THE INFLUENCE OF PRECEDING CONSONANT ON
PERCEPTUAL EPENTHESIS IN JAPANESE

Wakayo Mattingley (University of Canterbury), Elizabeth Hume (University of Canterbury)
Kathleen Currie Hall (University of British Columbia)
wakayo.mattingley@pg.canterbury.ac.nz; beth.hume@canterbury.ac.nz; kathleen.hall@ubc.ca

ABSTRACT

Research on perceptual epenthesis in Japanese has revealed high back [ɯ] to be the vowel commonly perceived in illicit consonant sequences. However, loanword studies suggest that there are three epenthetic vowels, which reflect phonotactic restrictions on certain consonant + vowel sequences. Expanding previous perception studies, this paper investigates the extent to which perceptual epenthesis in Japanese is also constrained by the language's phonotactic patterns. In particular, we seek to determine to what extent the preceding consonant influences perceptual epenthesis, reflecting native phonotactics. Our results show that the preceding consonant does influence the vowel perceived yet, at the same time, there is a strong bias toward perceiving [ɯ] in contexts not predicted by the language's phonotactic patterns.

Keywords: perceptual epenthesis, consonant-vowel interaction, native phonology, Japanese

1. INTRODUCTION

Loanword studies of Japanese epenthesis have found that the quality of a preceding consonant can influence the vowel produced in a phonotactically illicit sequence [10, 14]. Among the five different vowel qualities in Japanese [i e a o ɯ], the high front vowel [i] is observed after alveo-palatal affricates [tɕ dʑ], [o] occurs after alveolar stops [t d], and [ɯ] occurs elsewhere, i.e. after labial and velar consonants. The preference for epenthetic [o] as opposed to [i] or [ɯ] is assumed to follow from the observation that [tɯ], [dɯ], [ti], [di] are not well-formed native Japanese syllables [10, 14].

Results from research on perceptual epenthesis differ in some respects from the above. One contributing factor may be due to the fact that perception studies have thus far taken into account only a subset of the vowel qualities and preceding consonantal contexts examined in loanword and production studies. For example, in [3, 4] Japanese listeners were presented with stimuli that included a subset of the language's five vowel qualities: the

high front and/or high back vowels [i ɯ]. The studies found that Japanese listeners were more likely to perceive a high back as opposed to high front vowel in an illicit consonant sequence (see also [14]).

The influence of preceding consonants is considered in [11], which focusses on the perception of potential illusory vowels [ɯ, o] after alveolar [t d] and velar [k g] stops. They interpret their results as indicating that Japanese listeners did not perceive the contextually predicted vowel [o] after alveolars, nor did they perceive an illusory epenthetic [ɯ] in this context. The high back vowel was, however, perceived after velar stops, consistent with earlier production studies. The authors conclude that native language phonology alone cannot explain the perception of non-native speech.

The goal of the current study is to build on this previous research in order to provide a more comprehensive view of perceptual epenthesis in Japanese. To do so, we examine the potential influence of preceding consonants at four different places of articulation, using stimuli that include all five vowel qualities.

To anticipate the results, the vowel percept was found to be constrained by native phonotactics, yet in only some contexts. At the same time, [ɯ] was perceived in positions beyond those expected from the phonotactics, supporting the view that beyond native phonology, phonetics and potentially other factors are also at play. The results inform current debates about the nature of the epenthetic vowel as well as, more generally, the interaction of phonetics and phonology in speech perception.

2. METHOD

2.1. Stimuli

Stimuli for the perception experiment were produced by a 23-year-old male native speaker of Japanese, fluent in both Japanese and English. The speaker had no linguistic training.

In order to induce perceptual epenthesis, the stimuli included consonant sequences that do not occur in Japanese, i.e. non-homorganic consonant clusters. More specifically, the stimuli consisted of

pseudo-words of the form [aC$_1$(V)C$_2$a] with consonants selected from the set of voiced obstruents {b, d, g, dʑ} and C$_1 \neq$ C$_2$, giving a total of 12 different consonantal combinations. Voiced consonants were used in order to avoid potential challenges that could arise in analysing vowels between voiceless consonants since they can be devoiced in Japanese [17]. (V) represented one of five Japanese vowel qualities {a, e, i, o, ɯ} or no vowel.

