Japanese coda [m] elicits both perceptual assimilation and epenthesis

A. J. Kilpatrick[1], S. Kawahara[2], R. L. Bundgaard-Nielsen[3,4], B. J. Baker[1,] and J. Fletcher[1]

[1]University of Melbourne, Australia, [2]Keio University, Japan, [3]Western Sydney University, Australia,
[4]MARCS institute for Brain, Behaviour and development, Australia
[1]*alex.kilpatrick@unimelb.edu.au*

Abstract

When listeners are exposed to non-native speech, they sometimes perceptually repair the incoming signal to better adhere to the phonotactics of their native language. The present study examines two repair strategies—perceptual assimilation and epenthesis—in two experiments designed to test how Japanese listeners perceive non-homorganic coda [m]. In the first experiment, Japanese listeners categorized tokens that contain a medial coda [m] or a word-final coda [m] into categories represented by the Japanese Hiragana orthography with /mu/ or /ɴ/ in the target position. In Experiment 2, participants discriminated between tokens with coda [m] and tokens with [mu] or [n] in the target position in a series of AXB discrimination tests. The results show that Japanese listeners employ both perceptual epenthesis and perceptual assimilation when exposed to coda [m] sequences but favor assimilation over epenthesis, particularly in the word-final position.

Keywords: Phonetics, phonology, Japanese nasals, perceptual epenthesis, perceptual assimilation.

1. Introduction

Non-native listeners occasionally experience perceptual illusions when exposed to phonotactically prohibited sequences of sounds. These illusions typically repair phonotactic violations in non-native (L2) speech so that the signal adheres to native (L1) phonotactics. Two of the ways in which listeners repair phonotactically unattested sequences are perceptual assimilation and epenthesis. Perceptual assimilation is typically studied and discussed in terms of non-overlapping phonemic and phonetic inventories. However, perceptual assimilation can also occur when listeners are exposed to L2 speech that adheres to L1 phonemic and phonetic inventories but violates L1 phonotactics [e.g., 1,2]. For example, Japanese maintains several co-occurrence restrictions which limit how consonants and vowels may combine in CV sequences (e.g., */si/, */ti/ and */tu/) and when exposed to L2 speech that violates these restrictions, Japanese listeners sometimes repair the signal by assimilating one of the sounds into an unexpected phonemic category (e.g., /si/ is confused with /ʃi/) [1]. In AXB discrimination tests, Japanese listeners are significantly less accurate and slower to respond to trials when testing the discriminability of tokens when one token violates co-occurrence restrictions (e.g., /isi/ - /iʃi/) compared to when both tokens adhere to L1 phonotactics (e.g., /usu/ - /uʃu/). This result can be taken as evidence that for example, /isi/ is perceptually assimilated to /iʃi/.

Loanword assimilation behavior suggests that Japanese co-occurrence restrictions are undergoing weakening, as some of these restrictions are sometimes not being adhered to in some recent loanwords. This weakening is not uniform across restrictions; some restrictions are adhered to in most recent loanwords (e.g., */si/ and */tu/) while others are adhered to infrequently (e.g., */ti/) [3]. This difference is reflected in the above study [1], whereby tokens that contain infrequently adhered to restrictions are easier to discriminate from licit tokens (e.g., /iti/ - /itʃi/) compared to tokens that contain restrictions that are adhered to in most recent loans (e.g., /isi/ - /iʃi/).

Perceptual epenthesis involves illusory vowels which typically split unattested consonant clusters. L1 Japanese listeners are excellent participants in perceptual epenthesis studies because Japanese only allows consonant clusters if the first consonant is a so-called moraic nasal /ɴ/ or the first part of a geminate [4]. When Japanese listeners are exposed to L2 speech that contains non-homorganic consonant clusters, they sometimes perceive an epenthetic /u/ which serves to repair the violation [5,6]. In one study, French and Japanese participants were exposed to tokens that contained non-homorganic consonant clusters (e.g., /ebzo/). The Japanese participants reported hearing a medial /u/ amid consonant clusters and were quite inaccurate at discriminating between tokens that contained clusters and tokens with a /u/ in the epenthetic position (e.g., /ebzo/ - /ebuzo/). The French participants—for whom the non-homorganic consonant clusters are phonotactically licit—did not report hearing an epenthetic /u/ and were significantly

10.21437/ISAPh.2018-14

more accurate at discriminating between contrasts [5]. Japanese listeners typically perceive /u/—as opposed to other vowels—in these instances, arguably because /u/ is the most 'minimal'—shortest and most quiet—vowel in Japanese [6].

