The Perception of Epenthetic Vowels in Voiced and Voiceless Contexts in Japanese

By

Hsieh, Chih-Hsiang

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

_____

Co-Chairperson Joan Sereno

_____

Co-Chairperson Robert Fiorentino

_____

Allard Jongman

Date Defended: June 7th, 2013

The Thesis Committee for Hsieh, Chih-Hsiang

certifies that this is the approved version of the following thesis:

The Perception of Epenthetic Vowels in Voiced and Voiceless Contexts in Japanese

_____

Co-Chairperson Joan Sereno

_____

Co-Chairperson Robert Fiorentino

Date approved: June 7th, 2013

ABSTRACT

Dupoux et al. (1999) showed that Japanese native speakers perceptually inserted a vowel /u/ between consonant clusters when listening to nonce words containing consonant clusters. Our study aimed to examine the effect of voicing type of these clusters on perceptual vowel epenthesis and replicate the results of Dupoux et al. (1999) between English and Japanese. We constructed nonce words containing both voiced and voiceless consonant clusters and the stimuli were recorded by a female English native speaker. 10 English native speakers and 10 Japanese native speakers were recruited and asked to judge whether they heard a /u/ in the stimuli. The results showed that voicing does not have an effect on perceptual vowel epenthesis in Japanese and both Japanese and English participants in our study behaved similarly to their Japanese and French counterparts in Dupoux et al. (1999).

Acknowledgements

I would like to express my deepest appreciation to my thesis committee members, Dr. Sereno, Dr. Fiorentino and Dr. Jongman. I also would like to express my gratitude to all the friends who have helped and supported me.

TABLE OF CONTENTS

The Perception of Epenthetic Vowels in Voiced and Voiceless Contexts in Japanese

Introduction

The interest in understanding how humans understand spoken language has stimulated a wide array of studies on the relationship between the acoustic properties of speech and speech perception. What we think we hear depends on both the acoustic properties of each sound in the speech signal and how these sounds are combined. Massaro and Cohen (1983) showed that in addition to the acoustic properties, what listeners perceive is also influenced by the rules that govern the sound combination. Dupoux et al. (1999) further showed that speech perception depends on the unique rules that govern the sound combination of a language by studying Japanese and French listeners' perception of the same stimuli. While French listeners could perceive the stimuli as they physically were, Japanese listeners' perceived an extra vowel that had no acoustic correlates in the stimuli. This perceptual vowel epenthesis in Japanese proves to be a useful test case for studying how the rules governing the possible sound sequences shape our perception of speech. The focus of the current study is to further our understanding of how the rules governing the combination of speech sounds in a language interact with other rules of speech sounds to shape what we hear.

The present research investigates the possible interaction between the perception of epenthetic vowels in the context of illicit consonant clusters and vowel devoicing in Japanese. This study is inspired by the work of Dupoux et al. (1999). The Dupoux et al. study showed that when stimuli containing consonant clusters were presented to Japanese natives, they report hearing /u/ when the acoustic correlates of /u/ did not even present themselves in the stimuli. Their ABX experiments also showed that Japanese native speakers could not distinguish CC and

C[u]C clusters. This epenthetic vowel perception was viewed as a mechanism to resolve phonotactically illicit consonant clusters in Japanese, rendering illegal clusters legal by perceptually inserting a vowel in between them.

However, is phonotactic violation the only factor behind perceptual epenthesis? One of the possible contributing factors is the vowel devoicing rule. First, In Japanese, the high vowels /u/ and /i/ are commonly devoiced between voiceless obstruents. The vowel /u/ is thus the target of both the devoicing rule and the epenthesis rule. Secondly, the devoicing rule in Japanese creates a context that is similar to the context where perceptual vowel epenthesis is attested. It is thus natural to wonder if voicing type of the consonant clusters plays a role in the perceptual vowel epenthesis in Japanese. The primary goal of the present study is to extend Dupoux et al. (1999) and investigate the role of vowel devoicing in perceptual vowel epenthesis in Japanese.

Literature Review

English and Japanese phonotactics

Within a syllable, only 6 possible vowel-consonant combinations are allowed in Japanese. These are V, VV, CV, CVV, CVN and CVQ (where N and Q stand for nasal and the first half of a geminate, respectively). Consonant clusters such as /bz/ and /st/ are illicit in Japanese. On the contrary, English allows a wide range of vowel-consonant clusters in a syllable and places little restriction on possible vowel-consonant combinations. As far as word-medial consonant clusters like /bz/ in /ebzo/ are concerned, as long as a valid coda is followed by a valid onset, there is no constraint on coda-onset contacts (Lamontagne 1993).

Phonotactics and speech perception

In addition to phonemic differences across languages, how phonemes are combined also influences our speech perception to a great degree. Massaro and Cohen (1983) investigated English natives' perception of a 7-step /l/-/r/ continuum that varied in F3. These stimuli were preceded by different consonants and followed by /i/ and thus formed both legal and illegal consonant-vowel clusters in English. When the stimuli were preceded by /t/ (/tl/ is illicit in English), the participants reported hearing more/r/ (as /tri/) across stimuli; when the stimuli were preceded by /s/ (/sr/ is illicit in English), the participants reported hearing more /l/ as /sli/ across stimuli. This suggests that even when stimuli entirely consist of phonemes from native phoneme categories, perception may vary. Perception is influenced by how the phonemes in the stimuli are combined in English. However, since the only language tested was English, this study could not tease apart language-specific constraints on phoneme combinations from universal preferences for certain phoneme combinations. In other words, human language in general may prefer /tr/ and /sl/ over /tl/ and /sr/ respectively.

Many production errors made by second language learners support the idea that there is a language-specific constraint on possible phoneme combinations. For example, in Spanish, /sC/ (C=consonant) clusters always come after a vowel. When Spanish learners of English pronounce words like "special" in English, they tend to add the vowel /e/ before /sp/. When Japanese learners of English pronounce words containing consonant clusters in English, they tend to add vowels between consonants. For instance, the English word text /tekst/ is pronounced as /tekisuto/ in Japanese as a loan word.

In order to tease apart language-specific phontactic constraints from universal preferences for certain clusters, Dupoux et al. (1999) conducted a cross-linguistic study on perceptual vowel epenthesis, comparing Japanese and French listeners. While Japanese only allows consonant

clusters that consist of a nasal and a consonant or geminates, French allows a wide range of consonant clusters. Nonce stimuli in the form of VCCV, VC[u]CV, and VC[u]CV whose [u] differs in duration ranging from 0 ms to 72ms, were presented to both Japanese and French native speakers. The participants were asked to judge whether there was a [u] in the stimuli. The results show that while Japanese natives reported significantly more [u] and the presence of [u] even in the stimuli without [u], French natives only consistently reported the presence of [u] when the duration of [u] was longer than 38 milliseconds. The results not only further confirmed the claim that speech perception is altered by the legality of stimuli in terms of possible sound combinations in the language, they also showed that phonotactic constraints can be specific to a language rather than being universal.

In addition, further experiments in this study showed that coarticulatory cues in the presented stimuli do not alter the nature of perceptual vowel epenthesis in Japanese. First, if coarticulatory cues play a language-specific role, the stimuli recorded by French and Japanese native speakers should contain different cues and thus affect perception. The results showed that the stimuli recorded by both French and Japanese native speakers elicited equally significant amount of perceptual vowel epenthesis in Japanese. Change of speaker did not have an effect. Second, if the perceptual vowel epenthesis depends on the coarticulatory cues in the consonant preceding [u], then pure consonant clusters should elicit less vowel epenthesis. The results showed that digital consonant clusters whose acoustic correlates of [u] were completely removed and natural consonant clusters elicited equal amounts of /u/ perception.

