

Not All Epenthetic Contexts are Equal: Differential Effects in Japanese Illusory Vowel Perception

PHILIP MONAHAN, ERI TAKAHASHI, CHIZURU NAKAO AND WILLIAM
IDSARDI

University of Maryland

1. Introduction

Japanese speakers are notorious for epenthesis in consonant clusters (e.g. *makudonarudo*). Dupoux et al. (1999a) showed that this is also true at the level of speech perception and that native language phonology strongly affects how speech sounds are perceived. It has been found that Japanese listeners perceptually epenthesize [u] when hearing a consonant cluster (illusory epenthesis). On the other hand, it has not been shown what happens in contexts where [u] is not the appropriate epenthetic vowel, for example after coronal stops. This paper shows that in such contexts, no illusory epenthesis is found and that not all of phonology is employed in perceptual vowel epenthesis. Rather, there is a more intricate interplay between epenthetic context and vowel category and listeners are sensitive to both contexts that require repair and the relevant vowel category.

2. Japanese Syllable Structure

Japanese has a predominantly CV syllable structure that allows coda consonants only if it is a nasal (1) or if it is the first member of a geminate (2).

- | | | | |
|-----|----|------------|------------------|
| (1) | a. | [tom.bo] | ‘dragonfly’ |
| | b. | [kaN.ga.e] | ‘thought’ |
| (2) | a. | [kap.pa] | ‘a river sprite’ |
| | b. | [gak.koo] | ‘school’ |
| (3) | a. | *[kap.ta] | |
| | b. | *[tog.ba] | |
| | c. | *[pa.kat] | (from Itô 1989) |

As a result, when foreign words are borrowed into Japanese, they get altered so that they obey Japanese syllable structure (Ito & Mester 1995). In particular, a vowel is inserted in loan words to break up consonant clusters. In the following examples, the underlined letter represents an inserted vowel.

- | | | | |
|-----|----|---|-------------|
| (4) | a. | [fe.s <u>u</u> .ti.ba.ru] | ‘festival’ |
| | b. | [s <u>u</u> .fiN.k <u>u</u> .s <u>u</u>] | ‘Sphinx’ |
| | c. | [bi <u>i</u> .r <u>u</u>] | ‘beer’ |
| | d. | [k <u>u</u> .ri.s <u>u</u> .ma.s <u>u</u>] | ‘Christmas’ |

In most cases, [u] is inserted as in (4). It has been suggested that the reason why [u] is the most common epenthesis vowel is because it is least sonorant among the Japanese vowels (Lovins 1973, Katayama 1998, Peperkamp and Dupoux 2003) and is susceptible to devoicing (Jaeger 1978, Vance 1987, Tsuchida 1997, Varden 1998).

3. Illusory Epenthesis

Dupoux et al. (1999a) asked whether the vowel epenthesis effect observed in loan words occurs at the level of production or perception. Does the syllable structure of Japanese merely affect production or does it also affect speech perception? A continuum of nonword stimuli was created that ranged from items with no vowel between the consonants (e.g. *ebzo*), to items with a full vowel between the consonants (e.g. *ebu zo*). The interval of the continuum was 18 ms. In the offline judgment task, the Japanese subjects, but not French subjects, reported the presence of the vowel [u] between the consonants in all conditions. Even when there was no vowel pre-

sent, more than 70% of Japanese subjects reported that they heard [u]. In a speeded ABX discrimination task, Japanese participants had difficulty discriminating between *ebzo* and *ebuzo*. This study suggests that Japanese listeners perceptually epenthesize [u] when hearing a consonant cluster and that speech perception is affected by the phonotactic knowledge of the native language.