Japanese listeners' perception of coda [m] offers an interesting case study in phonotactically conditioned perceptual repair, because both assimilation and epenthesis are potentially viable repair strategies. The current study presents two experiments which are designed to examine whether coda [m]—produced medially and word-finally—elicits perceptual assimilation (whereby [m] is confused with [n]) or perceptual epenthesis (whereby [m] is confused with [mu]), or both, in L1 speakers of Japanese. Japanese has three phonemic nasals, /m/, /n/ and /ɴ/, the last of which is generally treated as phonologically placeless [7] (although recent research suggests that coda /ɴ/ can be realized with bilabial closure, particularly in the word-final position [8]).

Phonotactically conditioned repairs—such as perceptual assimilation and epenthesis—are not accounted for in speech perception models that only consider the effect of non-overlapping and/or partially overlapping phonemic and phonetic categories on L2 speech perception. One such model, PAM-L2 [9], proposes that the discriminability of non-native segments is predictable from the degree of articulatory or gestural similarity between native and non-native segments. These sounds are filtered by the listener's linguistic experience so that non-native sounds are assimilated to their nearest native phonemic category. When two non-native phones are assimilated to the same category, they become more difficult to distinguish. In the framework of PAM-L2, this is known as either 'single category' assimilation—when both non-native phones match equally well to the assimilated category—or 'category goodness' assimilation—when sounds at issue vary in how well they are perceived to fit. A recent extension of PAM-L2 expands this model, proposing that phonotactically conditioned repair follows a similar pattern to those outlined in PAM-L2 [1]. Kilpatrick et al. propose to incorporate transitional probability into PAM-L2 to predict and account for the perceptual difficulties listeners experience when exposed to sequences of speech that are unattested (or are improbable) in their L1 [1]. This model proposes that unattested sequences will sometimes be assimilated to sequences that adhere to L1 phonotactics. The frequency of such an assimilation is predicated on the articulatory distance between the signal and the target and the predictability of both.

This extension suggests that perceptual assimilation and epenthesis are similar processes where both involve the assimilation of an unattested sequence to its perceptually nearest, transitionally probable match. In the case of perceptual assimilation, the nearest match for the unattested sequence contains the same number of sounds but one or more of those sounds are assimilated to an otherwise unexpected phonemic category, as in the case of the Japanese /isi/ - /iʃi/ example discussed above. On the other hand, perceptual epenthesis involves the assimilation of an unattested sequence to a sequence that contains an additional sound (e.g., /ebzo/ - /ebuzo/). In the case of Japanese, the /u/ is perceived because it is the most minimal vowel, making the distance between it and the vowel-less sequences shorter than that of other Japanese vowels. If this model is accurate, then certain conditions should elicit both perceptual assimilation and epenthesis provided that the assimilation target and the epenthesis target are somewhat similar in terms of their transitional probability and/or their articulatory distance from the unattested (or improbable) sequence. Such conditions are tested for in the following experiments: Experiment 1: Categorisation and Goodness of Fit and Experiment 2: Discrimination.

2. Method

Twenty L1 Japanese students (female = 15) were recruited from Keio University in Tokyo to complete the two experiments in succession. Both tests took less than an hour and participants were paid JP¥1,500 for their time. Participants were selected based on a limited exposure to languages other than Japanese: all participants were L1 Japanese speakers born to L1 Japanese speaking parents and no participant had spent more than 1 month in a country other than Japan.

Tokens were produced by three female speakers of Australian English and recorded in a recording studio located at the University of Melbourne. Tokens consisted of six attested nonce words and two coda [m] nonce words which vary in whether coda-position [m] occurs medially or word-finally, as in Table 1. (In the following, we include the [eton] token in the "attested token" category in order to differentiate coda [m] tokens from other tokens.) Experiments were conducted in a quiet room located at the Mita campus of Keio University.

Experiment 1 required participants to listen to attested and coda [m] tokens and categorize them into categories presented in Hiragana, the most basic Japanese script (see Table 2). Hiragana was used to represent categories because it offers no way of representing coda [m], participants were therefore

forced to choose between categories that contain either /ɴ/ or /mu/. Participants were also asked to assign each token a Goodness of Fit rating out of seven (where 7 is an excellent fit). Each participant completed 144 categorization trials (18 per token).

Table 1: List of tokens used in Experiment 1 and 2.

Attested Tokens	Coda [m] Tokens
[etomu]	[etom]
[emuto]	[emto]
[eton]	
[ento]	
[teno]	
[temo]	

Experiment 2 consisted of five contrasts organized into AXB triads (see Table 3) made up of the same tokens used in Experiment 1. One contrast ([teno] – [temo]) was designed to test the discriminability of [n] and [m] in phonotactically attested sequences. The remaining four of the AXB contrasts were designed to test whether listeners were employing perceptual assimilation ([ento] – [emto] & [eton] – [etom]) or perceptual epenthesis ([emuto] – [emto] & [etomu] – [etom]), using discriminability as a proxy for this behaviour. All tests in Experiment 2 were presented in a single block from which AXB triads were drawn at random. Tokens were spaced with a 1,000 ms interstimulus interval and trials were spaced with a 1,500 ms intertrial interval. Each participant completed 240 AXB trials (48 per contrast).

