Lexical Statistics Determine the Choice of Epenthetic Vowel in Japanese Loanword Adaptation

Carmen Wilson

Advisors: Ryan Bennett & Jason Shaw

Submitted to the faculty of the Department of Linguistics
in partial fulfillment of the requirements for the degree of
Bachelor of Arts

Yale University
28 April 2017

Abstract

Perceptual epenthesis is when listeners perceive a vowel between consonant sequences that are illicit in their native language. Psycholinguistic studies about perceptual epenthesis in Japanese have shown that the high back vowel [u] is commonly chosen to repair illicit consonant clusters. However, based on patterns of Japanese loanwords, more recent research has shown there can be more than one epenthetic vowel for Japanese. These studies have also provided convincing evidence that the quality of the epenthetic vowel can be modulated by other factors, such as native phonological processes. While progress has been made towards understand perceptual epenthesis, there are still mysteries surrounding the phenomenon and its relationship to loanword adaptation.

The purpose of this study is to investigate other factors that could possibly affect the choice of epenthetic vowel in Japanese loanword adaptation. Assuming the quality of the epenthetic vowel can be modulated by context, does that context-dependence reflect categorical phonological rules, statistical facts about the lexicon, or both? To answer this question, I use the Corpus of Spontaneous Japanese (CSJ) to quantify vowel distribution in the Japanese lexicon. I then conducted an online transliteration experiment in which native speakers were asked to listen to nonce words containing illicit consonant clusters and write them in Japanese orthography (*katakana*). Epenthetic vowels in observed responses were compared to epenthetic vowels predicted by patterns in CSJ. I also examine interactions of word position and C1 by comparing results from two sets of stimuli, one containing word-medial consonant clusters and another set consisting of the same clusters in a word-initial position. A correlation between observed and predicted responses reveals that as the probability of a vowel (V) increased in the lexicon following a particular consonant C1, Japanese speakers were more likely to choose V as the epenthetic vowel following C1. These results indicate that the choice of epenthetic vowel is sensitive to the identity of the preceding consonant, and listeners access and apply native lexical statistics granularly when adapting foreign words into their native language.

1 Introduction

Perceptual epenthesis is when listeners perceive a vowel between consonant sequences that are illicit in their native language. This effect was first observed for Japanese listeners who encountered clusters such as [ebzo] and perceived them with an illusory vowel as in [ebuzo] (Dupoux et al., 1999, see Section 3 for details). In subsequent papers, Dupoux et al. (2001 & 2011) have argued that [u] (phonetically [ɯ] but here represented as [u]) is the default epenthetic vowel perceived in illicit consonant clusters because [u] is the phonetically minimal vowel of Japanese (i.e. shortest and prone to devoicing). Other studies have found that Japanese speakers perceive [i] as the epenthetic vowel following palatal affricates (Mattingley et al., 2015) and some suggest that its appearance is motivated by a palatalization rule in native Japanese phonology wherein coronal consonants are palatalized before [i] (Durvasula & Kahng, 2015). For example, when the verb stem [hanas] is followed by [i], the result is [hanaɕi]. However, palatalization alone cannot drive the presence of epenthetic [i] for Japanese listeners because palatal affricates occur in contexts independent of [i] (i.e. [tɕa, tɕo, tɕu] are all licit sequences in Japanese); neither is Dupoux's theory of phonetic minimality able to account for the appearance of epenthetic [i] (See Section 7.3 for detail.) The question of what exactly modulates the epenthetic vowel is still open for debate. The current study aims to investigate a speculation in Mattingley et al. (2015): could a language's statistical patterns bias listeners to perceive certain vowels in certain contexts?

The rest of the paper is structured as follows: Section 2 provides background information on Japanese syllable structure; Section 3 reviews past psycholinguistic studies on epenthetic vowels; Section 4 introduces my proposal and hypotheses; Section 5 details the experiment; results are in

Section 6. This is followed by discussion, possible sources of error, and future research in Section 7.

2 Japanese Syllable Structure

Japanese syllable structure is almost exclusively [(C)V]. Onsets are optional, and the only permissible codas are nasal consonants or the first part of a geminate, as shown in (1) and (2) respectively. No other consonants may appear as codas, as shown in (3).

(1) a. [tom.bo] 'dragonfly'
b. [kaŋ.ga.e] 'thought'

(2) a. [kap.pa] 'a river sprite'
b. [gak.koo] 'school'

(3) a. *[kap.ta]
b. *[tog.ba]
c. *[pa.kat] (Itô 1989)

When words are borrowed from one language to another they may undergo adaptation processes to satisfy the structural constraints of the borrowing language phonology. When Japanese borrows words from languages that allow more complex syllable types (e.g. English) it must change the English syllable structure to obey native phonotactics while remaining as faithful as possible to the input form. Japanese simplifies consonant clusters and illicit codas with vowel epenthesis, as shown in (4).

(4) a. [fu.ru.ta.i.mu] 'full-time'
b. [su.fiŋ.ku.su] 'Sphinx'
c. [ku.ri.su.ma.su] 'Christmas'

While [u] is frequently chosen as the epenthetic vowel in Japanese loans, it is not the only one. [o] appears after coronal consonants [t,d] because *[tu] and *[du] are illicit sequences in native Japanese as seen in (5). If [u] is epenthesized following [t], it triggers frication, turning [t] into [tsu] as in [tsurii] from the English ‘tree.’ [i] appears after coronal affricates [ʧ, dʑ] as shown in (6).

(5)	a. ‘fight’	[fa.i.to]	*[fa.i.tu]
	b. ‘drive’	[do.ra.i.bu]	*[du.ra.i.bu]
	c. ‘strawberry’	[su.to.ro.be.rii]	*[su.tu.ro.be.rii]
	d. ‘McDonald’s’	[ma.ku.do.na.ru.do]	*[ma.ku.do.na.ru.du]

(6)	a. ‘catch’	[kʲat.ʧi]	*[kʲat.ʧu]	
	b. ‘pitch’	[pit.ʧi]	*[pit.ʧu]	
	c. ‘fudge’	[fad.dʑi]	*[fad.dʑu]	
	d. ‘range’	[ren.dʑi]	*[ren.dʑu]	(Monahan 2009)

Previous studies of Japanese epenthetic vowels have investigated consonants that would induce [u] epenthesis (e.g. [k, g, b]) and attributed perceptual epenthesis to phonotactic constraints and phonetic minimality. Others have investigated fricatives and affricates and proposed phonology as a key factor in choosing the epenthetic vowel. In this study I posit there is yet a third influencing factor on the choice of the epenthetic vowel— statistical patterns in the language.

3 Previous Studies on Perceptual Epenthesis

Dupoux et al. (1999) found that Japanese speakers reported hearing the illusory vowel [u] between two consonants in illicit contact (e.g. [bz] in [ebzo]) even when there was no vowel present in the acoustic signal. Starting from ten sequences of VCuCV nonce words, they created a continuum of stimuli containing the full [u] vowel to no vowel at all by splicing out pitch periods of the medial vowel [u] from the original nonce words. In an offline phoneme detection task, Japanese participants reported a vowel present in all levels of vowel length, while French speakers were able to judge that the vowel was absent in [ebzo] and present in [ebuzo]. In seperate ABX tasks, Japanese participants had difficulty discriminating [ebzo] from [ebuzo]. They concluded that the phonotactics of a listener's native language influence speech perception, and that Japanese listeners perceive the presence of an illusory [u] when they encounter an illicit consonant cluster.

