Title: Vowel-coda interaction in spontaneous Japanese utterances

Author: Shigeto Kawahara

Affiliation:

The Institute of Cultural and Linguistic Studies

Keio University

2-15-45 Mita, Minato-ku, Tokyo, JAPAN

Corresponding email: kawahara@icl.keio.ac.jp

Abstract

Japanese is often considered to be mora-timed, and the existence of syllables is sometimes questioned. Evidence for syllables comes from interactions between a vowel and the following consonant. Previous experiments show that vowels get longer before coda consonants. Using the Corpus of Spontaneous Japanese, this study explores to what degree vowels lengthen before geminates and coda nasals. The results demonstrate that (i) vowels do lengthen before geminates and coda nasals, (ii) this lengthening occurs for all types of vowels, (iii) but degrees of lengthening differ by vowel, and (iv) vowels lengthen more before nasals than before geminates.

Keywords: Japanese, vowels, the CSJ, syllables, duration, lengthening

PACS number: 43.70.+i, 43.70.Bk, 43.70.Fq

1 Introduction

Since Trubetzkoy (1939/1969), Japanese has been classified as a "mora-timed" language, and the role of syllables in the language has been questioned from time to time. Labrune (2012*a*,*b*) offers the most recent and explicit version of this proposal, arguing that Japanese has no syllables at all. One basis of this claim comes from the (putative) "[a]bsence of phonetic clues for the existence of a rhyme-like constituent" (Labrune, 2012*b*, p. 120), implying that vowels and coda consonants do not seem to interact in Japanese, either phonetically or phonologically. Kawahara (2016) argues, based on previous phonetic studies, that this claim is empirically false, as vowels and coda consonants do interact phonetically (see also Ito and Mester 2015; Kubozono 2003; Tanaka 2013; Vance 2013 for more counterarguments). For example, many production studies have shown that vowels are longer before geminates than before singletons (Fukui, 1978; Han, 1994; Hirata, 2007; Idemaru and Guion, 2008; Kawahara, 2006, 2013; Kawahara and Braver, 2014; Port et al., 1987). In Idemaru and Guion (2008)[p. 176], arguably the most extensive phonetic study on the acoustic correlates of geminates in Japanese, vowel intervals were on average 75 ms before geminates, whereas they were 59 ms before singletons. Campbell (1999)[pp. 34-35] studied a corpus of Japanese speakers reading 503 phrases and sentences in Japanese newspapers and magazines, and showed that vowels are longer before a coda nasal than they are in open syllables: the median vowel durations were 90 ms for open syllables and 110 ms when closed by a coda nasal.

This paper reexamines these observations in further detail, given that a large corpus of spontaneous speech is now available (as we will see, the corpus contains more than 300,000 tokens of vowels: Kawahara and Shaw 2017). The previous studies were often limited in several aspects; no studies, for example, examined the interaction between vowels and coda consonants for each vowel type separately, often studying only a subset of the full list of Japanese vowels. Kawahara (2006) for example studied only non-high vowels, and collapsed their results; Idemaru and Guion (2008) studied only /e/. Moreover, most studies, except for Campbell (1999), are based on lab speech rather than spontaneous speech. (It is not meant to imply here that lab speech is unreliable for phonetic analyses, but it is interesting and important to examine natural speech as well.) The corpus that the current study deploys, the Corpus of Spontaneous Japanese (the CSJ: Maekawa 2003), is a large corpus of spoken Japanese, which allows us to overcome these limitations, thereby shedding new light on the vowel-coda interaction in spoken Japanese.

A brief background discussion of the syllable structure in Japanese is in order. Japanese has five vowels, /a/, /e/, /o/, /i/, and /u/. The coda consonants are limited to either nasal consonants (represented traditionally as /N/) or the first part of obstruent geminates (represented as /Q/). Japanese has a short vs. long contrast in vowels (Hirata, 2004; Hirata and Tsukada, 2009), but long vowels rarely occur before a coda consonant (Kubozono, 2003; Kubozono et al., 2008). This study examines the duration of each of the five vowels in open syllables as well as in syllables closed with /N/ and /Q/. Japanese is a pitch accent language, whose effects on duration are minimum (Beckman, 1986; Homma, 1981), and hence not taken into account in this study. See Vance (1987, 2008) for further details on the basic sound patterns of Japanese.