Dupoux et al. (2001) then asked whether such perceptual epenthesis was lexical in nature. That is, while there are no words in Japanese that contain the sequence [bz], there could be words containing the sequence [buz]. Therefore, the illusory perception of the vowel might be due to the fact that there are some potential words with such vowel between the consonants. To test this lexical effect, two sets of stimuli were created. Words in both sets had the shape CVCCV, such that the insertion of a vowel in the consonant cluster produces an existing Japanese word. One set (*u*-set; e.g. *sokdo*) was such that the insertion of [u] would produce a Japanese word (e.g. *sokudo*). The second set (non-*u* set; e.g. *mikdo*) was such that the insertion of non-epenthetic vowels [a], [i], [e] would produce an existing Japanese word (e.g. *mikado*), but the insertion of epenthetic [u] would produce a nonword (e.g. **mikudo*). If perceptual epenthesis were lexical in nature, the items in the non-*u* set would be perceived with a non-*u* vowel (e.g. *mikado*) because it would yield a Japanese word. On the other hand, if the effect is non-lexical in nature, words in both sets would be perceived with the epenthetic [u], and only the words in the *u*-set would be considered as a word. The results showed that in the lexical decision task, Japanese subjects classified the words in the *u*-set as words but the words in the non-*u* set as nonwords. Thus, perceptual epenthesis does not seem to be due to lexical feedback. In short, it was found that [a] is never an illusory vowel and that illusory epenthesis happened only with [u].

Dehaene-Lambertz et al. (2000) confirmed these results using event-related potentials. ERPs on Japanese and French subjects were recorded in a mismatch detection task. The stimuli consisted of four identical precursor items followed by the test item which was either the same as or different from the precursor. Subjects will show mismatch negativity (MMN) if they detect a mismatch between what is in the sensory memory left by the preceding stimuli and the currently perceived stimulus. French participants showed three MMN responses while Japanese participants showed lack of or weaker effects. This indicates that Japanese speakers did not perceive any difference between *ebzo* and *ebuzo*, and that speech input is processed fast and automatically according to phonotactic rules of the native language.

While above studies show that native language phonology affects speech perception, it does not tell us whether the perceptual epenthesis phenomena arises from violations of phonotactic constraints (e.g. the sequence

[bz] is illicit) or syllable structure (e.g. [b] is an illicit coda in Japanese). This is because all the stimuli used for these studies violated both of these restrictions. Using Korean, which has a phonological system that can tease these two possibilities apart, Kabak and Idsardi (2007) found that Korean listeners can distinguish consonantal sequences from their epenthesized counterparts only when first of the two consonants was a licit coda consonant. This suggests that perceptual epenthesis is caused by syllable structure violations and not by phonotactic violations. In other words, presence of an illicit consonantal sequence alone does not trigger perceptual epenthesis. Another important finding in Kabak and Idsardi (2007) was that sequences that never occur in Korean (e.g. [k.m], [l.n]) could nevertheless accurately be represented by the Korean listeners. The illicit sequences [k.m] and [l.n] never appear in Korean, nonetheless, the Korean listeners successfully discriminated them from the epenthesized counterparts.

In sum, previous studies show that native language phonology affects not only production but also speech perception and that illusory epenthesis is a perceptual phenomenon. Phonological constraints strongly and persistently influence how we perceive speech sounds.

4. Different Epenthetic Contexts and Vowels

All the consonants used in the previous studies (e.g. [g, k, b, ʃ]) are consonants that cause [u] epenthesis (i.e. [gu, ku, bu, ʃu]). However, [u] is not the only epenthetic vowel in Japanese. Different vowels are epenthesized depending on the context. For example, [tu] and [du] are illicit sequences in Japanese. So, in loan words that involve coda [t] or [d], [o] is epenthesized instead of [u].

- | | | | | |
|-----|----|--------------|---------------------------------------|--|
| (5) | a. | ‘fight’ | [fa.i.t <u>o</u>] | *[fa.i.t <u>u</u>] |
| | b. | ‘drive’ | [d <u>o</u> .ra.i.b <u>u</u>] | *[d <u>u</u> .ra.i.b <u>u</u>] |
| | c. | ‘strawberry’ | [s <u>u</u> .t <u>o</u> .ro.be.rii] | *[s <u>u</u> .t <u>u</u> .ro.be.rii] |
| | d. | ‘McDonald’s’ | [ma.k <u>u</u> .do.na.ru.d <u>o</u>] | *[ma.k <u>u</u> .do.na.ru.d <u>u</u>] |

Second, [i] is epenthesized after palatal affricates [tʃ, dʒ].

- | | | | | |
|-----|----|---------|---------------------|----------------------|
| (6) | a. | ‘catch’ | [kyat.tʃ <u>i</u>] | *[kyat.tʃ <u>u</u>] |
| | b. | ‘pitch’ | [pit.tʃ <u>i</u>] | *[pit.tʃ <u>u</u>] |
| | c. | ‘fudge’ | [fad.dʒ <u>i</u>] | *[fad.dʒ <u>u</u>] |
| | d. | ‘range’ | [ren.dʒ <u>i</u>] | *[ren.dʒ <u>u</u>] |

Lastly, [u] is inserted in all other environments.