In a later study, Dupoux et al. (2011) attribute the appearance of epenthetic vowels to an interaction between phonetics and phonotactic constraints. Specifically, they assert that the epenthetic vowel used to repair illegal clusters—in both perception and loanword adaptation—should be the phonetically minimal vowel of the language. This vowel "optimizes the product of the sequence probability on the one hand and the acoustic match between the segment and the relevant part of the signal (i.e., the transition between the consonants)"). This essentially means that the minimally phonetic vowel is the vowel that closest matches the phonetic information (i.e. spectral information, including formants) present in the transition part of the consonant cluster. To investigate which factors influence the quality of the epenthetic vowel, they ran several experiments with native speakers of Japanese, Brazilian Portuguese, and European Portuguese. In both Japanese and Brazilian Portuguese, the high vowels /i/ and /u/ become devoiced in certain contexts, and [u] is the shortest vowel in Japanese and [i] the shortest in Brazilian Portuguese. Because they are prone to deletion as well, these vowels, in certain contexts, are the closest to the absence of a vowel. Therefore Dupoux et al. predict that the epenthetic vowel should always be [u] for Japanese and [i] for Brazilian Portuguese. European Portuguese phonetically differs from Brazilian Portuguese because it optionally deletes unstressed vowels, so surface coda obstruents are common, and therefore clusters need not be repaired for European Portuguese speakers.

Experiment 1 tested the perception of illegal clusters in a vowel classification task, and Experiment 2 used an ABX discrimination task to test the perception of consonant clusters. Their findings mimicked the results of Dupoux et al. (1999) in which Japanese participants perceived the presence of [u] even when it was absent from the acoustic signal, and speakers of Brazilian Portuguese overwhelmingly perceived [i] even when it was absent from the acoustic signal. They propose a one-step model of speech perception in which segmental categorization and native language phonotactics interact to induce a vowel to repair illegal clusters. They assert this vowel (the "phonetically minimal" vowel) is always [i] for Brazilian Portuguese and [u] for Japanese.

These findings about perceptual epenthesis are closely related to arguments about the link between speech perception and loanword adaptation. Peperkamp & Dupoux (2003) attempts to formalize this connection. The authors propose that all loanword adaptations are phonetically minimal transformations that take place during speech perception; transformations that seem to arise from phonological transformations are attributed to perceptual processes involved in understanding non-native sounds. In the psycholinguistic model of speech perception they propose, non-native segments are mapped onto the closest phonetic category in the borrowing language by a "phonetic decoding module" that is part of the perceptual system. The authors propose that phonetic decoding module consists of complete word forms, not individual segments. They thus take the the cross-linguistic differences in loanword adaptations to be the result of differences in the "surface phonetic structure of individual languages." Peperkamp (2005) further asserts that loanword adaptations are coordinated by a separate (perception-based) system apart from the system that calculates phonological alternations. This model can account for repairs that occur in loanword adaptations but not in the native phonology. For example, in Korean [s] is not a licit coda of Korean and surfaces as [t] instead when it appears in the coda-position; however, loanwords that have [s] as the coda undergo epenthesis (e.g. English 'glass' is borrowed as [gɨrasɨ]). One of the author's main points is that a speech perception model that includes native phonological alternations is not the best way account for the patterns in loanword adaptation.

The strongest version of the Dupoux et al. (2011) and Peperkamp & Dupoux (2003) claim—that there is only one phonetically minimal vowel and it is the only vowel to appear in loanword adaptations—only predicts the appearance of [u] as the epenthetic vowel. The vowels [i, u, o] all appear as epenthetic vowels in Japanese loanword adaptations, as shown in the previous section. Therefore, the strongest version of their claim cannot account for all vowels in Japanese loanword adaptation. It is possible that a phonetically minimal vowel may differ depending on context; [u] may not be the vowel to best fit the transition between all consonant clusters. Alternatively, there is other machinery that can affect the quality of epenthetic vowel such as native language phonology. Other studies about perceptual epenthesis and Japanese loanword adaptation propose other factors that might influence the quality of an epenthetic vowel in addition to the vowel's phonetic properties. Durvasula & Kahng suggest that phonology is active during speech perception. They argue that phonological alternations can modulate the epenthetic vowel and that it is possible to elicit more than one epenthetic vowel. A study on perceptual epenthesis in Japanese by Monahan et al. (2009) tried to illicit more than one epenthetic vowel for Japanese speakers. Crucially both of these papers provide evidence that different consonantal contexts can induce epenthetic vowels other than the "phonetically minimal" one as Dupoux et al. (2011) asserts.

Based on loanword patterns in Japanese in which epenthetic [o] appears after coronal stops, Monahan et. al. (2009) hypothesized that Japanese speakers would hear the illusory vowel

[o] in consonant clusters where C_1 is [t] or [d]. They created nonce words of the form [eCVma], using coronal consonants [t, d], velar consonants [k, g], nasal consonants [m, n], and the vowels [u] and [o]. One set of stimuli contained the full vowel (either [o] or [u]) and another set was created by splicing out the vowel to form nonce words of the form [eCma]. In an AX discrimination task, participants were played pairs of stimuli and asked to decide whether the nonce words were the same or different. While participants confused [ekuma] and [ekma] as was predicted by Dupoux's 1999 findings, Japanese speakers correctly discriminated [edma-eduma], [etuma-etma], and [etoma-etma] pairs. Not only do their findings suggest that Japanese speakers do not illusorily epenthesize [o] after coronal segments, they also suggest that Japanese speakers do not perceive illusory [u] after coronal segments either. They conclude [o] cannot be illusorily epenthesized because it is too sonorant to mimic "absence" like [u] can, neither is it subject to phonological devoicing like Japanese [u]. As for the nature of perceptual epenthesis, they assert that native language phonology alone cannot account for perception of non-native speech.

Durvasula & Kahng directly address these results in a 2015 paper on perceptual epenthesis effects in Korean. They propose that it is possible to elicit more than one illusory vowel if the appearance of that vowel is supported by processes in the native phonology. Inspired by Bayesian models of speech perception, they take the task of the listener in speech perception to be one of reverse inference in which listeners "identify the best estimate of the intended underlying categories (phonemic/underlying representations) of the utterance given the acoustic token." At least two phonological processes could influence the quality of the illusory vowel: consistent deletion of a vowel and allophonic mappings. If a vowel is regularly deleted after a certain consonant ($V_1 \rightarrow [Ø] / C_1\_\_$), then the phonology of that language supports reverse inference of that vowel when there is no vowel present in the surface representation ($[Ø] \rightarrow V_1$ /

C_1__C_2). Therefore, the vowel chosen to repair a cluster containing C_1 will be V_1. If there is allophonic mapping at a consonant before a particular vowel (/C_1/ → [C_2] / __ V_2), a listener encountering C_2 would infer the presence of an underlying consonant and choose to repair the cluster with the vowel that triggers $C_1 \rightarrow C_2$. [ɨ] is the shortest vowel in Korean, deletes in context of vowel hiatus and in weak non-initial open syllables, and is often chosen as the epenthetic vowel in Korean adaptations of English words. It is thus the most probable illusory vowel for Korean speakers. Durvasula & Kahng also predict the presence of the illusory vowel [i] in the context of palatal segments like [ɕ] and [c^h] because Korean has a phonological process palatalizing alveolar consonants before /i/. They assert that when a Korean listener encounters a nonce word like [ec^hima], the surface consonant [c^h] could be interpreted as underlying phoneme /t^h/ or the phoneme /c^h/ which also exists as a contrastive phoneme of Korean; therefore, the nonce word could either be /ec^hima/ or /et^hima/ phonemically. If a Korean speaker interprets the underlying phoneme as /c^h/, then the following vowel would be [ɨ] because it is the phonetically minimal vowel, but if the Korean speaker interprets the phoneme to be underlying [t^h] then that the following vowel must be [i] because the only way to get phonetic [c^h] from phonemic /t^h/ is to have following vowel phoneme /i/. Therefore [c^h] could induce both [i] and [ɨ]; [t^h] or [s] should only induce the illusory vowel [ɨ]. Using AX and ABX discrimination tasks with stimuli containing palatal consonants (e.g. [ec^hima]), Durvasula & Kahng did manage to elicit two illusory vowels for Korean speakers ([i] and [ɨ]); they also hypothesize that it is possible to elicit more than one illusory vowel for Japanese listeners because Japanese also has a similar palatalization rule.