Dehaene-Lambertz et al. (2000) conducted an event-related potential (ERP) study to look for neural correlates of perceptual vowel epenthesis in Japanese based on the property of an auditory response in the brain called “mismatch response”. Mismatch responses are elicited

when a stimulus that has been presented over a period of time suddenly changes to a different stimulus. In this study, 12 Japanese and 12 French native speakers were recruited. They were told to judge whether the test item (e.g. igmo or igumo) was different from the precursor item (e.g. igumo) while their ERPs were measured. While French native speakers had three mismatch responses when the auditory stimulus changed from [igumo] to [igmo], Japanese native speakers only showed weaker versions of the 2nd and the 3rd response. The three mismatch responses the French speakers had indicates that they were able to perceive the difference between [igumo] and [igmo]. The absence of the first response in Japanese participants indicates that Japanese native speakers could not discriminate [igumo] from [igmo]. This means that phonotactic constraint affects the perception at a very early stage. The presence of weaker versions of the 2nd and the 3rd response indicates that Japanese listeners may consciously integrate various cues to perceive the difference between [igumo] and [igmo]. The results provide a piece of neurolinguistic evidence to the claim that Japanese native speakers perceptually epenthesize a vowel to resolve consonant clusters.

Although Dupoux et al. (1999) concluded that the perceptual vowel epenthesis is a solution for solving phonotactic illegality in Japanese, other studies suggested that this perceptual vowel epenthesis is the result of a complicated interaction between multiple factors. (Kabak and Idsardi 2007; Chang et al. 2007)

Kabak and Idsardi (2007) suggest that Dupoux et al. (1999) did not tease apart two kinds of phonotactic illegality. The first kind is that the illegality is due to illegal consonant contact, in which the consonants of concern are never in adjacent position in the native language. The other kind is that the illegality is due to the violation of syllable structure. Since all the stimuli in the study of Dupoux et al. (1999) violate both consonant contact rules and syllable structure in

Japanese, Kabak and Idsardi (2007) conducted another cross-linguistic study between Korean and English to tease apart the two kinds of illegality. On one hand, /k/ is a permitted coda while /c/ is never a permitted coda in Korean. On the other hand, both /km/ and /cm/ are not permitted consonant clusters. It has been shown that Korean native speakers resort to perceptual vowel epenthesis to resolve licit consonant clusters. If it is the violation of consonant contact rules that induces perceptual epenthesis, then perceptual epenthesis will be observed in the stimuli that contain both /km/ and /cm/. In contrast, if perceptual epenthesis is due to the violation of syllable structure, then the perceptual epenthesis will only be observed in the case of /cm/. Kabak and Idsardi (2007) showed that perceptual epenthesis does not occur in the case of /km/, which supports the syllable violation account.

Chang et al. (2007) conducted a perception study similar to that of Dupoux et al. (1999) on native speakers of Taiwan Mandarin (TM). Since the Mandarin syllable does not allow consonant clusters, it is similar to that of Japanese. They constructed 18 pairs of CəCV disyllabic and CCV monosyllabic English words. These stimuli were recorded by an American phonetician. Taiwan Mandarin learners of English were recruited and their ability to distinguish CəCV words from CCV words was tested by an AXB discrimination test. The results showed that Taiwan Mandarin speakers were able to correctly discriminate CəCV and CCV contrasts most of the time in spite of the syllable structure constraint of Mandarin. Only 6%-20% of CC clusters were perceived as CəC. This suggests that the vowel epenthesis is not the main solution to resolve phonotactic illegality and phonotactic constraints alone are not decisive in TM speakers' perception of non-native consonant clusters. They further argue that the perceptual vowel epenthesis found in Dupoux et al. (1999) was largely due to the Japanese writing system, in which each "kana" is a syllable consisting of a consonant and a vowel.

Davidson and Shaw (2011) conducted a perception study on English native speakers to examine the relationship between the manner of articulation of unattested consonant clusters in English and the types of perceptual alterations, including epenthesis, deletion, consonant change and prosthesis (an addition of a phoneme at the beginning of a word). English native speakers were recruited and their ability to discriminate non-native word-initial clusters and modified counterparts of the clusters was investigated through AX and ABX discrimination tasks (e.g. [zmatu]-[zəmatu] as epenthesis, [matu] as deletion, [smatu] as consonant change, or [əzmatu] as prosthesis). The results show that fricative-initial sequences tend to elicit prosthesis, stop-nasal sequences tend to invoke deletion or change of the first consonant and stop-stop sequences elicit vowel insertion. This means that phonotactically illegal clusters elicit different types of perceptual modifications based on their acoustic properties. The results suggest that perception of phonotactically illegal sequences is affected by factors other than the phonotactic constraints of a language, including language-specific phonetic knowledge and the acoustic similarity of the stimulus items. Phonotactic constraints only affect perception of unattested clusters indirectly through the language-specific interpretation of acoustic cues.

Although the above studies were conducted in languages other than Japanese, the results clearly suggested that phonotactic constraints are not the sole decisive factor in the perception of unattested clusters in one's native language. This motivates us to reexamine the factors that may contribute to the perceptual vowel epenthesis in Japanese.

Vowel Devoicing in Japanese

In Japanese, the high vowels /i/ and /u/ often undergo devoicing when immediately flanked by voiceless obstruents in an unaccented syllable (Vance, 1987, Cutler et al., 2008),

especially in the Tokyo dialect. Some authors even report observing complete vowel devoicing in this context (Tsuchida, 1997). For example, the first /u/ in /kusuri/ (medicine) is almost always devoiced but Japanese native speakers think they hear a /u/ between /k/ and /s/ if asked. The first /u/ is present in the native speakers' perception even though there are no acoustic correlates.

Imai (2005) conducted a sociolinguistic study on both linguistic and social factors that influence the occurrence of vowel devoicing. She recruited 42 participants from different age and sex groups and asked them to read a word list and a passage with words containing different devoicing environments. The frequency of devoicing in different contexts was calculated and compared. The identity of the voiceless obstruents flanking devoiceable vowels plays an important role. Imai showed that voiceless fricatives as a preceding consonant promote devoicing the best; voiceless stops as a following consonant promote devoicing the best. In terms of vowel identity, in addition to the general rule that the high vowels /i/ and /u/ are devoiced between voiceless obstruents, /i/ is more frequently devoiced. Devoicing is favored when the vowel is morpheme-internal. When there are other vowels in devoicing environments preceding the vowel in question, devoicing is not preferred. As for social factors, speech style plays the most significant role in vowel devoicing in Japanese. More devoicing was found in casual speech. In terms of age and sex, young males devoice the most while young females avoid devoicing.

Vowel Devoicing and Vowel Epenthesis in Japanese

Monahan et al. (2008) examined the role of vowel devoicing in perceptual vowel epenthesis in Japanese. They constructed three types of stimuli [etmo], [etomo] and [etumo] and conducted an AX discrimination experiment in Japanese native speakers. The results showed that

Japanese native speakers could readily discriminate [etmo] from both [etomo] and [etumo]. In other words, neither [u] nor [o] epenthesis occurred. This suggests that [u] epenthesis does not always occur when Japanese phonotactic constraint is violated. [u] epenthesis only occurs in the contexts where devoiceable vowels can be inserted. This suggests that the target vowel /u/ of the vowel devoicing rule plays a role in perceptual vowel epenthesis in Japanese. However, [etmo] does not constitute a vowel devoicing environment. Is it possible that the absence of epenthesis is due to the absence of a vowel devoicing context? Since in Dupoux's study (1999) [u] epenthesis was invoked in non vowel devoicing contexts, this possibility can be excluded.