2 Method

The analysis is based on the Corpus of Spontaneous Japanese (the CSJ: Maekawa et al. 2000; Maekawa 2003, 2015), one of the largest annotated speech corpora of Japanese. The CSJ contains several speech styles, including, but not limited to, Academic Presentation Style (APS) and Spontaneous Presentation Style (SPS). The former is based on real academic presentations, whereas the latter is based on solicited monologue, in which speakers were given a few topics as prompts. The gender in the corpus is more or less balanced, although there are slightly more male speakers than female speakers. The current analysis used the core portion of the corpus (known as the CSJ-RDB), which comes with rich annotation and phonetic duration information. The CSJ-RDB consists of 11,559 unique words which were produced by 70 speakers. The CSJ contains a hand-coded annotation tier, in which duration of each sound is specified. See `http://pj.ninjal.ac.jp/corpus_center/csj/en/` for further details of this corpus and the CSJ-RDB.

For the current study, duration of each vowel interval was extracted from the CSJ-RDB text file, and sorted by their syllabic environments, based on the annotations provided by the CSJ, which uses the traditional labels /N/ (a coda nasal) and /Q/ (a obstruent geminate) to mark coda consonants. Long vowels were excluded from this analysis, as long vowels rarely, if ever, occur in closed syllables (Kubozono, 2003; Kubozono et al., 2008). Within each vowel quality in the three separate environments, outliers—those that were more than 2 standard deviations away from the mean—were excluded.

3 Results

The boxplot in Figure 1 (provided at the end of the paper) shows the durational properties of each of the five vowels in three different syllabic environments; open syllable, before an obstruent geminate (=/Q/), and before a coda nasal (=/N/). Table 1 shows medians, means, and standard deviations around the means. Standard deviations are generally large, even after excluding outliers. Figure 1 also shows that the distributions are right-skewed, having outliers at the top; therefore medians should be more reliable measures than means for the current analysis.

Table 1: Duration of the five Japanese vowels in the three syllabic environments. Median, mean (SD) values (ms), and the number of relevant tokens.

	Open syll.			Pre-/Q/			Pre-/N/		
	median	mean (SD)	n	median	mean (SD)	n	median	mean (SD)	n
/a/	72.2	76.0 (28.0)	95,532	73.9	74.9 (22.3)	2,859	83.9	85.5 (24.6)	5,770
/e/	63.3	72.9 (38.4)	50,431	71.1	74.4 (24.4)	787	81.4	83.6 (26.4)	4,002
/o/	60.6	65.5 (28.2)	81,386	65.9	66.9 (20.7)	1,052	81.7	84.1 (27.3)	2,757
/i/	48.1	52.8 (23.4)	66,624	55.4	57.6 (19.4)	1,673	63.2	66.9 (25.4)	2,237
/u/	44.2	48.8 (23.3)	44,610	57.9	59.7 (22.6)	590	66.6	67.9 (24.6)	2213

As previous studies have shown, the median and mean duration of the five vowels follow the order of /a/ $>$ /e/ $>$ /o/ $>$ /i/ $>$ /u/ in Japanese (Arai et al., 2001; Campbell, 1992, 1999; Campbell and Isard, 1991; Han, 1962; Sagisaka and Tohkura, 1984; Sagisaka, 1985), and this order generally holds across the all three syllabic environments (except for the reversal between /i/ and /u/ in closed syllables, and the reversal between /e/ and /o/ in the pre-/N/ environment). Within each vowel, vowels are generally shortest in open syllables, slightly longer before a geminate, and longest before a coda nasal. These results replicate the general lengthening effects in closed syllables identified in the previous phonetic studies (Campbell, 1999; Fukui, 1978; Han, 1994; Hirata, 2007; Idemaru and Guion, 2008; Kawahara, 2006, 2013; Kawahara and Braver, 2014; Port et al., 1987). The only reversal is /a/ in the pre-/Q/ environment in terms of mean, but not in terms of median.

Table 2 shows the differences, in terms of median and mean, between open syllables and the two types of closed syllables. Since the distributions are right-skewed, the statistical significance of the differences was accessed by a non-parametric Wilcoxon rank-sum test (also known as a Mann-Whitney test). The

results show that lengthening before an obstruent geminate and lengthening before a coda nasal are both statistically significant at the $p < .001$ level, for all types of vowels.

Table 2: Differences in vowel durations between open syllables and the two types of closed syllables. Statistical significance was tested by a non-parametric Wilcoxon rank-sum test.

	Pre-/Q/			Pre-/N/		
	median	mean	Wilcox test	median	mean	Wilcox test
/a/	1.7	-1.0	$p < .001$	11.7	9.5	$p < .001$
/e/	7.7	1.6	$p < .001$	18.0	10.8	$p < .001$
/o/	5.4	1.4	$p < .001$	21.1	18.7	$p < .001$
/i/	7.3	4.8	$p < .001$	15.1	14.1	$p < .001$
/u/	13.8	10.9	$p < .001$	22.4	19.2	$p < .001$

Since we observe that lengthening is more extensive before a coda nasal than before a geminate, this difference was accessed for each vowel, again with a non-parametric Wilcoxon test. The differences are significant at $p < .001$ level for all of the vowels: vowels are statistically longer before nasals than before obstruent geminates.