Mattingley et al. (2015) further explored the factors that modulate the quality of the epenthetic vowel by conducting an experiment in which Japanese native speakers were asked to listen to stimuli containing illicit consonant clusters and select which vowel they heard between the two consonants. [i] was frequently chosen after [dʑ] in responses and [u] was most often identified as the vowel following [g], [d], and [b]. This was surprising because [o] often follows coronal consonants in Japanese loanword adaptations, but [o] was not the most popular choice following [d]. Instead, participants identified the vowel following [d] as [u] even though [du] is not an observed sequence in native Japanese. (It occasionally appears in recent loanword adaptations such as English 'doom' and French 'vin du pays'). The question is why [u] would appear as an epenthetic vowel in places not predicted by Japanese phonology. This result also contradicts the findings of Monahan et al. (2009) where speakers perceived neither [u] nor [o] after the alveolar consonants [t] and [d].

In a follow-up study, Mattingley (2016) conducts a perception experiment to determine if the Mattingley et al. (2015) results were affected by the nature of the forced-choice vowel identification task. In an AX discrimination task, the author found that Japanese speakers had a difficult time distinguishing between the [tɕɯC]-[tɕC] pair and the [tɕi]-[tɕC] pair. Participants were also more likely to perceive [u], rather than [o], as the epenthetic vowel following the coronal consonants [t] and [d]. Moreover, [u] appeared as the illusory vowel between voiced segments though that is not a context in which high vowel devoicing would occur. Because the appearance of [u] seems to be generalizing beyond what is predicted by native Japanese phonotactics, the author speculates that a language's statistical patterns could support a strong perceptual bias towards [u]. The author leaves this question open to further consideration.

4 The Role of Statistical Patterns

4.1 My proposal

This study aims to address the question left open by Mattingley et al (2015): Could a language's statistical patterns influence the choice of epenthetic vowel? This possibility does not necessarily contradict claims about the role of phonology as asserted by Durvasula & Kahng (2015) and Monahan et al. (2009). While it is true that Japanese does have a palatalization rule, this phonological process alone is not likely to motivate the appearance of [i] as an epenthetic vowel in perception or loanword adaptation because palatal consonants [tɕ, dʑ] are phonemic, not allophonic. That is, [tɕa, tɕo, tɕu] are all licit sequences in Japanese; the appearance of these consonants does not require [i] as the following vowel.

I believe that loanword adaptation is influenced by a listener's expectations, and these expectations are informed by patterns in the lexicon and phonological rules. The model of loanword adaptation I'm proposing would include two steps. First, phonological rules specify certain features that an epenthetic vowel should have. In Japanese, the rule would specify that a high vowel should break up the cluster. In other languages too, only a subset of vowels could possibly be chosen as epenthetic vowels. Second, listeners consult fine-grained the patterns in the lexicon to choose the likeliest epenthetic vowel. Under this model, it is possible to make predictions about which vowel is more likely to be epenthesized in specific contexts based on the how the vowel patterns in the lexicon.

4.2 Corpus Work

To understand the role of statistical patterns in the choice of epenthetic vowel, I first investigated vowel distributions in Japanese in the Corpus of Spontaneous Japanese. CSJ is an online database (http://pj.ninjal.ac.jp/corpus_center/csj/en/) containing a large amount of Japanese spoken language data developed jointly by NINJAL, NICT and the Tokyo Institute of Technology. The corpus contains about 650 hours of spontaneous speech and about 7 million words. The recordings span a variety of topics, from academic presentations to everyday topics spoken by a wide range of native Japanese speakers. Using the transcriptions of the speech in CSJ and formulas in Excel, I calculated probabilities of each vowel appearing after each consonant in the Japanese phonemic inventory. The statistics here were not controlled for word-position; nor is the identity of the following consonant (if any) taken into account. Table 1 shows the frequency of each vowel following the Japanese phonemes [p], [k], [n], [s], [tɕ], [dʑ], and [ɕ]. In the last column, I've indicated the most frequent vowel following each consonant.

Table 1: Vowel Distribution in Japanese (In Percent)

Consonant	[a]	[i]	[u]	[e]	[o]	Most Frequent
p	29	17	33	10	10	[u]
k	27	15	31	11	16	[u]
n	39	21	5	14	21	[a]
s	23	1	37	25	14	[u]
tɕ	6	63	10	0	22	[i]
dʑ	4	52	21	4	19	[i]
ɕ	5	65	12	0	17	[i]

According to the corpus statistics, each vowel can appear after each consonant. The only exception to this is [e], which does not appear after [tɕ] and [ɕ] and very infrequently after [dʑ]. This is due to the fact that [tɕe] and [ɕe] are not well-formed Japanese syllables (Ito & Meister 2003), though they may occasionally appear in loanword adaptations such as シェーク[ɕe:ku] from the English 'shake.' Similarly, [i] very rarely follows [s] because the [si] sequence in Japanese becomes [ɕ] through a palatalization rule. There are some general trends in the data. [u] appears more frequently than any other vowel after [p], [k], and [s]; however, it should be noted that [u] is not so much more probable than any other vowel following these consonants. For example, [a], which has a frequency of occurrence of 29% after [p], is a close second for the most frequent vowel. [i] appears more frequently than any other vowel following [tɕ], [dʑ], and [ɕ] by a wide margin. [a], [e], and [u] are notably lacking— but not completely absent—where [i] is most frequent, except for [dʑ]. [a] is the most frequent vowel following [n], and [i] and [o] are tied for second most frequent vowel.

4.3 Hypothesis and Predictions

Japanese speakers could follow any of the following four strategies when choosing an epenthetic vowel. Each of these possibilities has slightly different implications for the mental grammar.

Hypothesis I Japanese speakers follow a rule that inserts the same epenthetic vowel after C1 of an illicit consonant cluster. This rule is not sensitive to the identity of C1. This hypothesis predicts that as the probability (lexical statistics) of the predicted vowel increases, the observed responses containing that vowel remain constant. If this were true, I would expect only one epenthetic vowel (either [i] or [u]) to appear in all contexts.

Hypothesis II Japanese speakers first apply a rule that says the epenthetic vowel must be a high vowel. They then access their language's statistical patterns in a categorical fashion to find the most frequent high vowel following C1. Under this hypothesis, listeners would consistently choose [u] as the epenthetic vowel following [p], [k], and [s] because it is the likeliest vowel according to the corpus and never choose [i] as the epenthetic vowel in these conditions.

Hypothesis III Japanese speakers first apply a rule that says the epenthetic vowel must be a high vowel. Then, when choosing which high vowel to insert, they access their language's statistical patterns and apply them granularly. This hypothesis predicts that as the probability of a high vowel V gradually increases following a given consonant C1, responses containing V also gradually increase following C1.

Hypothesis IV Japanese speakers do not apply a rule specifying epenthesis of a high vowel. They access the probabilities of all Japanese vowels and apply this information granularly. This hypothesis predicts that any vowel can be chosen as the epenthetic vowel, and as the frequency of this vowel increase, responses containing V following C1 also increases. For example, [a] would appear most frequently after [n] in the responses, followed by [i] and [o].

To understand which hypothesis best represents the process of choosing an epenthetic vowel, I conducted an online transliteration experiment in which native Japanese speakers were presented with nonce words containing illicit consonant clusters and asked to write those words in Japanese *katakana*. These instructions will force participants to choose a vowel following C1 because it is impossible to write a singular consonant (except for the alveolar nasal [n]) in Japanese orthography. I then analyzed vowel distribution in their responses and compared them to the corpus vowel distribution.

Following the patterns in Japanese loanword adaptations and the results from previous studies, I do not expect all of these vowels to appear as epenthetic vowels. [i], [u], and [o] are the only vowels chosen to break up clusters in foreign words. In the perceptual studies by Monahan et al. (2009), [o] did not appear illusorily after the coronal consonants [t] or [d] as predicted by loanword adaptations. In the vowel identification experiment carried out by Mattingley et al.