Dupoux et al. (2011) conducted a study similar to Dupoux et al. (1999) to examine the relationship between phoneme categorization and perceptual vowel epenthesis in Japanese, Brazilian Portuguese and European Portuguese. These languages do not allow consonant clusters in their underlying representations of words but have vowel devoicing rules (Japanese and Brazilian Portuguese) and vowel deletion rules (European Portuguese). They constructed three types of nonce words stimuli, VCCV, VC/u/CV and VC/i/CV and these were recorded by a male native speaker of French. From VC/u/CV and VC/i/CV, six new stimuli were created (from each original item) by splicing out pitch periods of the vowel [u] and [i] at zero crossings. Participants were forced to choose which vowel ([i], [u], or no vowel) they hear in the middle (VC?CV) of each nonce word. The results indicated that while Japanese and Brazilian Portuguese show [u] and [i] epenthesis respectively, European Portuguese shows no epenthesis when there is no coarticulatory cue (VCCV, natural cluster). When there are coarticulatory cues, both Japanese and Brazilian Portuguese's responses vary according to the cues. For example, [i]-coarticulation has a strong effect on Japanese listeners, changing their predominant responses from /u/ to /i/. In regard to our interest in perceptual vowel epenthesis in Japanese, these results give us three

suggestions. First, the predominant epenthetic vowels are the phonetically minimal vowels in both Japanese and Brazilian Portuguese. [u] is inserted because it is shortest in Japanese; [i] is inserted because it is the shortest in Brazilian Portuguese. Second, perceptual epenthesis is not purely a repair mechanism resulting from phonotactic violation. Phonetic factors such as coarticulatory cues are involved. Finally, the finding that European Portuguese speakers can perceive consonant clusters suggests that phonotactic probabilities are computed based on the surface distributions of speech sounds. Given that some authors like Tsuchida (1997) reported observing complete vowel deletion in Japanese, one may expect that vowel epenthesis is absent or less strong in vowel devoicing environments.

Dupoux et al. (2001) conducted a study to test whether the perceptual vowel epenthesis in Japanese is due to a "top-down" lexical effect. The C_1C_2 in the stimuli in Dupoux et al. (1999) may activate lexical items that contain C_1/u/C_2 and elicit the perception of [u] due to the potential existence of real Japanese words and phonetic neighbours of the nonword stimuli. To test this, they manipulated the lexical neighborhood of the nonwords. They constructed non words that contain illegal consonant clusters. The "u-set" is defined as the nonwords whose [u] epenthetic counterparts are real words in Japanese; the "non u-set" is defined as the nonwords whose [u] epenthetic counterparts are not real words in Japanese. They presented these stimuli to Japanese native speakers and asked them to transcribe the stimuli and judge whether it is a word or not. If there is a lexical effect, then the non words whose epenthetic counterparts are real words in Japanese may elicit more perceptual vowel epenthesis in Japanese native speakers. The results showed that both the "u-set" and the "non u-set" elicited equal amounts of [u] perception. Although the results suggest that there is no lexical effect in perceptual vowel epenthesis, the authors suspect that there might be a strong lexical effect only in the stimuli that have both C_1C_2

as voiceless consonants because of the vowel devoicing rule in Japanese. For example, /kokuso/, a real word in Japanese, may be stored as both /kokso/ and /kokuso/ in the Japanese lexicon. Their post-hoc analysis comparing items containing voiceless obstruents to other items, however, showed that items containing voiceless obstruents did not yield more perceptual vowel epenthesis.

The three studies (Dupoux et al. 2011; Monahan et al. 2008; Dupoux et al. 2001) suggested three different roles that vowel devoicing may play in the perceptual vowel epenthesis. Vowel devoicing environments may have no effect on perceptual epenthesis, discourage perceptual vowel epenthesis or promote perceptual vowel epenthesis. However, none of these studies aimed to directly examine the effect of vowel devoicing on the perceptual vowel epenthesis in Japanese.

The present study

The primary aim of this study is to investigate the possible contribution of vowel devoicing to the perception of epenthetic vowels in Japanese. The possible influence of vowel devoicing in vowel epenthesis is discussed in Dupoux et al. (2001), Monahan et al. (2008) and Dupoux et al. (2011). Although Dupoux et al. (2001) conducted a post-hoc study and showed that devoicing contexts did not promote more vowel epenthesis than non-devoicing contexts, they did not systematically contrast these two environments by controlling factors such as dialect and age and social class (Imai 2005). Moreover, Monahan et al. (2008) and Dupoux et al. (2011) only made suggestions about the role of vowel devoicing contexts. Overall, no study has specifically investigated possible interactions between Japanese vowel devoicing and perceptual vowel epenthesis.

Based on Dupoux et al. (2001), a given consonant cluster C_1C_2 may activate lexical items containing $C_1/u/C_2$ due to its phonetic proximity and the lexical items in turn invoke the perception of /u/ between C_1C_2. If we assume that words in which vowel devoicing occurs are stored as both C_1C_2 and $C_1/u/C_2$ in the Japanese lexicon, then it is plausible to say that when C_1 and C_2 are both voiceless, it can directly activate the lexical items that contain $C_1/u/C_2$. The lexical effect thus activated is expected to be stronger than for the C_1C_2 that are both voiced. Our hypothesis is that vowel devoicing environments may exert a lexical effect on the perception of phonotactic illegal consonant clusters in Japanese by promoting more perceptual vowel epenthesis than non vowel devoicing environments.

In Dupoux et al. (1999), the stimuli containing phonotactically illegal consonant clusters like [egdo] elicited perceptual vowel epenthesis in Japanese native speakers. Our study will contrast items like [egdo] with stimuli containing vowel devoicing environments such as [ekto].Both [egdo] and [ekto] will elicit perceptual vowel epenthesis in Japanese because they are phonotactically illegal, but we expect that [ekto] will elicit more perceptual vowel epenthesis than [egdo] because of the following logic. Words like /kutabireta/ ("worn-out") may be stored as both /ktabireta/ and /kutabireta/ in the Japanese lexicon because of the vowel devoicing applied to /u/ between the two voiceless obstruents. [kt] can thus directly activate the lexical items like /ktabireta/ and the lexical items in turn invoke the perception of /u/. Unlike [kt], [gd] cannot directly activate the lexical items like /kutabireta/ because [gd] is not stored as part of a word in the Japanese lexicon. If it does activate any lexical item, the activation should be less strong and promote less perceptual vowel epenthesis.

Our objective is two-fold. First we aim to replicate the results of the French natives in Dupoux et al. 1999 with English natives and thus establish a paradigm for English-Japanese

cross-linguistic studies along this line of research. Secondly, we will test the contribution of vowel devoicing contexts to perceptual vowel epenthesis. Our hypothesis is that, if vowel devoicing contexts have a lexical effect, the number of epenthetic vowels perceived will be higher for nonce clusters like [ekto] and lower for items like [egdo].

Design

Our experiment basically follows Experiment 1 of Dupoux et al. (1999) with some minor adjustments.

1. Participants

Ten native English and ten native Japanese speakers were recruited at the University of Kansas. The English participants were all college students and their age ranged from 20-25 years old. They all had some knowledge in one or more foreign languages but none of them had any knowledge of the Japanese language. The Japanese subjects were all college students and their age ranged from 20-25 years old. Seven participants spoke the Tokyo-dialect as their first language. The remaining three participants spoke Mie, Kyusyu and Okayama dialects, respectively, as their first language. All the Japanese participants spoke English as a foreign language. Their length of residence in the U.S. ranged from less than 6 months (8) to 5 years (2). They were all able to communicate in English. Based on our impression and their self-evaluation of their English proficiency, none of them had a native-like or nearly native proficiency in English.

2. Stimuli

The stimuli of this study consist of nonce words in the form of $V_1C_1[u]C_2V_2$ and $V_1C_1C_2V_2$. V1 and V2 are either [e], [i], [a] or [o]. 6 pairs of stimuli were created by choosing different C_1 and C_2. C1 is either a bilabial stop [p] or a velar stop [k]; C2 is a fricative or any of the stops [p, t, k]. Each pair of C_1C_2 differs only in voicing type. For example, the pair [egudo] - [egdo] and [ekuto] - [ekto] differ only in the voicing of the stops. To prevent that the Japanese always report hearing /u/, 50 filler items that only contain the vowels /a/, /i/, /e/ and /o/ were created. The filler items contain 2 to 4 syllables in the form of CVCV, CVCVCV and (C)VCVCVCV. The full list of our stimuli and fillers is shown in Appendix I.