In this way, there are three kinds of epenthetic contexts: after coronal stops (which requires [o] epenthesis), after affricates (which requires [i]

epenthesis), and the rest (which requires [u] epenthesis). Previous studies on Japanese illusory perception have only looked at the cases where [u] epenthesis should occur and the Japanese listeners were found to perceive the illusory [u]. In this paper, we are going to ask what do Japanese listeners perceive in environments where [o] epenthesis is required by the phonology? In other words, are all epenthetic contexts equal? Do people always perceive an illusory vowel that is appropriate for the given environment? Or do they always perceive [u] by default regardless of the environment?

5. Experiment

5.1. Method

**** native speakers of Japanese and ***** native speakers of English participated in the experiment. *** Japanese participants were recruited and run at Sophia University in Tokyo, Japan. The rest of the Japanese participants were recruited and run at the University of Maryland, College Park. All of the English participants were undergraduate students at the University of Maryland.

Following Kabak and Idsardi (2007), an AX discrimination task was used. Each participant was tested individually in a small room with a computer. Participants were instructed to listen to sound pairs over the headphones and make a judgment as to whether the two sounds were the same or different from each other. If they think the two sounds were the same, they were told to press <f> on the keyboard, and <j> if different. Participants were instructed to answer as fast as possible. No feedback was given during the experiment.

All tokens had the form [eC(V)ma]. Tokens were recorded in the form of [eCVma] with the full vowel present (e.g. *etoma*, *ekuma*) by a female native speaker of Japanese, who was blind to the purpose of the study. The recorded consonants and vowels are given in (7) and (8) respectively.

(7) C = [t, d, k, g, m, n]

(8) V = [o, u]

The recorded tokens were transferred into the Praat sound editor. Two types of stimuli were then created from the recorded material. The first type was of the form [eCma] without the medial vowel. These stimuli were created by splicing the entire vowel from the original signal. This means that a token [etma] could derive from either [etoma] with [o] spliced out (transcribed henceforth with the double strikethrough as in [et~~o~~ma]) or from [etuma] with [u] spliced out ([et~~u~~ma]). Another type of stimuli had the form [eCVma] with a full vowel. These were created by splicing three pitch peri-

ods of the vowel from the original signal and leaving 10 ms of the burst of the previous consonant. This was done to address the concern that participants might be able to tell the difference between the two types of stimuli based on which stimuli had been edited and which had not. By splicing out some pitch periods from all tokens, we ensured that all stimuli had some editing done to it.

Table 1: Words used in the experiment

		Vowels		
		[o]	[u]	None
Consonants	Coronal	[etoma]	[etuma]	[etma]
		[edoma]	[eduma]	[edma]
	Velar	[ekoma]	[ekuma]	[ekma]
		[egoma]	[eguma]	[egma]
	Nasal	[emoma]	[emuma]	[emma]
		[enoma]	[enuma]	[enma]

Table 1 lists all the words used in the experiment. Of the six consonants, the coronals [t, d] had the direct relevance to our question. The velars [k, g] acted as a control condition to see whether we could replicate previous findings (the ‘*ebzo* effect’) by Dupoux and colleagues. The nasals [m, n] acted as fillers for the experiment.

No cross-place or cross-voicing contrasts were tested (e.g. *[etma-ekma], *[etma-edma]), but all combinations of vowels ([o, u] or no vowel) were tested (e.g. [etuma-etuma], [etuma-etoma], [etuma-etma], [etoma-etuma], [etoma-etoma], [etoma-etma], [etma-etma]). The order of presentation was counterbalanced so that tokens were equally likely to be the first or the second of the pair. Each sound pair was repeated ten times and altogether there were 480 trials in the experiment (360 test pairs plus 120 filler pairs). All items were automatically randomized for each participant. There were three self-timed rest periods and the experiment session lasted approximately 40 minutes.

5.2. Hypotheses and predictions

There are at least three different hypotheses we can formulate with regard to illusory epenthesis in Japanese. First hypothesis (Hypothesis 1) holds that illusory epenthesis is indifferent to context and that Japanese listeners hear illusory [u] regardless of the environment. This hypothesis predicts that when presented with a token like [etma], Japanese would hear [etuma], even though [u] is not the appropriate epenthetic vowel after coronal stops and the sequence [tu] is illicit in Japanese. Under this view, [u] is always the default illusory epenthetic vowel.