(2015), very few (~10%) participants identified [o] as the vowel following coronals and almost none identified [o] as the epenthetic vowel following non-coronal consonants. Therefore, I do not expect [o] to appear after non-coronal consonants. Cross-linguistically, centralized or high vowels are often chosen as epenthetic vowels and are very rarely rounded (Hall 2011; de Lacy 1999; de Lacy 2006). It seems that the appearance of [o] as an epenthetic vowel in loanword adaptations is not mostly based in perception, if at all. Furthermore, I do not expect [o] to appear as an epenthetic vowel in the results because the contexts in which it appears in loanword adaptation (after coronal consonants) is not investigated here. I posit that Japanese speakers only have two possible epenthetic vowels to choose from: [i] and [u]. Because the vowel distributions of [i] and [u] are of particular interest, I have reworked the lexical statistics to reflect the ratio of [i] to [u] in the corpus. These reworked probabilities are shown in Table 2. These are the corpus statistics I will compare to the vowel distribution on the responses.

Table 2: Ratio of [i]/[i u] and [u]/ [i u] in Japanese (in Percent)

Consonant	[i]	[u]
p	34	66
k	33	67
n	81	19
s	1	99
tɕ	86	14
dʑ	71	29
ɕ	84	16

Because the choice of epenthetic vowel has shown to be affected by the position of the cluster in the word (Yun 2012), I will compare the results for word-medial position to the pattern of epenthetic vowels in word-initial position. Based on the theory of positional faithfulness (Beckman 1998) I posit that consonant clusters in word-initial position are more salient and have stronger perceptual cues. If there are stronger perceptual cues, then speakers will rely less on their expectations, which are informed by lexical statistics, to perceive and comprehend the speech signal. This would result in speakers choosing the phonetically minimal vowel [u] more frequently as the epenthetic vowel in word-initial consonant clusters. Conversely, in contexts with weak perceptual cues, word-medial position in this case, speakers must rely more on their expectations to fill in the gap left by insufficient perceptual cues. I hypothesize that this would result in speakers choosing the most frequent vowel (as predicted by the lexicon) as the epenthetic vowel in word-medial consonant clusters. Again, the lexical statistics shown in the previous section do not control for word-position. Therefore the vowel distributions in word-medial and word-initial positions will both be compared to the corpus vowel distribution shown in Table 2. While it may not be the most accurate comparison, I believe an analysis of the vowel distribution in the responses is a good starting point in determining the effects of word-position.

5 Experiment Design

5.1 Materials

I created 42 nonce words for the transliteration experiment. Auditory stimuli were recorded in Praat by a female native speaker of Modern Israeli Hebrew. The speaker was shown one nonce word at a time on a slide in Microsoft PowerPoint and read through the PowerPoint slides three times. The stimuli were then checked for quality and accuracy in Praat. Amplitude was normalized and spectrograms were scrutinized to ensure the absence of a transitional vowel between the consonant clusters. In the event that one stimulus was pronounced incorrectly, another repetition of the word was used in its place.

The stimuli contained the illicit consonant clusters [pt], [kt], [nt], [st], [tɕt], [ɕt], and [dʑd], where C2 was always [t] except when C1 was a voiced obstruent. The consonants [p] and [k] were chosen because they are known to induce [u] epenthesis; the fricatives and affricates were chosen because they are most frequently followed by the vowel [i]. The nasal consonant [n] is of interest because there are very few instances of a nasal cluster inducing epenthesis in loanword adaptations and as such it is not clear which vowel would be chosen as the epenthetic vowel following [n] if any at all. Therefore it was a good place to test whether the "phonetically minimal" vowel or the most frequent vowel would be chosen, if any vowel would be chosen.

The total 42 stimulus items were split into sublists (called Block A and Block B), each containing 21 questions. Each block contains three sets of stimuli which were created from these clusters. In Set A, each cluster appears in word-medial position; in Set B, each appears in word-initial position. Set C consists of CVCVCV fillers. Consonant clusters were placed within phonetic frames to distract participants from the goal of the experiment. For word-medial position, one phonetic frame was [ge__afa], creating [geptafa], [gektafa], and [gentafa], [gestafa], [getɕtafa], [gedʑdafa], and [geɕtafa]. The other frame was [de__am] was used to create stimulus items [deptam], [dektam], [dentam], [destam], [detɕtam], [dedʑdam], and [deɕtam]. These nonce words were then shuffled so that Set A of Block A contained [pt], [kt], and [nt] in the [ge__afa] frame and the other consonant clusters in the [de__am] frame; Set A of Block B contained [pt], [kt], and [nt] in the [de__am] frame and the other consonant clusters in the [ge__afa] frame. For word initial position (Set B), the consonant clusters were followed by two phonetic "tails", -am and –afi. Set B of Block A contained nonce words in which [pt], [kt], and [nt] were placed before –am and the other consonant clusters before –afi; Set B of Block B contained nonce words in which [pt], [kt], and [nt] were combined with phonetic tail –afi and the rest of the clusters were combined with –am. Set C consists of CVCVCV fillers, which lacked consonant clusters of any kind. C1 and C2 of the fillers correspond with C1 and C2 of the target consonant cluster. The two phonetic frames used for Set C were [__o__aru] and [__o__afu]. Set C of Block A contained all [__o__aru] fillers, and Set C of Block B contained all [__o__afu] fillers. The fillers follow the phonotactic constraints of native Japanese words and therefore are not expected to trigger epenthesis. Below I have listed examples of stimuli experiment; for the full list of stimuli, see the Appendix.

Table 3: Example Stimuli used in the Experiment

Consonant Cluster	Word-medial	Word-initial	Filler
pt	geptafa	ptam	potaru
kt	gektafa	ktam	kotaru
nt	gentafa	ntam	notaru
st	destam	stafi	sotaru
tɕt	detɕtam	tɕtafi	tɕotaru
dʑd	dedʑdam	dʑdafi	dʑodaru
ɕt	deɕtam	ɕtafi	ɕotaru

5.2 Procedure

The experiment was conducted online using an online survey platform called Qualtrics; the experiment can be found here: https://yalesurvey.qualtrics.com/SE/?SID=SV_8HOOzAxpJwSCQAd. The landing page, which describes the experiment and explains the task, is shown below.

Participants had to click the blue button to begin the survey. The survey contained a total of 27 questions and took about 10 minutes to complete. For questions 1-21, participants were randomly presented with stimulus items from either Block A or Block B. They were instructed to listen to one “foreign word” at a time and write the word in Japanese *katakana* (外国語の言葉を聞いてカタカナで書いてください). No constraints were placed on the possible responses other than being written in *katakana*. A screenshot of a sample item is shown below.

Stimulus items within each block were also randomized. The experiment was self-paced, and there was not a limit to how many times participants could listen to each stimulus. In the remaining questions, participants provided information about their age, biological sex, hometown, current country of residence, and languages spoken other than Japanese. No feedback was given during the experiment. Both full (100% completion) and partial responses (~100% completion) were recorded, but only full responses were analyzed.

5.3 Participants

The survey link was distributed to friends and colleagues who are native speakers of Japanese. 35 native speakers of Japanese participated in the online experiment over the course of two weeks. 21 identified as monolingual; that is, they reported limited exposure to any foreign language (e.g. English). 14 identified as bilingual, reporting a high exposure to a foreign language through time lived abroad or a bilingual upbringing. 11 participants were between the ages of 18 and 25; 7 were between 26 and 33; 8 were between 34 and 41; 6 were between 42 and

49, and 4 were over the age of 50. 19 were female and 16 were male. Participants were not compensated for their time.

6 Results

The results for bilingual and monolingual participants pattern similarly and aggregating the responses did not qualitatively change the results. Therefore I present the results together here.