The stimuli were recorded by a female native English speaker with training in phonetics. She was instructed to read [egdo], [egudo], [ekto] and [ekuto] written in IPA and put the accent on the first syllable. The original stimulus (e.g. [egudo] and [ekuto]) in addition to five new files were created (from each original item) by splicing out pitch periods of the vowel [u] at zero crossings. Stimulus 0 is the baseline condition, containing either natural or digital consonant clusters. Stimulus 1 was created by removing all the pitch periods of [u] after the burst. Stimulus 2 contains the two most extreme pitch periods of the vowel (vowel transitions at both consonant ends). Stimulus 3 contains the 4 most extreme pitch periods (two on each side). Stimuli 4 and 5 contain 6 and 8 pitch periods, respectively. Stimulus 6 is the original stimulus. For each voiced-voiceless pair of $V_1C_1[u]C_2V_2$ there are two 7-step continua ranging from $V_1C_1C_2V_2$ to $V_1C_1[u]C_2V_2$. We thus have a total of 84 items [6 (pairs) x 2 (voicing) x 7 (vowel length)]

3. Recording

The female speaker was recorded in an anechoic chamber, using a microphone (Electrovoice, model 767) and a solid state recorder (Marantz PMD-671). The sampling rate was 22050 Hz.

The speaker was instructed to say each $V_1C_1[u]C_2V_2$ or $V_1C_1C_2V_2$ (e.g., [egudo]-[egdo] and [ekuto]-[ekto]) five times respectively. There was no fixed interval between tokens or between the two types of stimuli. The speaker self-paced her speech.

4. Perception experiment procedure

The perception experiment was conducted using Paradigm on a desktop PC in The University of Kansas Phonetics and Psycholinguistics Laboratory (KUPPL). Before the experiment, participants filled out a consent form and a questionnaire about their linguistic background. Participants were asked to listen to the stimuli through headphones and judge whether the vowel [u] was present by clicking either "yes" or "no". They were instructed that this [u] does not have to be exactly like the Japanese [u] in *udon* (noodle) (for Japanese participants) or English [u] in *boo* (for English participants). Any [u]-like sound will be considered [u] in this experiment. They were also told that they would hear a list of words from a foreign language. Before the actual experiment, they were given a practice session with real English words containing /u/ or no /u/. Each participant heard each item three times. A total of 252 trials [6 (pairs) x 2 (voicing) x 7 (six created continuum stimuli and one natural cluster) x 3 (repetitions)] were presented to the participants in a randomized fashion with an ISI of 3s.

Results

Statistical Analysis including all participants

The mean percentages of vowel responses elicited by both voicing types combined as a function of language and vowel length are shown in Figure 1. This function resembles that of

Dupoux et al. (1999). We performed an analysis of variance on percentages of vowel responses with participants as random variables; we performed the same analysis within each language group. Language was a between-subjects factor and both Vowel Duration and Voicing Type were within-subjects factors (Language x vowel duration (7) x voicing (2)). We also performed paired *t*-tests within each language group to compare the mean percentages of vowel responses for each vowel duration condition.

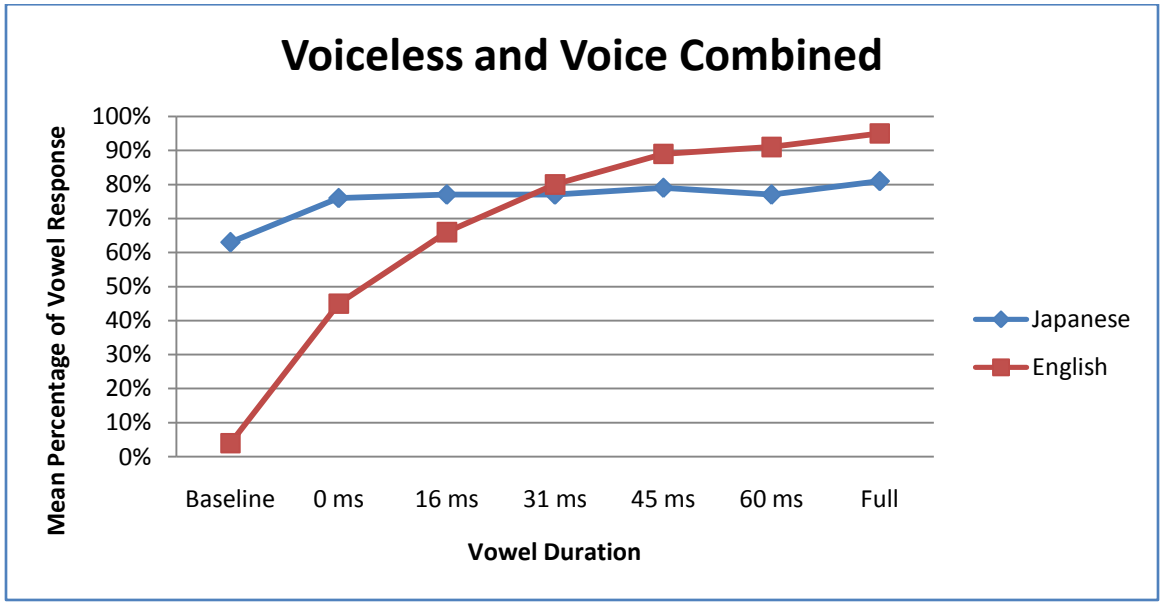

Figure 1. Percent [u] vowel judgments for English and Japanese participants as a function of vowel duration.

Overall, language does not have a significant effect on vowel responses [F (1, 18) = .90 p=.354]. English participants gave 67% /u/ responses and Japanese participants gave 76% /u/ responses. There was a significant Vowel Duration effect [F (6, 13) = 35.40 p<.001]. This suggests that the longer the vowel duration is, the more /u/ responses are elicited. There was an interaction between Language and Vowel Duration [F (6, 13) = 16.60 p<.001]. This suggests that although more /u/ responses are elicited by longer vowels, this effect is different in Japanese and

English. There was also an interaction between Voicing and Vowel Duration [F (6, 13) = 6.51 p=.002]. This suggests that although more /u/ responses are elicited by longer vowels, this effect varies according to the voicing type, voiceless or voiced. The interaction between Voicing and Vowel Duration and Language is also significant [F (6, 13) = 5.57 p=.005). The significant interaction suggests that that Voicing and Vowel Duration together may exert different effects on each language group. In other words, Japanese and English behaved differently from each other.

The mean percentages of vowel responses elicited by both voicing types as a function of language and vowel length are shown in Figure 2 (voiced) and Figure 3 (voiceless). Overall, voiced stimuli elicited 62% /u/ responses in English participants and 74% in Japanese participants. Voiceless stimuli elicited 73% /u/ responses in English and 77% in Japanese.