Hypothesis 2 states that all of native language phonology is employed to constrain speech perception. Under this view, appropriate vowel category is perceived for each environment. So after coronal stops, illusory [o] is perceived, whereas after velars, [u] is perceived. Therefore, when presented with [etma], Japanese participants are predicted to hear [etoma].

Hypothesis 3 holds that illusory epenthesis takes some of phonology into account but not all of it. In particular, appropriate vowel is epenthesized only when context licenses an ‘illusory’ vowel, for example, a vowel that undergoes devoicing. According to this hypothesis, when Japanese listeners encounter a speech stream like [etma], they first attempt to parse it into syllables.

Figure 1: Initial attempt to parse into syllables

PUT SOME KIND OF PICTURE/FIGURE HERE

At this point, the identity of the vowel is unknown, but Japanese listeners know that there must be a vowel between [t] and [m], because [t] cannot be parsed as a coda consonant, so an unknown vowel V is assigned to a slot in the syllable. Trying to determine the category of the vowel V, [o] is the appropriate category following [t]. However, [o] does not undergo devoicing (and perhaps other reasons too, e.g. it is quite sonorous) and therefore it cannot be perceived illusory. Next candidates are [u] and [i] which undergoes devoicing and are common epenthetic vowels. However, neither [tu] nor [ti] is an acceptable sequence in Japanese. In Japanese loan words, the nature of the consonant [t] is usually changed to accommodate such sequences (/tu/ → [tsu]; /ti/ → [tʃi]). But in the case of [etma], there is no evidence that [t] has changed to [ts] or [tʃ]. Therefore, lack of evidence for consonant change prohibits those possible percepts (i.e. *[etuma], *[etima]). Therefore in this case, when presented with [etma], Japanese speakers are predicted to perceive something like [etVma] and they will be able to distinguish it from both [etuma] and [etoma].

5.3. Results

We report the results of our experiment in terms of A' scores. A' is a non-parametric version of d' scores that factors in subject response biases (**CITATION**). An A' score of 1.0 represents perfect performance while 0.5 represents chance-level performance.

(11) **PUT IN** A' score equations **

Figure 2 shows A' scores of Japanese and English subjects.

PUT FIGURE 2 HERE

Figure 2: Japanese and English mean A' scores

The mean A' scores for English and Japanese participants ranged between $A' = 0.998$ and $A' = 0.583$.

We will first report participants' performance on the velar conditions. As shown in Figure 3, Japanese subjects showed significantly poorer discrimination between *eguma/eguma* ($A' = 0.83$) and *ekuma/ekuma* ($A' = 0.882$) than English subjects ($A' = 0.948$, $\Delta A' = 0.12$; $t(28) = -2.29$, $p < 0.01$; $A' = 0.972$, $\Delta A' = 0.09$, $t(28) = -1.52$, $p < 0.05$, respectively). In other words, Japanese listeners could not discriminate [ekuma] and [ekma] as well as English subjects. This is a replication of Dupoux et al's (1999a) original *ebzo* effect, which confirms that Japanese speakers perceive illusory [u].

Figure 3

Sensitivity to presence/absence of vowel [u] with velar consonants

Next, as shown in Figure 4, Japanese participants successfully discriminated between *egoma/egoma* ($A' = 0.967$) and *ekoma/ekoma* ($A' = 0.937$). This result replicates Dupoux et al (2001) in that illusory epenthesis only happened with [u] and not with [o], in contexts where [u] is the appropriate

epenthetic category. Here, the Japanese listeners can be interpreted to have perceived [ekma] as [ekuma] and that is why they could discriminate [ekma] and [ekoma].

Figure 4
Sensitivity to presence/absence of vowel [o] with velar consonants

Our results so far replicate previous studies. But our main question in this paper is what Japanese speakers perceive in contexts where [o] is the appropriate epenthetic vowel. As Figure 5 shows, on the coronal conditions, Japanese subjects showed better discrimination between *edoma/edoma* ($A' = 0.953$) and *etoma/etoma* ($A' = 0.955$) than with velars. This result suggests that Japanese listeners could distinguish between [etoma] and [etma], and that Japanese listeners do not perceive illusory [o] in the environments where [o] should be epenthesized, contrary to what was predicted by Hypothesis 2.