In the majority of responses, participants broke up the clusters via epenthesis. One special case, however, is the [nt] cluster. As expected, [nt] in medial position was not subject to vowel epenthesis because [n] could be parsed as a coda in conformity with Japanese phonotactic constraints and can be written as a single character in Japanese orthography. Surprisingly, the [nt] cluster was tolerated in initial position and written with the character for a placeless nasal [N] followed by the character for [ta] (ンタ) 87% of the time even though this is not a licit word-initial sequence in Japanese. This result will be further discussed in Section 7. Other repairs to the cluster were more predictable. In some cases when the consonant cluster was word-initial (such as [ktam]) participants deleted the first consonant or had trouble perceiving C1 in the first place resulting in [tamu] or [tan]. There were very few instances of deletion in word-medial position. There were other cases in which consonant substitution occurred ([ptam] being written as [kutamu], for instance) and where changes of voicing occurred ([gekutafa] → [geggudafa]). All of these cases were excluded from the analysis because they do not reveal anything about the choice of epenthetic vowel after the consonants of interest.

Table 4 shows the number of times each vowel [i] or [u] appeared after each consonant in the responses.

Table 4: Vowel Distribution in Responses (Raw Counts)

Consonant	Vowel	
	[i]	**[u]**
p	8	36
k	18	44
n	2	7
s	2	65
tɕ	59	11
dʑ	51	18
ɕ	39	31

[a], [e], and [o] never appeared as epenthetic vowels after the target consonants. This is in agreement with previous studies on perceptual epenthesis in Japanese speakers and loanword adaptation patterns in Japanese. Overwhelmingly, [u] was chosen as the epenthetic vowel after the consonants [p], [k], [s], and [n] in the few cases it triggered epenthesis. [i] was consistently chosen as the epenthetic vowel following the affricates [tɕ] and [dʑ] and after fricative [ɕ] by a small margin. A Pearson's chi-squared test on the exact numbers shown in Table 4 reveals that these vowel distributions are very likely not random ($p < 2.2e-16$), meaning there is likely an interaction between C1 and the following consonant. Binominal tests were run in the R statistical computing environment (R Core Team 2016) to understand the significance of these raw data; the results are shown in Table 5. Table 5 focuses on the likelihood of [i] appearing as an epenthetic vowel under the hypothesis that [i] and [u] are the only two possible epenthetic vowels.

Table 5: Binomial Test Results for [i] Responses

Consonant	Success	Trials	Expected	Probability of Success	p
p	8	44	0.5	0.1818182	2.545e-05*
k	18	62	0.5	0.2903226	0.001299*
n	2	9	0.5	0.2222222	0.1797
s	2	67	0.5	0.02985075	< 2.2e-16*
tɕ	59	70	0.5	0.8428571	4.466e-09*
dʑ	51	69	0.5	0.7391304	8.769e-05*
ɕ	39	70	0.5	0.5571429	0.403

The binomial tests were calculated with a hypothesized probability of success ("Expected") of 0.5 because only two vowels can be chosen to break up the consonant clusters. For [p], [i] was chosen as the epenthetic vowel 8 out of 44 times; the p-value from the binomial test (p=2.545e-05) shows that this result is significantly different from the expected response, and I reject the null hypothesis that [i] would be chosen 50% of the time after [p]. Binomial tests for [k] and [s] show a similar result. For the affricates [tɕ] and [dʑ], [i] was chosen as the epenthetic vowel 59 out of 70 times and 51 out of 69 times respectively. The p-values from the binomial tests are significant (p= 4.466e-09 and p= 8.769e-05), meaning I reject the null hypothesis that [i] would be chosen 50% of the time after [tɕ] and [dʑ]. The estimated probability of success (of [i] appearing as the epenthetic vowel) is relatively high for both consonants (0.8428571 and 0.7391304). The p-values from the binomial tests for [n] and [ɕ], on the other hand, were not significant ($p > 0.05$); therefore for these consonants I cannot reject the null hypothesis. There is a possibility that the distribution of [i] after [n] and [ɕ] is due to the 50-50 chance of [i] being randomly chosen.

The data in Tables 4 and 5 allows us to make some preliminary conclusions about the quality of the epenthetic vowel. First, by the significant p-values, it appears that the choice of epenthetic vowel is influenced by the preceding consonant and not determined at random (with only [n] and [ɕ] as possible exceptions.) This contradicts Hypothesis I, which says the epenthetic vowel is not influenced by the preceding consonant. This supports claims of Durvasula & Kahng (2015) and Mattingley et al. (2015), given that phonetic miniality is not context dependent; the data contradicts Dupoux et al. (2011) which asserts that the epenthetic vowel arises from the perception system and is not sensitive to contextual information. The fact that [i] appears at all as an epenthetic vowel calls for revision of Dupoux et al.'s model of epenthesis and loanword adaptation. However, this data does not necessarily contradict Dupoux et al's (2011) other claim—that [u] is the "phonetically minimal" vowel of Japanese. It could be that Japanese has two "phonetically minimal" vowels according to the definition set forth by Dupoux, or [u] is not actually a "phonetically minimal vowel" in all cases or epenthesis takes factors other than phonetic minimality into account. The phonetics of [i] and [u] will be discussed in Section 7.

The response data alone does not allow us to evaluate Hypotheses II and III because it does not include corpus data. I need to compare the observed vowel distribution to the predicted vowel distribution to evaluate how Japanese speakers might be applying patterns derived from the lexicon. Table 6 shows these distributions in percentages. [n] has been excluded because of the insufficient number of epenthetic responses.

Table 6: Vowel Distribution of [i] in Responses (in Percent)

	p	k	s	tɕ	dʑ	ɕ
Observed	18	29	3	84	74	56
Predicted	34	33	1	86	71	84

The 'Predicted' percentages here reflect the ratio of [i] to [u] in the corpus. Table 6 shows the observed distribution of [i] (1st row) versus the predicted distribution of [i] (2nd row). [i] appeared half as frequently as expected after [p]. For the consonants [k], [s], [tɕ], and [dʑ], the frequency of [i] is remarkably close to the frequency in the corpus, with a differences no greater than 4%. [u] appeared more frequently than expected after [p] and [ɕ], and appeared as expected following [k], [s], [tɕ], and [dʑ].

Figure 1 shows the data in Table 6 graphically, which will allow us to ascertain the strength of the correlation.

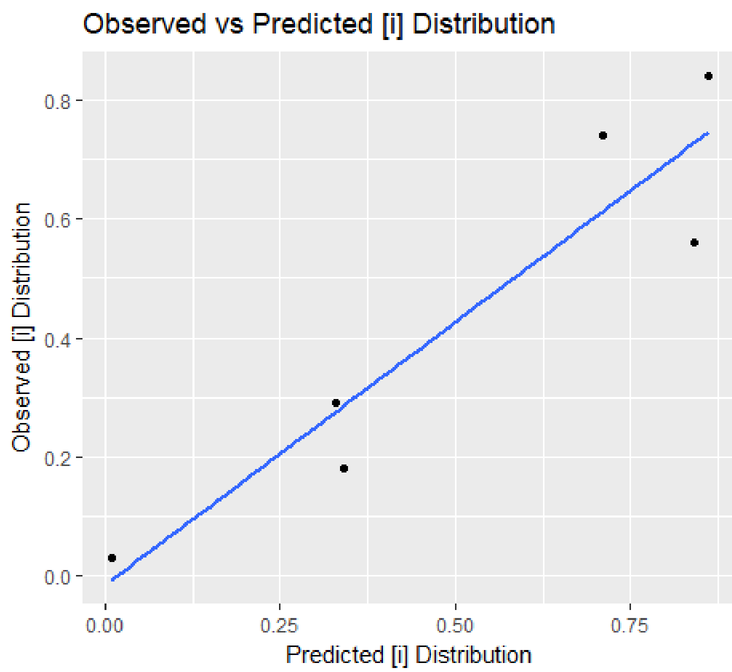

Figure 1: Correlation of observed (in responses) vs. predicted (corpus) [i] distribution

Figure 1 plots the observed percentage of epenthetic [i] and the percentage of [i] appearing after each consonant according to the Corpus of Spontaneous Japanese. This scatterplot was created with the ggplot2 function in R (Wickham, Hadley 2009), and the blue solid line was created by ggplot2's linear scatterplot smoother (geom_smooth) to aid the eye in seeing the general pattern in the scatterplot data.