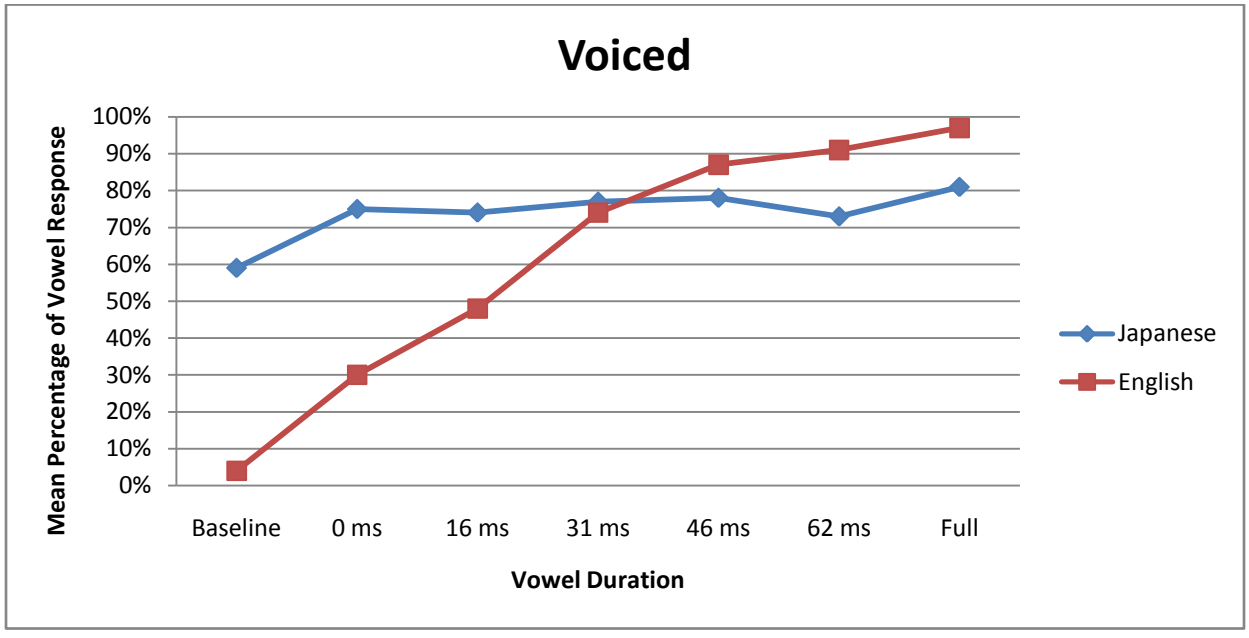

Figure 2. Percent [u] vowel judgments in the context of voiced consonants for English and Japanese participants as a function of vowel duration.

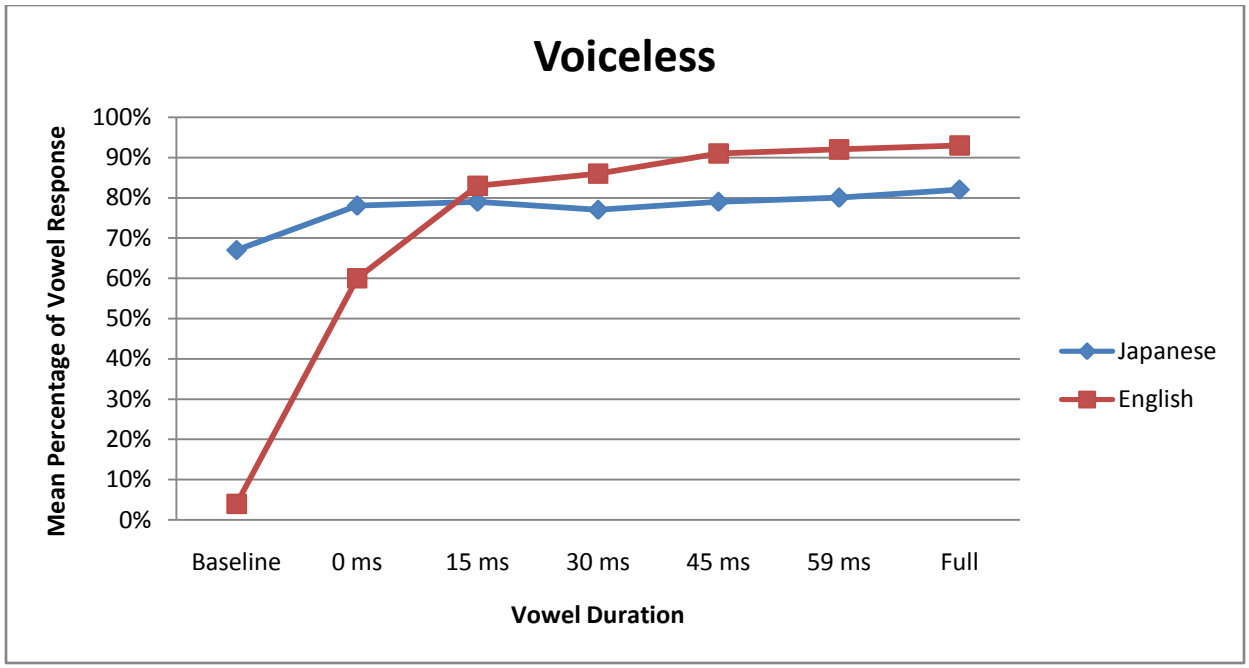

Figure 3. Percent [u] vowel judgments in the context of voiceless consonants for English and Japanese participants as a function of vowel duration.

Figure 4 shows the perception function for both voiceless and voiced stimuli in English participants. Figure 5 shows the perception function for both voiceless and voiced stimuli in Japanese participants.

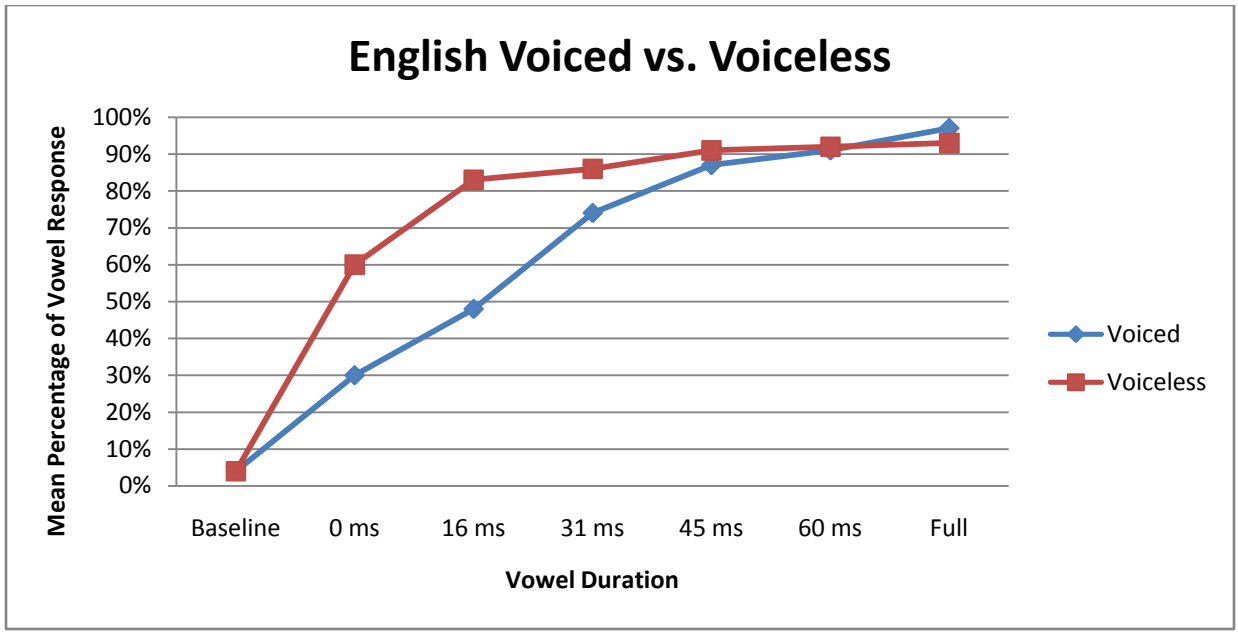

Figure 4. Percent [u] vowel judgments in voiced and voiceless stimuli for English participants as a function of vowel duration.

Within the English group, there was a significant Voicing effect [F (1, 9) = 11.82 p=.007], with the voiceless stimuli eliciting more vowel responses than voiced stimuli. Voiceless stimuli elicited 73% vowel responses, more than the 62% elicited by voiced ones.

There was a significant Vowel Duration effect [F (6, 4) = 47.38 p=.001], with longer vowels eliciting more vowel responses. The interaction between voicing and vowel duration was significant [F (6, 4) = 10.64 p=.019]. Paired t-tests showed that voiceless stimuli elicited more vowel responses (72%) than voiced ones (40%) when vowel duration is 16 ms and 0 ms. Digital clusters (0 ms) elicited more vowel responses (voiced 30%; voiceless 60%) than their respective natural counterparts (baseline: voiced 4%; voiceless 4%).