We also observe that different vowels lengthen to different degrees. Shorter vowels (e.g. /u/) tend to be affected more, and longer vowels tend to be affected less (e.g. /a/). It may be the case that there is a limit on how long a short vowel can be—given that Japanese has a short/long contrast (Hirata, 2004; Hirata and Tsukada, 2009)—so that a vowel like /a/, which is already long, cannot be lengthened too much.

One prominent difference between the current result and the previous studies is the magnitude of difference between vowels in open syllables and vowels before obstruent geminates. As reviewed in the introduction, Idemaru and Guion (2008) found that vowels get longer on average by 16 ms before an obstruent geminate. Their stimuli only contained [e]; in Table 2, the differences that the current study found for /e/ were 1.6 ms in terms of means and 7.7 ms in terms of medians. Kawahara (2006), like Idemaru and Guion (2008), found that pre-geminate lengthening can on average be as long as 20 ms, using the stimuli containing /a/, /e/, and /o/. The largest difference that the current study found in terms of means is 1.6 ms for /e/, or 7.7 for /e/ ms in terms of medians. These comparisons indicate that pre-geminate lengthening may be exaggerated in lab speech, and is less clearly observed in spontaneous speech.

4 Discussion

The results confirm the previous findings, mostly based on lab-speech, that vowels are longer before obstruent geminates than in open syllables. The results also confirm the results of Campbell (1999) that vowels are longer before a coda nasal than in open syllables. Not only does the current study replicate these findings based on a very large spoken-speech corpus, it offers several new findings. First, the data on pre-geminate lengthening has mainly come from lab-read speech, and this effect was replicated in spontaneous speech. Second, it demonstrates that lengthening before a geminate and lengthening before a coda nasal hold for each of the five vowels in Japanese. These findings are important in supporting the role of syllables in the prosodic organization of Japanese (Kawahara, 2016), as they instantiate the phonetic interaction between a vowel and a coda consonant, which is not predicted by the syllable-less theory of Japanese (Labrune, 2012*a*,*b*).

Third, the current study found that lengthening before a coda nasal is more extensive than lengthening before a geminate, which is a new finding that no previous studies have found. Even Campbell (1999), who reported both pre-geminate and pre-nasal lengthening, did not find this systematic difference. Fourth, the current study shows that vowels lengthen to different degrees; [a] seems least susceptible to closed-syllable lengthening, whereas [u] is most "stretchable" (see Kawahara and Shaw 2017 for a similar observation). Overall, we conclude that vowels and coda consonants in Japanese do interact phonetically, despite the claim to the contrary by Labrune (2012*a*,*b*), although the way they interact is more complex than previously thought: the degree of lengthening depends on which vowel interacts with which consonant.

References

Arai, T., Warner, N., and Greenberg, S. (2001), "OGI tagengo denwa onsei koopasu-ni okeru nihongo shizen hatsuwa onsei no bunseki [Analysis of spontaneous Japanese in OGI multi-language telephone speech corpus]," *Nihon Onkyoo Gakkai Shunki Happyoukai vol.1 [The Spring Meeting of the Acoustical Society of Japan]*, pp. 361–362.

Beckman, M. (1986), *Stress and Non-Stress Accent*, Dordrecht: Foris.

Campbell, N. (1992), "Segmental elasticity and timing in Japanese," in *Speech Perception, Production and Linguistic Structure*, eds. Y. Tohkura, E. V. Vatikiotis-Bateson, and Y. Sagisaka, Tokyo: Ohmsha, pp. 403–418.

Campbell, N. (1999), "A study of Japanese speech timing from the syllable perspective," *Onsei Kenkyu [Journal of the Phonetic Society of Japan]*, 3(2), 29–39.

Campbell, N., and Isard, S. (1991), "Segment durations in a syllable frame," *Journal of Phonetics*, 19, 37–47.