According to Pearson's correlation test, there is a strong correlation between the observed and predicted percentages of [i] epenthesis (t = 5.265, df = 4, p-value = 0.006233) and a correlation coefficient of 0.9348244. So, as X increases, Y increases. That is, as the probability of [i] following C1 increased in the lexicon, Japanese speakers were more likely to choose [i] as the epenthetic vowel. This supports Hypothesis III, which says that listeners follow a rule that specifies the height feature of the epenthetic vowel and then apply the distribution of high vowels in the lexicon granularly.

Now I address the second part of my main hypothesis, which states that in contexts with weak perceptual cues, listeners are biased to perceive the most expected vowel in the language, and when perceptual cues are strong, listeners are biased to perceive the phonetically minimal vowel in the language. This means I expect more occurrences of [u] (and fewer occurrences of [i]) in word-initial position, which is known to be more perceptually salient (Beckman 1998). I used subsetting code in R to extract the word-medial responses and the word-initial responses and formatted the data into two contingency tables. Simplified versions of these contingency tables can be found in the Appendix; though the contingency tables themselves are omitted here, the raw count data will be discussed below alongside the results of the binomial tests.

Table 7: Binomial Test Results for [i] Responses (Word-medial)

Consonant	Success	Trials	Expected	Probability of Success	p
p	3	23	0.5	0.1304348	0.0004883*
k	8	35	0.5	0.2285714	0.001878*
s	1	32	0.5	0.03125	1.537e-08*
tɕ	29	35	0.5	0.8285714	0.0001168*
dʑ	27	34	0.5	0.7941176	0.0008214*
ɕ	22	36	0.5	0.6111111	0.243

Table 8: Binomial Test Results for [i] Responses (Word-initial)

Consonant	Success	Trials	Expected	Probability of Success	p
p	5	21	0.5	0.2380952	0.0266*
k	10	27	0.5	0.3703704	0.2478
s	1	35	0.5	0.02857143	2.095e-09*
tɕ	30	35	0.5	0.8571429	2.236e-05*
dʑ	24	35	0.5	0.6857143	0.04096*
ɕ	17	34	0.5	0.5	1

As with the previous data, the binomial tests for word-initial and word-medial conditions were calculated with a hypothesized probability of success of 0.5 because there were only two possible epenthetic vowels. For the word-medial condition in Table 7, binomial tests yield significant p-values for all consonants except [ɕ]; these results are similar to the ones found in Table 5. Therefore I reject the null hypothesis that [i] would be chosen 50% of the time after all consonants except [ɕ]. The results for the word-initial condition are slightly different. Binomial tests for [p], [s], and [tɕ] yield significant p-values, but the p-value for [k] is insignificant in word-initial position and [dʑ] is barely significant.

The vowel distribution in word-medial stimuli is very similar to the word-initial vowel distribution. There are noticeably more occurrences of [i] after [ɕ] in word-medial position, where perceptual cues are hypothesized to be weaker and more occurrences of the phonetically minimal vowel [u] in word-initial position. This could be an indication that listeners are biased to perceive the "phonetically minimal" vowel when perceptual cues are strong as our hypothesis asserts. To understand whether the effects of word-position are significant, I examine the interactions of variables using a mixed logistic regression models.

This analysis only includes responses in which epenthetic occurred in target items. As previously mentioned, responses to the [nt] cluster were not included. The levels of the factor POSITION were coded as 0 for the word-medial condition and 1 for the word-initial condition, which makes word-medial the baseline. The levels of the factor VOWEL were coded as 1 for [i] and 0 for [u] under the expectation that [u] is the "default" epenthetic vowel, occurring after a wider range of consonants. The probability of [i] occurring in the corpus was coded as a continuous predictor. There were two fixed effects terms, POSITION and [I] PROBABILITY, and a random effect was included for participant intercepts to account for inter-listener variability. The model was fit by maximum likelihood using the glmer function in the lmerTest library of R (Alexandra Kuznetsova, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen, 2016). The initial model tries to quantify how well the vowel can be predicted by the interaction between word-position and the corpus [i] probability. Estimates for fixed-effects, standard errors, and z-scores are shown in Table 9 below.

Table 9: Model #1 Summary

	Estimate	Std. Error	z-value	p
Position	0.4398	0.7501	0.586	0.558
prob_i	1.1066	0.1517	7.296	2.96e-13
position:prob_i	-0.1381	0.1850	-0.747	0.455

According to the insignificant p-value neither word-position nor the interaction of word-position and corpus probability were significant factors in determining the epenthetic vowel. The significant p-value for prob_i (p=2.96e-13) indicates a significant effect of corpus probability. I then reduced the model by omitting non-significant terms starting with the interaction of word position and vowel probability. An ANOVA comparison of the first and second model reveals that Model #2 without the interaction (log Likelihood = -182.78) is able to account for the data just as well as Model #1 (logLik = -182.50) with one less degree of freedom and a p-value of 0.4546; AIC and BIC scores agree with logLik values. This means that including the interaction of word-position does not help significantly to explain patterns in the data. In the third model I removed position as a factor entirely. This model (logLik = -182.82) accounts for the data just as well as Model #2 (logLik = -182.78) with a p-value of 0.7773. The reduced model's Estimates for fixed-effects, standard errors, and z-scores are shown in Table 10.

Table 10: Model #3 Summary

	Estimate	Std. Error	z-value	p
Intercept	-4.0686	0.5242	-7.762	8.36e-15
prob_i	1.0346	0.1141	9.069	< 2e-16

The simplest model that captures the data only takes the corpus probability into account. This model confirms that as the predicted probability of [i] increases, the percentage of [i] in the responses increases as predicted by Hypothesis III. These results do not support positional effects, but neither do they strongly suggest the absence of them either. It is possible that my data set was simply too small to see the effect of word-position. Therefore, I cannot draw any conclusions about the effect of word-position on the choice of epenthetic vowel.

7 Discussion

7.1 Evaluating Hypotheses

This paper addresses the question of whether a language's statistical patterns affect the choice of epenthetic vowel and has also attempted to address whether the strength of perceptual cues can bias listeners toward statistical patterns. I formulated four hypotheses regarding the role of a language's statistical patterns. Hypothesis I predicts that listeners would choose the same epenthetic vowel (either [i] or [u]) no matter what the preceding consonant. Hypothesis II says that listeners apply statistical patterns categorically when choosing an epenthetic vowel in a winner-takes-all fashion. If, for example, [u] appears after [s] 50% of the time and [i] appears

20% of the time in the lexicon, listeners would only choose [u] as the epenthetic vowel because it is the most frequent vowel following that consonant. Hypothesis III says that listeners follow a rule that specifies the height feature of the epenthetic vowel and then apply the distribution of high vowels in the lexicon granularly. Hypothesis IV says that listeners apply statistical patterns granularly without first filtering out non-high vowels. My inspection of Japanese statistical patterns via the Corpus of Spontaneous Japanese revealed different vowel probabilities depending on the preceding consonant. If Japanese listeners are more likely to choose a vowel V after C1 as the probability of that vowel increases in the corpus data, this could be an indication that this fine-grained lexical data informs the listener's expectations.

