: Mouton de Gruyter.

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.

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.

Kuijpers, C., Donselaar, W. van, & Cutler, A. (1996). Phonological variation: Epenthesis and deletion of schwa in Dutch. In *Proceeding the Fourth International Conference on Spoken Language Proceedings* (ICSLP). *1*, 149-152. Philadelpia, USA.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. Retrieved from https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world’s languages*. Oxford: Blackwell.

Ladefoged, P., & Johnson, K. (2015). *A course in phonetics*. USA: Cengage Learning.

Lin, Y. -H. (2003). Interphonology variability: Sociolinguistic factors affecting L2 simplification strategies. *Applied Linguistics*, *24*(4), 439-464.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *The journal of the Acoustical society of America*, *35*(11), 1773-1781.

Maekawa, K. (2015). Corpus-based phonetics. In H. Kubozono (Ed.), *Handbook of Japanese phonetics and phonology* (pp. 651-680). Berlin: Mouton de Gruyter.

Maekawa, K., Koiso, H., Furui, S., & Isahara, H. (2000). Spontaneous speech corpus of Japanese. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC), 2*, 947-952.

Mattingley, W (2014). *The Influence of Preceding Consonant on Perceptual Epenthesis in Japanese*. Unpublished manuscript.

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

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician, 32*(1), 12-16.

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* (Vol. 17, pp. 391-405). Stanford, CA: CSLI Publications.

Otaki, Y. (2012). A phonological account of vowel epenthesis in Japanese loanwords:Synchrinic and dyachronic perspectives. *Phonological studeis, 15,* 35-42.

Peperkamp, S. (2005). A psycholinguistic theory of loanword adaptations. In *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*. 341–352.

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.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*(2), 285-290.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*(3), 510-532

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

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.

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.

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).

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

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.

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: 10.1121/1.3651080

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

Tsuchida, A. (1997). *Phonetics and phonology of Japanese vowel devoicing*. Doctoral dissertation: University of Cornell.

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:10.1016/j.lingua.2005.06.009

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.

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

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.