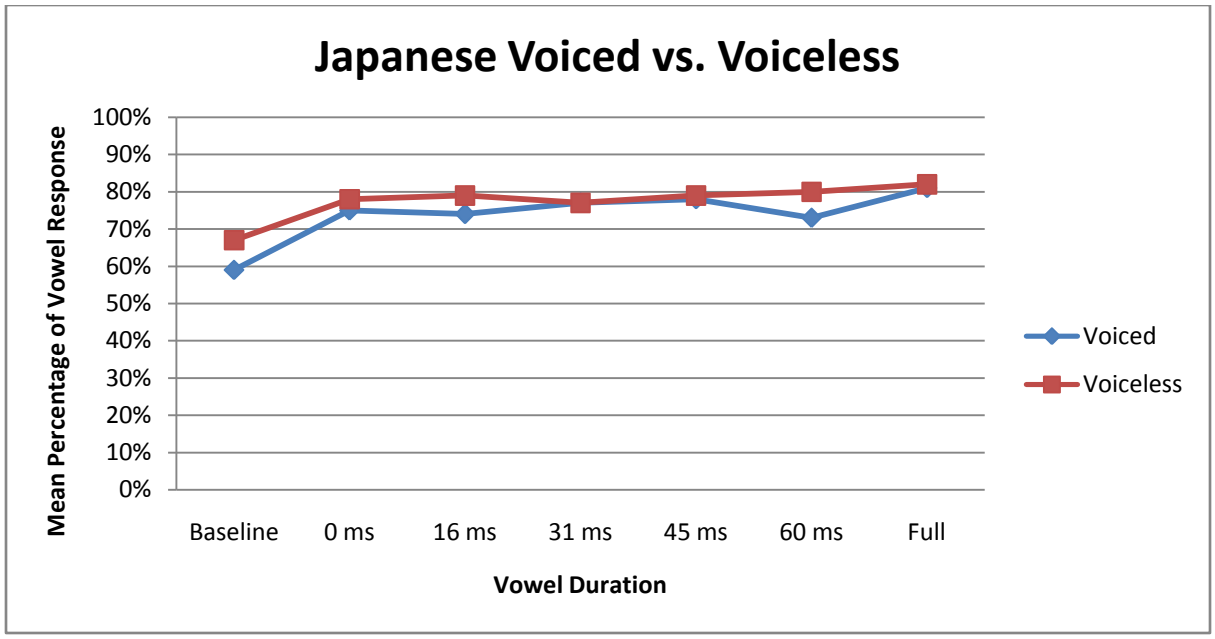

Figure 5. Percent [u] vowel judgments in voiced and voiceless stimuli for Japanese participants as a function of vowel duration.

Within the Japanese group, there was only a marginal Vowel Duration effect [F (6, 4) = 4.98 p=.07]. This suggests that for Japanese listeners, longer vowel duration in general elicits more /u/ responses. Voicing does not have a significant effect on vowel responses [F (1, 9) = .61 p=.455]. Regarding our main hypothesis that voiceless consonant clusters elicit more vowel responses than voiced consonant clusters, a paired t-test shows that there is no significant difference between the two [t(9)= -1.329 p=.217].

Statistical Analysis including only Tokyo participants

We performed the same analysis excluding 3 participants who do not speak Tokyo-dialect as their first language. We thus compared the results of 7 Tokyo-dialect speaking participants against the results of 10 English native speakers. The results of the Tokyo dialect

speaking participants do not differ from the results that include all the Japanese participants. The results are summarized in Figure 6.

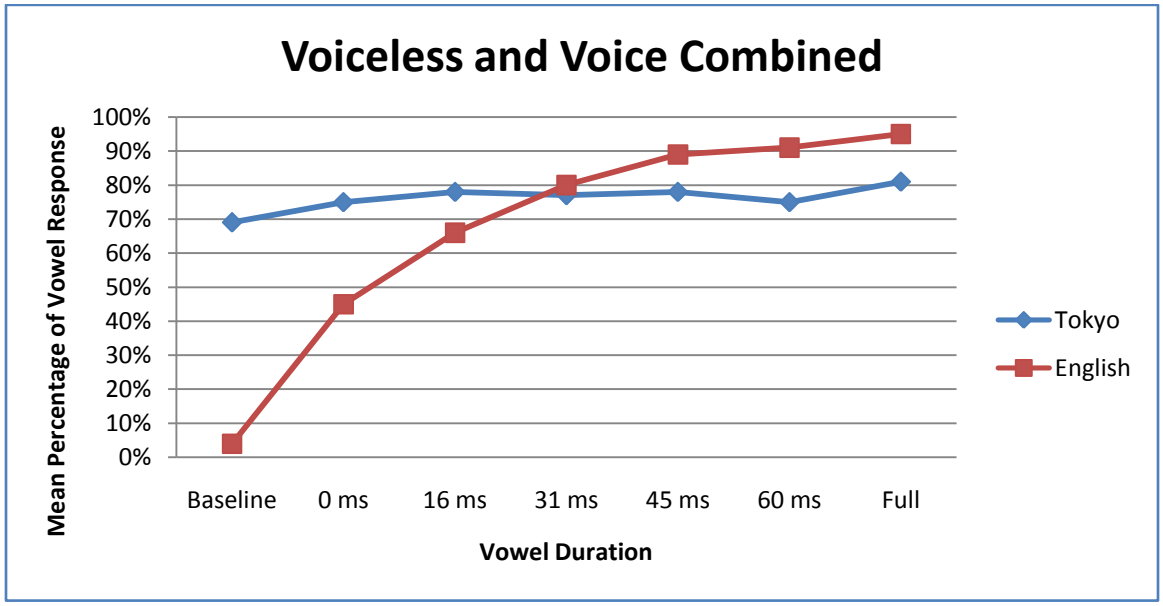

Figure 6. Percent [u] vowel judgments for English and Tokyo-dialect speaking participants as a function of vowel duration.

Overall, language does not have a significant effect on vowel responses [F (1, 15) = .95 p=.354]. English participants gave 67% /u/ responses and Japanese participants gave 76% /u/ responses. There was a significant Vowel Duration effect [F (6, 10) = 24.74 p<.001]. This suggests that the longer the vowel duration is, the more /u/ responses are elicited. There was an interaction between Language and Vowel Duration [F (6, 10) = 11.52 p=001]. This suggests that although more /u/ responses are elicited by longer vowels, vowel duration exerts different effects in both languages. There was an interaction between Voicing and Vowel Duration [F (6, 10) = 6.58 p=.005]. This suggests that although more /u/ responses are elicited by longer vowels, this

effect varies according to the voicing type, voiceless or voiced. The interaction between Voicing and Vowel Duration and Language is significant [F (6, 10) = 4.90 p=.01). The significant interaction suggests that Voicing and Vowel Duration together may exert different effects on each language group. In other words, Tokyo-dialect and English behaved differently from each other.

The results of the Tokyo-dialect speaking participants are summarized in Figure 7. In contrast to a marginal effect obtained in the group including all the Japanese participants, vowel duration does not have any effect on the vowel response [F (6, 1) = 2.83 p=.426] in the Tokyo-dialect speaking participants. Voicing does not have an effect on vowel responses [F (1, 6) = .38 p=.559] Regarding our main hypothesis that natural voiceless consonant clusters elicit more vowel responses than natural voiced consonant clusters, paired t-test shows that there is no significant difference between the two [t(6)= -.66 p=.534].