I tested these hypotheses through an online transliteration experiment. Native speakers of Japanese were asked to listen to nonce words containing illicit consonant clusters [pt], [kt], [nt], [st], [tɕt], [ɕt], and [dʑd] and write what they heard in Japanese orthography. Because it is impossible to write a consonant without a vowel (excluding the nasal [N]), participants are forced to choose an epenthetic vowel to break up the cluster. I found that the vowel distribution in their responses overall reflected the vowel distribution (of possible epenthetic vowels) in the corpus. For example, the probabilities of [u] in the corpus following [k], [s], [tɕ], and [dʑ] were 0.67, 0.99, 0.14, and 0.29 respectively and the distribution of [u] in the responses following the same consonants were 0.71, 0.97, 0.16, and 0.26. Chi-squared and binomial tests on this data show there is an interaction between C1 and V. Plotting these distributions reveals a strong correlation between the observed vowel distribution and the predicted vowel distribution. These results indicate that there are two possible epenthetic vowels and the choice of epenthetic vowel is influenced by the preceding consonant, contra Hypothesis I. Only two vowels out of five appeared as epenthetic vowels in the responses, contra Hypothesis IV. The strong, nearly one-to-one correlation between observed and predicted responses supports Hypothesis III which says that listeners access fine-grained statistical patterns in their language to inform the perceptual system when encountering sequences that violate native phonotactics.

However, the data does not completely discount Hypothesis II, which says that Japanese speakers apply a rule that specifies height of the epenthetic vowel and then apply statistical patterns categorically. This hypothesis allows for a series of very specific phonological rules regarding the epenthetic vowel. For example, there could be a rule that says to insert [i] following palatal affricates and fricatives and another rule to insert [u] following every other non-palatal consonant. Once the predicted probability becomes extremely high (say, 70%) then the observed responses containing that vowel would also suddenly increase. One could imagine that a best fit line for such data would share some similarities with the correlation shown in Figure 1; however, instead of a gradual increase in both y and x values, I would expect points gathered near (0,0) and another set of points clustered near (1,1) because one high vowel is always chosen while the other is excluded. The data presented here does not allow me to make a strong statement about which of the two hypotheses is more accurate, but the fact that there are some [i] responses following non-palatal consonants (20-30%) and some [u] responses following palatals (20-30%) could mean these kinds of statistically-informed categorical rules are not present in the mental grammar.

7.2 The behavior of [n] and [ɕ]

The observed responses generally corresponded to the vowel distribution within the corpus. However, responses to the [nt] and [ɕt] clusters were significantly different from what was expected. The predicted ratio of [i] to [u] responses following [ɕ] was roughly 5:1 (0.84 and 0.16), but the ratio in the observed responses was closer to 1:1 (0.56 and 0.44). This is surprising considering [i] was by far the most frequent vowel after [dʑ] and [ʧ] in the responses. Why would [ɕ] pattern differently? To my knowledge there are no outstanding differences between the phonetic properties of fricative [ɕ] and the affricates that would induce a different epenthetic vowel. It could be related to asymmetries in the nativization of palatal clusters discussed by Ito & Mester (1995). In a discussion about the differences in native Japanese phonology and loanword phonology, the authors turn to discuss the changing behavior of palatals in loanword adaptation. Originally the constraint *TI was active in loanword phonology; when the source word contained a [nonpalatal] + [i] sequence, the consonant was either palatalized (English 'team' became [ʧi:mu], not [ti:mu]) or the height of the following vowel was changed (English 'disco' was borrowed as [desuko] as opposed to [disuko].) In more recent loanwords, the [ti] sequence remains unchanged, but there is a related constraint that remains active: the *SI constraint. This constraint requires that coronal fricatives are palatal before [i]. As a result, "a fricative coronal followed by [i] always appears as palatal, even though its stop counterpart does not." (Ito & Mester 1995: 833). If it is the case that [i] triggers palatalization of the preceding fricative [s] even in recent loans, then I would expect [i] to appear more often as the epenthetic vowel following [ɕ]. One would also expect fewer [i] responses following [dʑ] and [ʧ] because [di] and [ti] sequences are acceptable in recent loanwords. But this is not the pattern that surfaces in the results. There are fewer [i] responses following [ɕ] and more [i] responses following [dʑ] and [ʧ]. The asymmetry described by Ito & Mester, then, does not seem to be the best way to explain the treatment of [ɕ] in the experiment. I leave this open for future investigations.

Because [n] is a licit coda of Japanese, I expected that the word-medial [nt] cluster would not undergo epenthesis. To my knowledge [#nt] is not a licit onset cluster of Japanese, so I expected a high percentage of participants to repair the cluster with epenthesis or deletion. Yet, only 9 (out of 69) responses show repair via epenthesis. Most participants remained faithful to the stimulus, transcribing it as [ntam]. Why would participants allow this cluster but repair every other cluster? This could be an effect of the orthography. After all, [nt] is the only cluster that can actually be represented faithfully in Japanese *katakana*. Perhaps participants heard the [nt] cluster and all other clusters accurately, but were only able to write [nt]. Though this is unlikely because previous perceptual experiments have shown that Japanese speakers perceive epenthetic vowels in real time, this is worth further investigation. It would be best to conduct a perception experiment alongside a transliteration experiment. This way one could determine whether there are differences in what speakers hear versus what they write. If participants were to actually perceive all consonant clusters correctly, then the appearance of consonants in written answers would be the result of the Japanese orthography. If participants were to actually perceive vowels between consonants in the perceptual experiment, then we can conclude that the orthography accurately represents what is perceived.

There is yet a third possibility. If [nt] is the only cluster that does not induce epenthesis in perception, then there might be something about the properties of the sequence that differentiate it from others. It could be that the sonority of [n] makes it a more acceptable C1 than fricatives or plosive consonants. Participants possibly felt that it would be worse to epenthesize than to remain faithful to the original speech signal. This phenomenon could possibly be represented with phonetically-based constraint rankings in an OT analysis, but I leave this to future research.

7.3 The "phonetically minimal" vowel

Finally, I will address a point raised in Section 6 about the possibility of more than one "phonetically minimal" vowel. Recall that one of the main claims of Dupoux et al. (2011) was that there is only one phonetically minimal vowel per language. In their discussion of Japanese vowels, the authors conclude that [u] is the phonetically minimal vowel of Japanese because it is the shortest in duration and prone to devoicing in certain contexts. This claim is supported by Yoshida (2006) and Mattingley (2016) who report that [u] is the shortest of the five Japanese vowels in both voiced and voiceless contexts (between [k] and [t] and [g] and [d], for example) and it is the least likely to be accented in Tokyo Japanese. These phonetic characteristics of [u] are what make it the closet vowel to zero, or no vowel; as such it is believed to be the best vowel to match the transitional space between two consonants in both perception and loanword adaptation. Steriade (2001, 2008) also supports the similar view that loanword adaptation is driven by perceptual factors, and the choice of epenthetic vowel is based on the similarity between an individual vowel segment and no segment. Steriade asserts that schwa is a popular epenthetic vowel cross-linguistically because of its short duration and its variability in certain contexts.

The results of the present study do not necessarily negate the claim the epenthetic vowels are only “phonetically minimal” vowels. It could be entirely possible, though, that the phonetically minimal vowel is not fixed, as the results of Dupoux et al. (2011) might suggest. In theory, both [i] and [u] could be phonetically minimal vowels if, in certain contexts, [i] is shorter than [u]. Overall, [i] only seems to be slightly longer than [u] on average or about the same duration. Mattingley et al. (2016), for instance, reports that [i] has an average duration of about 55ms in voiceless contexts and [u] was reported to have an average duration of about 40ms. Nishi et al. (2008) report [i] to have an average duration of 38ms and [u] to have an average duration of 37ms. Kawahara & Shaw (submitted) report the average duration of [i] and [u] as 54ms and 52ms respectively. However, vowel duration has been shown to be affected by a number of factors, including the voiced feature of the adjacent consonants, the length of the preceding consonants, and syllable structure (Campbell 1999, Kawahara 2006). Kawahara & Shaw discuss how these factors as well as consonant entropy influence vowel duration, and also chart the differences in vowel duration following certain Japanese phonemes. The values they report confirm that [i] and [u] are shorter than the other 3 vowels after every target consonant. The two vowels have similar durations following most of the consonantal contexts reported (which include: [w], [hy], [sy], [y], [zy], [d], [z], [n], [b], [s], [g], [k], [t], [r] [h], and [m]). [u] is noticeably shorter than [i] following [g], [n], [hy], and [i] is slightly shorter than [u] following [b] and [s]. Most notable however, is how much the duration of [u] is affected by the preceding consonantal factors. The vowel [u] is shortened by 12~15ms depending on the consonant's place of articulation and by about 11ms following voiceless consonants. While the vowel [i] is also shortened by about the same amount following voiceless consonants (-12ms), the preceding consonant's place of articulation never changed the vowel duration of [i].