Following [4], four different conditions were created for each stimulus by manipulating the original recordings in Praat [1]. These conditions involved differing the duration of the target vowel, yielding durations at 0%, 15%, 25%, and 50% of the the original duration of the target vowel. Duration was measured using a zero-crossing method with consideration given to vowel formants. The complete set of stimuli was made up of a total of 312 audio files: (a) 60 with full vowels (12 consonant combinations * 5 vowels), (b) 240 with reduced vowel durations (60 basic sound stimuli * 4 duration types); and (c) 12 with no original vowel between the consonants (i.e. aCCa-forms). Although our subjects listened to all sound files, in this paper we only present the analyses of tokens with full vowels (a), and those with consonant clusters in the original stimuli (c).

2.2. Participants

Seventeen native speakers of Japanese (seven males, ten females, ages 19-31, mean age 24.37) currently living in Christchurch, New Zealand were tested at the University of Canterbury. All participants had been in New Zealand for less than one year.

2.3. Procedure

Participants were tested individually in a sound-attenuated room at the University of Canterbury. Each participant was situated in front of a computer screen wearing Panasonic RP-HT265 headphones. The computer screen displayed the pseudo-word template 'aC?Ca' and six options (written in Japanese): 'a', 'e', 'i', 'o', 'u', 'no vowel'. Stimuli were presented one at a time, in random order to the participant who listened to each stimulus twice over headphones. The participant was asked to click on the option that best represented what she/he heard between the consonants of the pseudo-word. A brief training phase preceded the test phase. The entire experiment took approximately 20 minutes to complete. All aspects of the experiment were conducted in Japanese.

3. RESULTS AND ANALYSIS

3.1. Identification accuracy for each vowel type

Table 1 displays how accurately the Japanese subjects identified the vowel stimuli in the full vowel tokens. Cells provide the percentage of responses for each stimulus along with the number of responses (in parentheses). The diagonal (boldface) indicates percentages of correct responses.

Table 1: Percent identification accuracy in tokens with full vowel stimuli

		Given Response (n=1020)						
		[a]	[e]	[i]	[o]	[ɯ]	no vowel	Total
Correct Response	[a]	**100%** (204)						100% (204)
	[e]	.5% (1)	**99.5%** (203)					100% (204)
	[i]		1% (2)	**99%** (202)				100% (204)
	[o]	1% (2)			**98%** (200)	.5% (1)	.5% (1)	100% (204)
	[ɯ]	.5% (1)				**99%** (202)	.5% (1)	100% (204)

From the table above we can see that all of the vowels were correctly identified at least 98% of the time. This suggests that native speakers of Japanese are able to distinguish each vowel clearly from others and identify it in the task used here. With this as a basis, we now examine the perception of vowels in non-native consonant clusters.
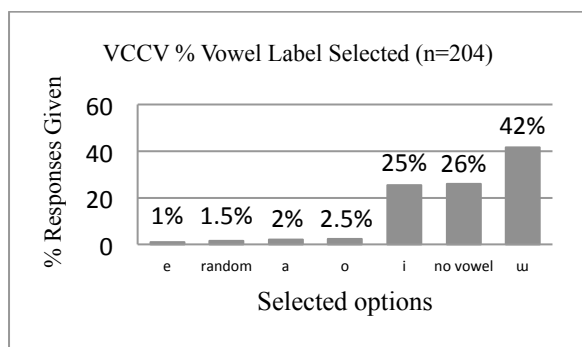
3.2. Perceptual epenthesis
3.2.1. Perceptual Epenthesis in VCCV

We begin by presenting the results for stimuli with consonant clusters, analysed with all consonant types grouped together. Recall that all stimuli had the form [VCCV] and V was always [a]. Seventeen participants listened to 12 tokens each, for a total of 204 tokens.

As seen in Figure 1, the stimuli were correctly identified as having no vowel only 26% of the time. The results thus show that listeners are highly likely to perceive a vowel between the two consonants, consistent with the findings of [3, 4].

The figure also shows that the illusory vowel was overwhelmingly perceived as [ɯ] 42% of the time and as [i] 25% of the time. On the other hand, [a], [e] and [o] were selected less than 3% of the time each. These results are consistent with the claim, based on the analysis of loanword data [10], that the Japanese epenthetic vowels are commonly high vowels. This finding is also consistent with [3, 4, 14] who found Japanese listeners biased to perceive an illusory [ɯ] when there was no vowel present in the stimuli.

VCCV % Vowel Label Selected (n=204)

| | C₁ | Given Response (n=204) | | | | | | | |
		[a]	[e]	[i]	[o]	[ɯ]	no vowel	random	Total
Correct Response= no	**[b]**	2% (1)	2% (1)	6% (3)		**55% (28)**	31% (16)	4% (2)	100% (51)
	[d]	2% (1)	2% (1)	4% (2)	**10% (5)**	43% (22)	37% (19)	2% (1)	100% (51)
	[g]	4% (2)				**63% (32)**	33% (17)		100% (51)
	[dʑ]			**92% (47)**		6% (3)	2% (1)		100% (51)

We turn now to results concerning the potential influence of the preceding consonant on the perception of the illusory vowel in VCCV stimuli.