3. Results

In Experiment 1, participants were more likely to assign [emto] to the epenthesis (/emuto/) category (75%), while [etom] was significantly more likely to be assigned to the assimilation (/etoɴ/) category (68%). A *t*-test calculated on the rate at which coda [m] tokens were assigned either to assimilation or epenthesis categories revealed a significant difference between the two conditions ($t(19) = 5.66$, $p < 0.001$). Goodness of Fit ratings do not entirely follow this pattern, however. [emto] tokens which were assigned to the /emuto/ category received a lower Goodness of Fit rating (3.74) than those that were assigned to the /ento/ category (4.65). However, this difference in Goodness of Fit ratings was less different than [etom] tokens that were assigned to the /etomu/ (4.04) and /etoɴ/ categories (5.43). An ANOVA calculated on the average Goodness of Fit scores that each participant assigned to coda [m] tokens revealed a significant effect ($F(3,72) = 11.42$, $p < 0.001$). A post-hoc comparison with Bonferroni correction revealed a significant difference between Goodness of Fit scores for [etom] assigned to /etomu/ and /etoɴ/ ($p < 0.001$), but not for [ento] assigned to /emuto/ and /eɴto/ ($p = 0.06$)

In Experiment 2, participants were more accurate at discriminating between pairs that differed in terms of perceptual epenthesis ([etomu] - [etom] 93%; [emuto] - [emto] 83%) than they were at those that tested perceptual assimilation ([ento] - [emto] 70%; [eton] - [etom] 60%). A one-way ANOVA was conducted to compare these five conditions, which revealed a significant effect on accuracy ($F(4, 99) = 38$, $p < 0.001$). A post-hoc comparison with Bonferroni correction revealed a significant difference between all conditions ($p < 0.05$), except for [teno] - [temo] & [etomu] - [etom] contrasts ($p = 1$) and [teno] - [temo] & [emuto] - [emto] contrasts ($p = 0.81$). ANOVA was also run on the response times for all participant responses, which also revealed a significant effect ($F(4, 4799) = 95$, $p < 0.001$). A post-hoc comparison with Bonferroni correction was conducted on this ANOVA, which revealed a significant difference in response times for all comparisons ($p < 0.001$) except for [emuto] - [emto] & [ento] - [emto] contrasts ($p = 1$).

Table 2: Results of Experiment 1: Categorisation and Goodness of Fit. Responses and their Hiragana categories are presented horizontally across the top of the table, stimuli are presented vertically on the left. Goodness of Fit scores are presented in parentheses.

	えとむ **etomu**	えむと **emuto**	えとん **etoɴ**	えんと **eɴto**	ての **teno**	ても **temo**
[etomu]	99% (5.16)		1% (1.50)			
[etom]	32% (4.04)		68% (5.43)			
[emuto]		100% (5.47)				
[emto]		75% (3.74)		25% (4.65)		
[eton]	1% (2.20)		99% (5.26)			
[ento]		2% (3.38)		98% (5.41)		
[teno]					99% (5.41)	1% (2.67)
[temo]						100% (4.82)

Table 3: Results of Experiment 2: Discrimination. Responses and their Hiragana categories are presented horizontally across the top of the table, stimuli are presented vertically on the left.

AXB Test Pairs	Accuracy	Accuracy St.Dev	Response Time	Response Time St.Dev
[etomu] - [etom]	93%	5%	1,185 ms	103 ms
[teno] - [temo]	88%	7%	1,087 ms	103 ms
[emuto] - [emto]	83%	12%	1,222 ms	128 ms
[ento] - [emto]	70%	13%	1,241 ms	130 ms
[eton] - [etom]	60%	8%	1,324 ms	133 ms

4. Discussion

The results show that Japanese listeners experience both perceptual epenthesis and perceptual assimilation when repairing sequences that contain illicit coda nasal consonants. Perceptual assimilation is more likely to occur when [m] occurs word-finally than when it occurs word-medially. The word-final coda [m] token, [etom], was both more likely to be categorized as /etoɴ/ and most difficult to discriminate from [eton] suggesting that both word-final [m] and [n] are typically assimilated to /ɴ/. The medial [m] token, [emto], was more likely to be categorized as /emuto/. However, those tokens that were assigned to the /emuto/ category achieved lower Goodness of Fit scores than those assigned to the /eɴto/ category, and participants were more accurate and faster to respond when discriminating between [emuto] and [emto] then they were at discriminating between [ento] and [emto].