From these results, it seems that average vowel duration and the tendency to devoice in certain contexts may not be the best determining factors in defining a "phonetically minimal" vowel because [i] and [u] are about the same length and both devoice or delete. Where the vowels seem to differ is in their malleability; that is, [u] is more variable in general than [i]. This makes the appearance of [i] as the epenthetic vowel in perceptual epenthesis and loanword adaptation even more surprising, since it does not have the variability that Steirade (2008) claims epenthetic vowels should have. A purely phonetic account of loanword epenthesis—wherein there is only one epenthetic vowel and it is chosen for its short duration and variability— is insufficient for explaining the patterns shown here and by others; yet a phonological account alone is also insufficient because it cannot explain why [u] appears in so many epenthetic contexts not predicted by Japanese phonotactics, like after coronal consonants. It is impossible to account for perceptual epenthesis with a purely phonetic model or a purely phonological rule-based model, or a statistical model. The observed patterns are best explained with a combination of all three mechanisms.

8 Conclusion

In conclusion, the current study compared the vowel distribution in transcriptions of nonce words to the vowel distribution in CSJ to understand the influence of lexical statistics on the choice of epenthetic vowel in loanword adaptations. Though each vowel can appear after every target consonant, only a subset of these can appear as epenthetic vowels. I therefore posited that there is a phonetically-informed phonological rule which specifies the height of the epenthetic vowel. There was also a strong correlation between the observed and predicted vowel distributions. This result supports the claim that listeners access lexical statistics and apply them granularly. Here, I would like to note that the results of this study cannot necessarily be generalized to perceptual epenthesis. As noted by the two aforementioned Mattingley et al. studies, results can be affected by the type of experimental task. The granular application of lexical statistics seems to happen in loanword adaptation, or at least, tasks that involve vowel identification and/or transcription, but whether listeners apply these same patterns in the same manner in real-time perception remains unseen.

References

Alexandra Kuznetsova, Per Bruun Brockhoff and Rune Haubo Bojesen Christensen (2016). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-33. https://CRAN.R project.org/package=lmerTest

Beckman, Jill. 1998. Positional faithfulness. PhD dissertation, University of Massachusetts, Amherst.

Davidson, Lisa & Jason Shaw. 2012. Sources of illusion in consonant cluster perception. *Journal of Phonetics*, 40, 234–248.

de Lacy, Paul. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge Studies in Linguistics 112. Cambridge University Press.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Cristophe Pallier, & Jaques Mehler. 1999. Epenthetic Vowels in Japanese: A Perceptual Illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568-1578.

Dupoux, Emmanuel, Cristophe Pallier, Kazuhiko Kakehi, & Jacques Mehler. 2001. New Evidence for Prelexical Phonological Processing in Word Recognition. *Language and Cognitive Processes* 16, 491-505.

Dupoux, Emmanuel, Erika Parlato, Sonio Frota, Yuki Hirose, & Sharon Peperkamp. 2011. Where do illusory vowels come from? *Journal of Memory and Language* 64, 199–210.

Durvasula, Karthik. & J. Kahng. 2015. Illusory vowels in perceptual epenthesis: the role of phonological alternations. *Phonology* 32.3, 385-416.

Hall, Nancy. 2011. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.) *The Blackwell companion to phonology*. Malden, MA & Oxford: Wiley-Blackwell. pp.~1576--1596.

Itô, Junko. 1989. A Prosodic Theory of Epenthesis. *Natural Language and Linguistic Theory* 7, 217-259.

Itô, Junko & Armin Mester. 1995. Japanese phonology. In John A. Goldsmith (ed.) *The handbook of phonological theory*, 817–838. Cambridge, MA & Oxford: Blackwell.

Ito, Junko & Armin Mester. 2006. Systemic markedness and faithfulness. In J. Cihlar, A. Franklin, D. Kaiser and I. Kimbara, eds. *Proceedings of the 39th Meeting of the Chicago Linguistics Society*. Chicago, Department of Linguistics, University of Chicago. 665-689.

Kabak, Baris & William J. Idsardi. 2007. Perceptual Distortions in the Adaptation of English Consonant Clusters: Syllable Structure or Consonantal Constraints? *Language and Speech* 50, 23-52.

Kawahara, Shigeto & Jason Shaw. 2016. Effects of vowel informativity on vowel duration in Japanese. Submitted.

Kitto, Catherine & Paul de Lacy. 1999. Correspondence theory and epenthetic quality. In Catherine Kitto and Carolyn Smallwood (eds.) *Proceedings of AFLA (Austronesian Formal Linguistics Association) VI*. Toronto: Toronto Working Papers in Linguistics, pp.181-200.ROA

Monahan Philip J, Eri Takahashi, C. Nakao, & Idsardi William J. 2009. Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. In *Japanese/Korean Linguistics*, Vol. 17. eds. Shoishi Iwasaki, Hajime Hoji, Patricia M Clancy & Sung-Ock Sohn. 391-405. Stanford, CA: CSLI.

Peperkamp, Sharon & Emmanuel Dupoux. 2003. Reinterpreting loanword adaptations: the role of perception. *Proceedings of ICPhS* 15, 367-370.

Peperkamp, Sharon. 2005. A psycholinguistic theory of loanword adaptations. In: M. Ettlinger, N. Fleischer & M. Park-Doob (eds.) *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: The Society, 341-352.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: a perceptual account. in Elizabeth Hume and Keith Johnson (eds.) *Perception in Phonology*, Academic Press 219-250.

Steriade, Donca. 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.

Wickham, Hadley. 2009. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag.

Wilson, Colin & Lisa Davidson. 2013. Bayesian analysis of non-native cluster production. S. Kan, C. Moore-Cantwell, R. Staubs (Eds.), *Proceedings of the Northeast linguistics society 40*, Graduate Linguistic Student Association, Amherst, MA (2013), pp. 265–278

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.

Yun, Suyeon. 2012. Perceptual similarity and epenthesis positioning in loan adaptation. In *Chicago Linguistic Society*, Vol. 48.

Appendix

Stimuli used in the experiment

	Block A	Block B
Word-medial		
pt	geptafa	deptam
kt	gektafa	dektam
nt	gentafa	dentam
st	destam	gestafa
tɕt	detɕtam	getɕtafa
dʑd	dedʑdam	gedʑdafa
ɕt	deɕtam	geɕtafa
Word-initial		
pt	ptam	ptafi
kt	ktam	ktafi
nt	ntam	ntafi
st	stafi	stam
tɕt	tɕtafi	tɕtam
dʑd	dʑdafi	dʑdam
ɕt	ɕtafi	ɕtam
Fillers		
pt	potaru	potafu
kt	kotaru	kotafu
nt	notaru	notafu
st	sotaru	sotafu
tɕt	tɕotaru	tɕotafu
dʑd	dʑodaru	dʑodafu
ɕt	ɕotaru	ɕotafu

Vowel Distribution in Responses (Word-medial)

Consonant	Vowel	
	[i]	**[u]**
p	3	20
k	8	27
n	0	2
s	1	31
tɕ	29	6
dʑ	27	7
ɕ	22	14

Vowel Distribution in Corpus (Word-initial)

Consonant	Vowel	
	[i]	**[u]**
p	5	16
k	10	17
n	2	5
s	1	34
tɕ	30	5
dʑ	24	11
ɕ	17	17