Fukui, S. (1978), "Nihongo heisaon-no encho/tanshuku-niyoru sokuon/hisokuon toshite-no chooshu [Per-

ception for the Japanese stop consonants with reduced and extended durations]," *Onsei Gakkai Kaihou [Phonetic Society Reports]*, 159, 9–12.
Han, M. (1962), "The feature of duration in Japanese," *Onsei no Kenkyuu [Studies in Phonetics]*, 10, 65–80.
Han, M. (1994), "Acoustic manifestations of mora timing in Japanese," *Journal of the Acoustical Society of America*, 96, 73–82.
Hirata, Y. (2004), "Effects of speaking rate on the vowel length distinction in Japanese," *Journal of Phonetics*, 32(4), 565–589.
Hirata, Y. (2007), "Durational variability and invariance in Japanese stop quantity distinction: Roles of adjacent vowels," *Onsei Kenkyu [Journal of the Phonetic Society of Japan]*, 11(1), 9–22.
Hirata, Y., and Tsukada, K. (2009), "Effects of speaking rate and vowel length on formant frequency displacement in Japanese," *Phonetica*, 66(3), 129–149.
Homma, Y. (1981), "Durational relationship between Japanese stops and vowels," *Journal of Phonetics*, 9, 273–281.
Idemaru, K., and Guion, S. (2008), "Acoustic covariants of length contrast in Japanese stops," *Journal of International Phonetic Association*, 38(2), 167–186.
Ito, J., and Mester, A. (2015), "Sino-Japanese phonology," in *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. H. Kubozono, Berlin: Mouton, pp. 289–312.
Kawahara, S. (2006), "A faithfulness ranking projected from a perceptibility scale: The case of [+voice] in Japanese," *Language*, 82(3), 536–574.
Kawahara, S. (2013), "Emphatic gemination in Japanese mimetic words: A wug-test with auditory stimuli," *Language Sciences*, 40, 24–35.
Kawahara, S. (2016), "Japanese has syllables: A reply to Labrune (2012)," *Phonology*, 33.
Kawahara, S., and Braver, A. (2014), "Durational properties of emphatically-lengthened consonants in Japanese," *Journal of International Phonetic Association*, 44(3), 237–260.
Kawahara, S., and Shaw, J. (2017), "Effects of consonant-conditioned informativity on vowel duration in Japanese,", Ms. Keio University [Revision submitted to Language and Speech].
Kubozono, H. (2003), "The syllable as a unit of prosodic organization in Japanese," in *The Syllable in Optimality Theory*, eds. C. Fery, and R. van der Vijver, Cambridge: Cambridge University Press., pp. 99–122.
Kubozono, H., Ito, J., and Mester, A. (2008), "Consonant gemination in Japanese loanword phonology," in *Current Issues in Unity and Diversity of Languages. Collection of Papers Selected from the 18th International Congress of Linguists*, ed. The Linguistic Society of Korea, Republic of Korea: Dongam Publishing Co, pp. 953–973.
Labrune, L. (2012*a*), *The phonology of Japanese*, Oxford: Oxford University Press.
Labrune, L. (2012*b*), "Questioning the universality of the syllable: Evidence from Japanese," *Phonology*, 29(1), 113–152.
Maekawa, K. (2003), "Corpus of Spontaneous Japanese: Its Design and Evaluation," *Proceedings of ISCA and IEEE Workshop on Spontaneous Speech Processing and Recognition (SSPR2003)*, pp. 7–12.
Maekawa, K. (2015), "Corpus-based studies," in *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. H. Kubozono, Berlin: Mouton, pp. 651–680.
Maekawa, K., Koiso, H., Furui, S., and Isahara, H. (2000), "Spontaneous speech corpus of Japanese," *Proceedings of the Second International Conference of Language Resources and Evaluation*, pp. 947–952.
Port, R., Dalby, J., and O'Dell, M. (1987), "Evidence for mora timing in Japanese," *Journal of the Acoustical Society of America*, 81, 1574–1585.
Sagisaka, Y. (1985), Onsei Gousei-no Tame-no Inritsu Seigyo-no Kenkyuu [A Study on Prosodic Controls for Speech Synethesis], Doctoral dissertation, Waseda University.

Sagisaka, Y., and Tohkura, Y. (1984), "Kisoku-niyoru onsei goosei-no tame-no onin jikan seigyo [Phoneme duration control for speech synthesis by rule]," *Denshi Tsuushin Gakkai Ronbunshi*, 67, 629–636.
Tanaka, S.-I. (2013), "Review of Labrune, Laurence (2012) *The Phonology of Japanese*, Oxford University Press, 320p," *Onsei Kenkyu [Journal of the Phonetic Society of Japan]*, 17(1), 70–80.
Trubetzkoy, N. S. (1939/1969), *Grundzüge der phonologie*, Güttingen: Vandenhoeck and Ruprecht [Translated by Christiane A. M. Baltaxe 1969, University of California Press].
Vance, T. (1987), *An Introduction to Japanese Phonology*, New York: SUNY Press.
Vance, T. (2008), *The Sounds of Japanese*, Cambridge: Cambridge University Press.
Vance, T. (2013), "Book review: The Phonology of Japanese," *Lingua*, 123, 168–174.

Figure 1: Boxplots showing the durations of each of the five vowels in Japanese in three syllabic environments.