3.2.2. Effect of preceding consonant on the epenthetic vowel

Table 2 provides the breakdown of responses to VCCV stimuli in order to determine to what extent a consonant's place of articulation influences the epenthetic vowel percept. Percentages indicate each response reported for a given stimulus with row totals adding up to 100%. The numbers in parentheses show the actual number of responses, with row totals adding up to 51. A chi-square test with 2000 Monte Carlo simulations to correct for small sample sizes indicates that this distribution of results is (not surprisingly) significantly different from chance [χ^2=174.96, p<0.001].

We extrapolate from loanword and production studies on the influence of preceding consonant [10, 14] to make the following predictions. All else being equal, if a vowel is perceived between the two consonants we would expect the vowel percept to be [i] after [dʑ], [o] after [d], and [ɯ] after [b] and [g], respectively; boldfaced percentages (shaded cells) highlight expected responses.

As the results show, listeners correctly reported no vowel present at least 31% of time for each context, except after [dʑ] when no vowel was perceived only 2% of the time. When a vowel was perceived, the findings are relatively consistent with the expected patterns for [b], [g] and [dʑ], but less so for [d], as we discuss below.

[ɯ] is the only vowel to be identified after all four consonants, albeit marginally after [dʑ], and as expected, it is the most common vowel selected after [b] (55%) and [g] (63%). An exact binomial test comparing the percentage of actual [ɯ] responses (42.29%) to what would be expected by chance (16.67%) indicates that this is a significant bias (p<0.001). There is also a strong bias toward [i] after the alveopalatal affricate (92% vs. expected 16.67%; p<0.001). Relating this finding back to Figure 1 where [i] was shown to be the second most common illusory vowel (after [ɯ]), when preceding context is taken into account, almost all [i] responses are seen to occur with a preceding [dʑ]. After other consonants, [i] was reported at most only 6% of the time. This is indicative of a significant bias *against* [i] in these other contexts (actual 6% vs. expected 16.67%; p<0.001). This finding supports the view of [i] as a context-dependent default epenthetic vowel after [dʑ] [14].

For [o], the results are also as predicted in that the vowel is expected after [d], and this is the only context in which [o] was selected. However, [o] was perceived in this context only 10% of the time. That is, although Japanese listeners show a statistically significant bias to use [o] after [d] as opposed to other consonants (actual 100%, expected 25%, p=0.001), there is no statistically significant bias toward (or against) using [o] as the epenthetic vowel in this context (actual 10%, expected 16.67%, n.s.). In fact, the illusory vowel after [d] is overwhelmingly identified as [ɯ] (43%) or as 'no vowel' (37%). These results run counter to the claim that [o] is the context-dependent epenthetic vowel after [d] [14]. They also differ from [11], where it was claimed that Japanese listeners did not perceive an illusory epenthetic [ɯ] nor the contextually predicted vowel [o] after a coronal consonant. We suspect that these differences relate to the use of different tasks from that of the present study. Consequently, it may be that subjects were accessing different levels of knowledge. In [14], stimuli with illicit consonant clusters were presented to Japanese subjects visually who were asked to transcribe the words in Japanese, thus emphasizing the language's

writing system and constraints thereon. In [11], listeners performed an AX discrimination task which arguably taps into a more purely auditory level of discrimination than the identification task used in this study [2].

4. DISCUSSION

This study investigated the extent to which perceptual epenthesis in Japanese is influenced by the quality of the preceding consonant and constrained by native phonotactic patterns. Consistent with previous loanword studies, [i] was predominantly selected as the epenthetic vowel after the alveopalatal affricate [dʑ]. On the other hand, while the mid back vowel [o] was identified after the alveolar stop [d] as expected, it was by no means the most common choice; instead, when a vowel was perceived, listeners overwhelmingly identified [ɯ] as the illusory vowel. Thus, even though [dɯ] is not an observed sequence in native Japanese, listeners were strongly biased to perceive [ɯ] in this context.