Figure 5
Sensitivity to presence/absence of vowel [o] with coronal consonants

On the other hand, Hypothesis 1 predicted that Japanese speakers perceive [etma] as [etuma]. However, Japanese subjects showed a good discrimination of *eduma/eduma* ($A' = 0.974$) and *etuma/etuma* ($A' = 0.95$), as shown in Figure 6. This shows that, with coronals, [u] does not get perceptually epenthesized either, contrary to the predictions of Hypothesis 1.

Figure 6

Sensitivity to presence/absence of vowel [u] with coronal consonants

As Hypothesis 3 predicted, Japanese speakers appear to have perceived neither illusory [o] nor [u]. Even though Japanese participants' performance on *edoma/edoma* and *etoma/etoma* may have been poorer than English participants', the important point is that Japanese participants' performance on the coronal condition is significantly better than the velar condition, as shown in Figure 7. In specific, collapsing across voicing (no effects of voicing were found), Japanese speakers performed significantly better at discriminating coronals with epenthetic [o] (*e[d,t]oma/e[d,t]ma*) than velars with epenthetic [u] (*e[g,k]uma/e[g,k]ma*; $\Delta A' = 0.1$; $t(15) = 3.15$; $p < 0.01$).

Figure 7

Sensitivity to presence/absence of vowel with velar vs. coronal consonants

6. Discussion

The results from our experiment found no illusory epenthesis after coronal stops. It seems that, unlike [u], the [o] vowel does not get perceptually epenthesized. Why might this be? Recall that the explanation for illusory [u] epenthesis in Japanese was that (a) it is least sonorant among the Japa-

nese vowels and (b) it is susceptible to devoicing. With [o], however, these conditions are not met. First, unlike [u], [o] does not undergo devoicing. So, it could be that only vowels that undergo devoicing get perceptually epenthesized. Second, [o] is higher than [u] in the sonority hierarchy. That is, [o] is more sonorous and thus may be too perceptible to be illusory (i.e. lack of [o] would be too noticeable).

In this way, there could be a number of reasons why Japanese speakers do not perceive illusory [o] after coronals. What we found was that after coronals, Japanese listeners do not perceptually epenthesize [u] like in other contexts. This suggests that contexts do play a role in illusory epenthesis and that there is no single 'default' epenthetic vowel. Phonological constraints such as a ban on sequences [tu] and [du] are in play as seen in the lack of [u] epenthesis after coronals. Moreover, there seems to be a complex interplay between epenthetic context and vowel category. In other words, some phonology matters for speech perception, but not all. In specific, perceptual illusion seems to be sensitive to the fact that [u] and not [o] is the appropriate vowel category after velars. It is also sensitive to the fact that [u] is not the appropriate vowel category after coronals. But it is impossible to epenthesize [o] after coronals in order to salvage the syllable structure even though it is the appropriate category. We speculated the reasons for this above. It seems that not all epenthetic contexts are equal with respect to illusory epenthetic vowel perception. Some contexts (e.g. after velars) allow illusory epenthesis while others (i.e. after coronals) do not even with the appropriate vowel category.

7. Conclusion

In environments where [u] is not the appropriate epenthetic vowel (after coronal stops), no illusory vowel perception was found. Japanese listeners are sensitive to both contexts that require repair and the relevant vowel category. Questions remain as to what Japanese listeners are perceiving when they hear a cluster such as *etma* and what their representation is.

References

- Bloomfield, L. 1933. *Language*. New York: Holt.
- Bolinger, D. 1965. The Atomization of Meaning. *Language* 41:555–573.
- Brugmann, K. 1906. *Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*. 2nd ed., vol. 2, part 1. Strassburg: Trubner.
- Chafe, W. L. 1965. Review of Grammar Discovery Procedures, by R.E. Longacre. *Language* 41: 640–647.

- Chomsky, N. 1957. *Syntactic Structures*. (Janua Linguarum, 4.) The Hague: Mouton.
- Hockett, C. F. 1964. The Proto Central Algonquin Kinship System. *Explorations in Cultural Anthropology*, ed. W. Goodenough, 329-358. New York: McGraw-Hill.
- Ross, J. R. 1967. Constraints on Variables in Syntax. Doctoral dissertation, MIT.