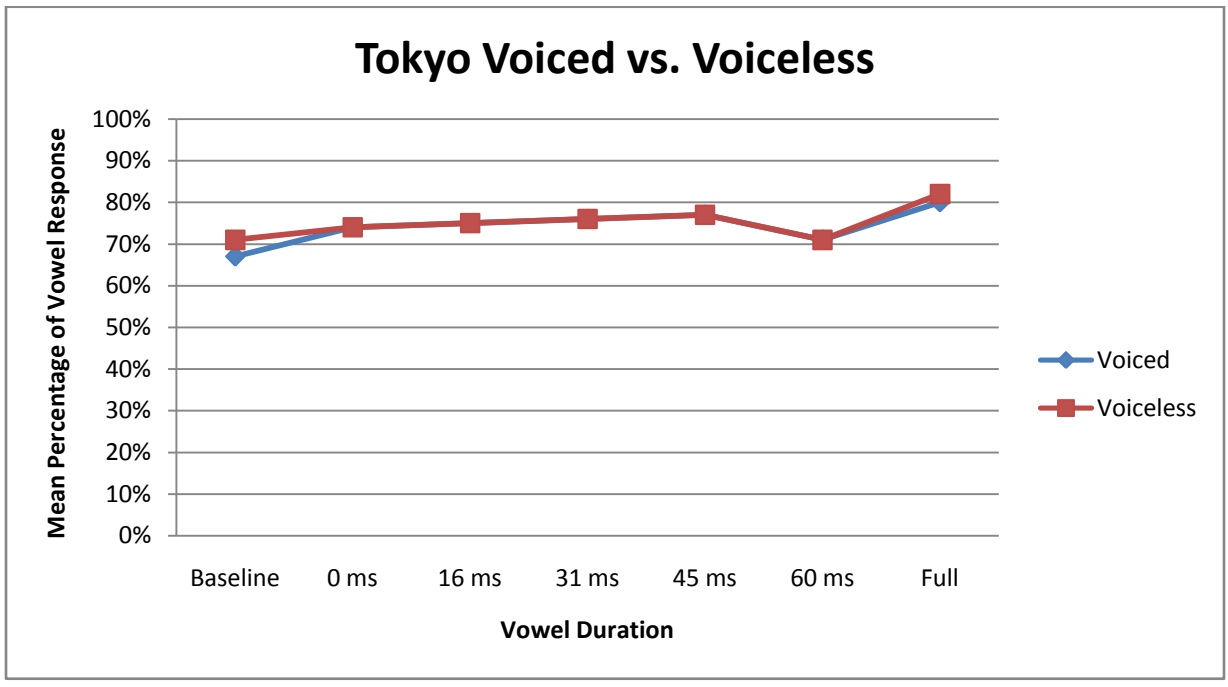

Figure 7. Percent [u] vowel judgments in voiced and voiceless stimuli for Tokyo-dialect speaking participants as a function of vowel duration.

Discussion

The effect of voicing

In this study, our first goal was to replicate and extend the results of Dupoux et al. (1999) looking at Japanese native speakers and comparing them to English native speakers. Secondly, we investigated effects of voicing in perceptual vowel epenthesis.

The main objective of our study was to examine the effect of voicing on perceptual epenthesis in Japanese. We hypothesized that the voiceless consonant clusters would elicit more vowel responses in Japanese participants than voiced consonant clusters because of the naturally occurring vowel devoicing rule in the Japanese language. In other words, voiceless consonant clusters may activate lexical items that contain vowel devoicing environments. These activated lexical items may in turn elicit /u/ perception. However, in our study we did not find this effect in Japanese. The mean percentage of vowel responses for voiceless natural clusters (voiceless baseline) is 67%, slightly higher than that for voiced natural clusters (voiced baseline), 59%, but this difference is not significant. Voicing context does not seem to play a role and vowel devoicing does not contribute to the elicitation of perceptual vowel epenthesis. If we only include the data of the Tokyo-dialect speaking participants, the difference becomes even smaller. The voiceless clusters elicited 71% vowel responses and the voiced clusters elicited 67%. This supports the conclusion of Dupoux's (2001) post-hoc examination of the voiceless items in their study that the devoicing context does not play a role.

The unique organization of the Japanese lexicon suggests a different interpretation of the absence of a lexical effect in our study. In Japanese, phonological rules only apply to a certain lexicon stratum. For example, rules concerning "rendaku" (sequential voicing) only apply to Yamato words (Tsujimura 2007). Although researchers have not reached a consensus about the relation between lexicon stratum and vowel devoicing, it is plausible to assume that if the vowel devoicing rule does not apply to foreign words, vowel devoicing should not affect perceptual vowel epenthesis in our study that only included nonce words.

On the other hand, if we look at the devoicing rule in terms of phonotactic probability as mentioned in Dupoux et al. (2011), our results suggest that complete vowel deletion may not exist in Japanese because our voiceless items did not elicit less perceptual vowel epenthesis. In Dupoux et al. (2011), since European Portuguese participants did not show perceptual vowel epenthesis when listening to consonant clusters, Dupoux et al. (2011) claimed that it is an effect of the complete vowel deletion rule in European Portuguese. Due to this deletion rule, consonant clusters appear at the surface level and European Portuguese speakers are used to consonant clusters. Therefore, phonotactic probability is calculated at the surface level. On top of that, Tsuchida (1997) claimed that complete vowel deletion sometimes happens in Japanese. If complete vowel deletion sometimes happens in Japanese and produces voiceless consonant clusters at the surface level, then Japanese native speakers are expected to be better at perceiving voiceless consonant clusters. In other words, voiceless consonant clusters should invoke less perceptual epenthesis than voiced ones. However, our results do not support this. This suggests that complete vowel deletion may not be present in Japanese. Alternatively, it may be the case that although complete vowel deletion occurs in Japanese, the probability is too low to affect perception.

To sum up, we observed consistent [u] epenthesis in phonotactically illegal clusters in Japanese. However, the voicing type of these illegal clusters did not affect the amount of perceptual epenthesis. Our results support Monahan's (2008) claim that while the target of the devoicing rule (/u/) plays a role in perceptual vowel epenthesis, the environment (flanking voiceless consonants) does not affect this illusory vowel perception.

Replication of Dupoux et al. (1999)

We successfully replicated the pattern found in Dupoux et al. (1999) between English and Japanese participants. Our English participants behaved like French participants in Dupoux's study. Like the French participants, they judged that the vowel was absent in the natural consonant clusters most of the time (More than 95% in both studies). Within the 6-step digital continuum, they gave more vowel responses to the stimuli with longer vowel duration and fewer to the stimuli with shorter vowel duration. Japanese participants judged that the vowel was present at all levels of vowel duration, even in the stimuli without actual acoustic correlates of the vowel.

Our results differ from those of Dupoux et al. (1999) in three ways. First, the perception function of Japanese is flat in our study while it is slightly positively correlated to vowel duration in Dupoux's study. It might be due to the fact that we had 50 fillers. Dupoux et al. (1999) did not use any filler. In addition, since our experiment had about 100 more trials (302) than Dupoux's (210). Japanese native speakers, faced of a larger number of stimuli, may have become insensitive to vowel duration due to the fillers. The difference in performance between our Japanese participants and those of Dupoux might be due to the difference in the number of fillers in both experiments.

Secondly, the cutoff point, defined by Dupoux as the point at which participants judged the vowel to be present in 50% of the cases, for the English participants is less than 1 pitch period (4 ms) in our study, instead of 6 pitch periods (44 ms) in Experiment 2 and 4 pitch periods (38 ms) in Experiment 1 of Dupoux's study. In other words, our digital consonant clusters elicited more /u/ perceptions when vowel duration is short, especially the voiceless clusters. In English, voiceless stimuli had an even stronger effect than voiced ones. Digital consonant clusters (0 ms) elicited already 60% vowel responses compared to 30% of voiced clusters. This "voiceless boost of /u/ perception" for digital consonant clusters must have originated from the allophonic cues in the voiceless consonants preceding the vowel in our voiceless stimuli. Although both voiced consonants and voiceless consonants preceding the vowel /u/ contain coarticulatory cues for /u/, the voiceless consonants may have contained additional allophonic cues that signal to English native speakers that they precede /u/. In English, /p/, /t/ and /k/ are usually aspirated in initial position in full-vowel syllable. Aspiration thus gives the English native speakers a cue that there is a vowel following these stops. For example, in our study, the [k] in the stimulus [ekuto] has a VOT of +50 ms. This aspirated [k] hints English listeners that the [k] precedes a vowel. Moreover, the [k] contains coarticulatory cues of [u]. Both coarticulatory cues and allophonic cues in the voiceless stimuli gave our English native participants the impression that they heard a /u/ when there were no or almost no acoustic correlates of /u/. On the contrary, since [g], the voiced counterpart of [k], only contains coarticulatory cues but no allophonic cues, it does not elicit as many vowel responses as [u] does. In Dupoux's study, all the C_1 and C_2 are voiced except one C_1 ([k]). It is thus difficult to evaluate whether such an effect can be seen in French native speakers under the same experimental conditions.