This finding indicates that the perception of [ɯ] as the epenthetic vowel is generalising to contexts beyond what would be predicted by the language's phonotactics. What would lead to this particular vowel being selected as opposed to any other vowel? Why would Japanese listeners perceive [ɯ] even though the occurrence of [o] after [d] is most common in native Japanese? One potential explanation relates to the weak phonetic nature of [ɯ]. For example, [ɯ] is considered to have the shortest duration among the five vowel types [13], consistent with acoustic analyses of the stimuli in this study. Further, [ɯ] is less likely to be accented than other vowels [16]. Under the assumption that shorter duration and unaccentedness correlate with weaker salience, listeners are thus associating the illusory vowel with the least salient vowel [14, 15].

Yet, why select this vowel after [d] even though listeners know that [dɯ] is not a viable sequence in native Japanese? One potential explanation may relate to the fact that since the stimuli were nonsense words, subjects considered them to be foreign and thus not subject to the phonotactic sequencing constraints of their own language ([5] reports, for example, 7 instances of [dɯ] sequences in the NTT lexicon of Japanese, all loanwords). As a result, they associated the phonetically weak vowel percept with the vowel with the most similar qualities given their linguistic experience, i.e. [ɯ]. This account is consistent with the view that perceptual adaptations to non-native speech are phonetically minimal [12].

It may also be the case that the language's statistical patterns support a strong perceptual bias toward [ɯ]. As such, in contexts with weak perceptual cues, subjects are biased to perceive the most expected vowel in the language [7]. We leave this open for future consideration.

Such a bias was not observed after [dʑ] where the vowel percept was almost exclusively [i]. We interpret this as being due to the observations that first, the pattern accords with Japanese phonotactics, and second, it is phonetically natural cross-linguistically [9]. Indeed, given that the consonant and vowel share similar articulatory and perceptual properties [6], the identification of [ɯ] instead of [i] would have been highly surprising.

5. CONCLUSION

This study has extended previous research on perceptual epenthesis in Japanese by taking into account a broader range of potential vowel percepts as well as preceding consonantal context. Consistent with the language's phonotactic patterns, a preceding consonant is shown to influence the perception of an illusory vowel, though not in all cases. Contra native phonotactics, an arguably less salient vowel is perceived as epenthetic in an otherwise illicit consonant sequence. This supports the view that an adequate account of perceptual epenthesis requires reference to the interaction of phonological, phonetic, as well as potentially other factors [11, 8].

6. REFERENCES

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

[2] Boomershine, A., Hall, K.C., Hume, E., Johnson, K. 2008. The impact of allophony vs. contrast on speech perception. In Avery, P., Dresher , B.E., Rice, K. (eds.), *Phonological Contrast*. Mouton.

[3] 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.

[4] 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.

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

[6] Hume, E. 1992. *Front Vowel, Coronal Consonants and their Interaction in Non-Linear Phonology*. New York: Garland Publishing.

[7] Hume, E., Bromberg, I. 2005. *Predicting Epenthesis: An Information-theoretic Account*. 7èmes *Journées internationales du réseau français de phonologie (7th*

Annual Meeting of the French Network of Phonology). Aix-en-Provence, France.

[8] Hume, Elizabeth, Kathleen Currie Hall, Andrew Wedel, Adam Ussishkin, Martine Adda-Decker and Cédric Gendrot. 2013. Anti-Markedness Patterns in French Epenthesis: An Information Theoretic Approach. In Chundra Cathcart, I-Hsuan Chen, Greg Finley, Shinae Kang, Clare S. Sandy, and Elise Stickles, eds., *Proceedings of the Thirty-Seventh Annual Meeting of the Berkeley Linguistics Society*. Berkeley: Berkeley Linguistics Society. pp.104-123.

[9] Kochetov, A. 2011. Palatalization. In M. van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (eds), *Companion to Phonology*. Wiley-Blackwell. 1666-1690.

[10] Irwin, M. 2011. *Loanwords in Japanese*. Philadelphia: John Benjamins Publishing Company.

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

[12] Peperkamp, S., Dupoux, E. 2003. Reinterpreting loanword adaptations: the role of perception. *Proceedings of ICPhS 15*, 367-370.

[13] Sagisaka , Y., Tokuhara, Y. 1984. Phoneme duration control for speech synthesis by rule. *Trans. Inst. Electron.Commun*. J67-A , 629-636.

[14] Shoji, S., Shoji, K. 2014. Vowel epenthesis and consonant deletion in Japanese loanwords from English. *Proceedings of the Annual Meeting on Phonology*.

[15] Steriade, D. 2001. 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. Cambridge, MA: MIT Press.

[16] Yoshida, Y.Z. 2004. *Asymmetric distribution of accents and the related issue of vowel duration.* Paper presented at the *Speech Prosody 2004*, International Conference.

[17] Vance, T.J. 2008. *The Sounds of Japanese*. Cambridge, UK: Cambridge University Press.