We propose that the disparity between categorization rates, Goodness of Fit scores and discriminability is likely due to listeners categorizing tokens that they have recognized as illicit in accordance with loanword assimilation patterns which prefers epenthesis in these contexts [10]. These results suggest that participants favor assimilation over epenthesis in both medial and word-final contexts but that this rate varies between medial and word-final positions.

In terms of PAM-L2 framework, medial coda [m] constitutes a category goodness contrast with both [n] and [mu]. While listeners apparently prefer coda [m] assimilation to epenthesis, it is confused with both. Listeners assign attested tokens to Hiragana labeled categories at a higher rate and with higher Goodness of Fit scores than coda [m] tokens suggesting that while medial coda [m] is being assimilated to both /ɴ/ and /mu/, it is perceived as a less canonical variant when compared with [n] and [mu]. Similarly, word-final coda [m] and [n] constitute a category goodness contrast, where [m] is perceived as a less canonical version of /ɴ/ when compared to [n]. Here, the difference between the two is the smallest of all conditions tested and we find evidence for this in Experiment 1 where [etom] tokens assigned to the /etoɴ/ category achieved the highest goodness of fit of all coda [m] tokens, and in Experiment 2 where the [eton] - [etom] contrast was most difficult to distinguish and required the most amount of time to respond to. However, whether word-final coda [m] and [mu] are a category goodness contrasts is more difficult to ascertain. The [etomu] - [etom] contrast achieved the highest accuracy, higher even than that of the [teno] - [temo] contrast, where both tokens constitute phonotactically attested sequences. However, the relatively high categorisation rate of [etom] to /etomu/ (32%) and Goodness of Fit score (4.04) in Experiment 1 and the time it took for participants to respond to [etomu] - [etom] tests in Experient 2 suggest that some confusion between word-final coda [m] and [mu] is occurring.

Word position appears to be an important factor as participants were significantly better at discriminating between [etomu] - [etom] than [emuto] - [emto] and significantly better at discriminating between [ento] - [emto] than [eton] - [etom]. We propose that this is due to the unconditioned allophonic variation of word-final /ɴ/ [8]. Where coda-position medial [m] is licensed by labial assimilation to the following segment, word-final [m] requires no such licensing, making [etom] a more valid representation of /etoɴ/ than [emto] is of /eɴto/. This makes the perceptual distance between [m] and /ɴ/ smaller when [m] occurs word-finally than when it occurs medially. Perceptual distance modulates the way in which illicit speech is assimilated to phonotactically legal (transitionally probable) sequences, thus making [m] more difficult to distinguish from /mu/ and /ɴ/ word-medially and more difficult to distinguish from /ɴ/ word-finally.

5. Acknowledgments

We wish to thank our participants and the staff at Keio University. We also thank the W. T. Mollison scholarship committee as well as Cathleen Benevanto, Rosey Billington, Katie Jepson and Eleanor Lewis for lending their voices

6. References

[1] Kilpatrick, A., Bundgaard-Nielsen, R., Baker, B. 2018. Japanese co-occurrence restrictions influence second language perception. *Applied Psycholinguistics*. doi:10.1017/S0142716418000711

[2] Hallé, P., Segui, J., Frauenfelder, U., Meunier, C. 1988. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of experimental psychology: Human perception and performance* 24(2), 592-608.
[3] Pintér, G. 2015. The emergence of new consonant contrasts. In Kubozono, H. (Ed.) Handbook of Japanese phonetics and phonology. De Gruyter Mouton, 121-165.
[4] Tsujimura, N. 2013. *An introduction to Japanese linguistics*, 3rd ed., John Wiley and Sons: West Sussex.
[5] 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(6), 1568-1578.
[6] Dupoux, E., Parlato, E., Frota, S., Hirose, Y., Peperkamp, S. 2011. Where do illusory vowels come from? *Journal of memory and language* 64(3), 199-210.
[7] Itô, J. 1989. A prosodic theory of epenthesis. *Natural Language & Linguistic Theory* 7(2), 217-259.
[8] Nogita, A., Yamane, N. 2015. Japanese moraic dorsalised nasal stop. *Phonological studies* 18, 75-84.
[9] Best, T., Tyler, M. 2008. Non-native and second-language speech perception: Commonalities and complementarities. In Munro, J. & Bohn, O. (Eds.), *Second language speech learning: The role of language experience in speech perception and production*, Amsterdam: John Benjamins. 13-34.
[10] Peperkamp, S., Vendelin, I., Nakamura, K. 2008. On the perceptual origin of loanword adaptations: experimental evidence from Japanese. *Phonology* 25, 129-164.