Thirdly, while the Japanese participants showed more total vowel responses than French participants in Dupoux's study, our Japanese participants did not give significantly more vowel responses than our English participants [F (1, 18) = .90 p=.354]. On one hand, this is because English shows the "voiceless boost" effect mentioned above. On the other, the Japanese does not respond to vowel duration so longer vowels do not elicit more vowel responses in Japanese

Finally, individual differences were found in our study, especially in the Japanese group. Some Japanese participants were able to correctly perceive consonant clusters despite their poor English proficiency and short length of residence. For example, participant DH had a length of residence (LOR) of 0 years and spoke limited English. He gave only 19% vowel responses to the baseline stimuli. On the other hand, some Japanese participants judged /u/ to be present almost all the time despite their high English proficiency and longer length of residence. Participant HY had 5 years of LOR and spoke English fairly well. She gave, however, 75% vowel responses to the baseline stimuli. This suggests that there are some non-linguistic factors involved. Individual differences in sensitivity to acoustic cues may play an important role. For example, at the phonemic level, although the perception is largely categorical, individuals show varied sensitivity to subphonemic features and this sensitivity is related to their sensitivity to acoustic cues. (Kong and Edwards, 2011). In this study, a /ta/-/da/ continuum differing in VOT and *f*0 was constructed. English native speakers were recruited and their sensitivity to both VOT and *f*0 were tested. The results showed that people who were more sensitive to *f*0 change were also sensitive to within-category VOT differences.

Conclusion

Our experiment has shown that voicing context has no effect on perceptual vowel epenthesis in Japanese. Both voiced and voiceless consonant clusters elicited the same amount of /u/ perception in Japanese listeners. Our English native speakers behaved like the French native speakers in Dupoux et al. (1999) in responding to vowel duration. Voicing does have an effect on English participants. Voiceless stimuli elicited more vowel responses than voiced ones in English listeners.

Our results support the claim that the environment of devoicing does not affect the perceptual vowel epenthesis in Japanese (Monahan et al. 2008). Our results do not support the claim that the vowel devoicing environment has a lexical effect on vowel epenthesis. The claim that the devoicing rule in Japanese increases the phonotactic probability of voiceless consonant clusters and thus affects perceptual vowel epenthesis is not supported. The conclusion of Dupoux et al. (1999) that the perceptual vowel epenthesis in Japanese is phonological and is a way to resolve phonotactically illegal sequences is supported.

Our results support the hypothesis that English native speakers' behavior should be similar to that of French native speakers due to the fact that both languages allow complex consonant clusters. This adds one new language to the repertoire of this line of research.

Voicing context did have an effect on vowel responses in English. This was caused by the allophonic cue "aspiration" within the voiceless stimuli. This voicing effect is an extra variable that needs to be considered when designing stimuli for this type of perception study.

Allophonic cues in the voiceless stimuli gave our English participants a "perception boost". This was not observed in the Japanese group. Both voiceless and voiced digital clusters

elicited the same percentage of vowel responses in the Japanese. This was expected because our stimuli were recorded by an English native speaker. Our Japanese native speakers could not access these allophonic cues or they had only limited access to these cues depending on their English proficiency. This raises an interesting question in second language acquisition. Will highly proficienct Japanese L2 learners of English be able to use these cues eventually? Does it come after or before they learn to perceive consonant clusters? Additional studies should be conducted by taking L2 proficiency into consideration. For example, we can recruit L2 Japanese learners of English, test their proficiency and group them into low, intermediate and high proficiency groups. While the high proficiency group is expected to be able to use allophonic cues and may show minimal epenthesis, the low proficiency group is expected to behave just like monolingual Japanese. The intermediate group will either show mixed results or just behave like the low proficiency group.

Globally, our findings buttress the claim that specific portions of the native language phonology are involved in speech perception as mentioned by Monahan et al. (2008) and listeners evaluate features (e.g. voicing) as relevant or irrelevant as a function of position in the syllables (Flege, 1995). Our study also indicates that speech perception is the results of the interaction between phonemes and phonotactics. Additional research is needed to tease apart the different roles that phoneme-level factors and phonological-level factors play in speech perception.

Reference

Berent, I., Steriade, D., Lennertz, T., Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104, 591-630.

Chang, Y., Hong, J. & Halleb, P. (2007). English cluster perception by Taiwanese Mandarin speakers. ICPhS XVI, 797-800.

Cutler, A., Otake, T. & McQueen, J. (2008). Vowel devoicing and the perception of spoken Japanese words. J. Acoust. Soc. Am. 125 (3), 1693-1703.

Davidson, L. & Shaw, J.A. (2011). Sources of illusion in consonant cluster perception. Journal of Phonetics 40, 234-248

Dehaene-Lambertz, D., Dupoux, E. & Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience* 12:4, 635-647.

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.

Dupoux, E. & Pallier, C., Kakehi, K., Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and cognitive processes,* 2001, 16 (5/6), 491-505

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.

Flege, J.E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277)

Imai, T. (2005). *Vowel Devoicing in Tokyo Japanese: A variationist approach*. Dissertation, Michigan State University.

Kabak, B. & Idsardi, K.W. (2007). Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints. *Language and Speech*, 50 (1), 23-52.

Kong, E.J. & Edwards, J. (2011). Individual differences in speech perception: evidence from visual analogue scaling and eye-tracking. *ICPhS XVII*. 1126-1129

Lamontagne, G. (1993). *Syllabification and consonant concurrence restrictions.* Unpublished PhD dissertation, Universtiy of Massachusetts, Amherst.

Massaro, D.W. & Cohen, M.M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, 34 (4), 338-348

Monahan, P., Takahashi, E., Nakao, C. & Idsardi, W. (2008). Not all epenthetic contexts are equal: differential effects in Japanese illusory vowel perception, *Proceedings of the 17th Japanese/Korean Linguistics Conference*.

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

Tsujimura, N. (2007) *An Introduction to Japanese Linguistics*, Second Edition, Blackwell.

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

Appendix A --- List of Stimuli

VC(Stop)[u]C(Stop)V

	Epenthesis	Devoicing
bilabial/velar	/abge/-/abuge/	/apke/-/apuke/
velar/alveolar	/egdo/-/egudo/	/ekto/-/ekuto/
bilabial/alveolar	/ibdo/-/ibudo/	/ipto/-/iputo/
velar stop/bilabial stop*	/egba/-/eguba/	/ekpa/-/ekupa/

VC(stop)[u]C(fricative)V

	Epenthesis	Devoicing
bilabial	/ebzo-ebuzo/	/epso-epuso/
velar	/egza-eguza/	/eksa-ekusa/

Appendix B --- List of Fillers

disyllabic	trisyllabic	quadrisyllabic
/rete/	/kanima/	/gigehada/
/neto/	/melimi/	/aezali/
/nena/	/keteke/	/tebigoba/
/mena/	/egeki/	/taamalo/
/kema/	/igiki/	/satebaki/
/lema/	/ritaka/	/jisimi/
/lomo/	/ritaki/	/mosaji/
/lite/	/rihaka/	
/moni/	/mihani/	
/tona/	/mehani/	
/bamo/	/nehani/	
/mota/	/kehake/	
	/kokige/	
	/ligako/	
	/galika/	
	/kelika/	
	/kelita/	
	/talama/	
	/tamala/	
	/selasa/	
	/lasesa/	
	/hikaka/	
	/hilisa/	
	/nilisa/	
	/takisa/	
	/logebe/	
	/segobo/	
	/nimato/	
	/ogika/	
	/datete/	
	/kabini/	