Title: Durational compensation within a CV mora in spontaneous Japanese: Evidence from the Corpus of Spontaneous Japanese

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

Note to the Lingbuzz version: Thanks to Michael Becker for pointing out my embarrassing mistake for my first submission (my CV was mistakenly uploaded…) and encouraging me to make Figure 3 more informative.

Abstract

Previous experimental studies showed that in Japanese, vowels are longer after shorter onset consonants; there is durational compensation within a CV-mora. In order to address whether this compensation occurs in natural speech, this study re-examines this observation using the Corpus of Spontaneous Japanese. The results, which are based on about 200,000 tokens, show that there is a negative correlation between the onset consonant and the following vowel, which is shown to be significant by a bootstrap resampling analysis. The compensation was not perfect, however, suggesting that it is a stochastic tendency rather than an absolute principle.

Keywords: Japanese, vowels, the CSJ, duration, compensation, mora-timing

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

1 Introduction

One of the phonetic characteristics of Japanese is a duration compensation effect within a CV-mora, which is sometimes taken to be evidence for mora-timing—a CV unit functions as a synchronous rhythmic unit in Japanese. More concretely, previous studies have shown that after longer consonants, vowels tend to get shorter (Port et al., 1980, 1987). Port et al. (1980) used CVCV stimuli by varying the medial consonant (/s, t, d, r/) and showed that before a short consonant, the following vowel gets longer. Likewise, Port et al. (1987), again using CVCV stimuli, systematically varied the second consonant using /k, g, t, d, s, z/ and found that different duration of these consonants is compensated for by adjusting the following vowel duration. Minagawa-Kawai (1999) compared Japanese, Korean, and Chinese using /r, b, s/ and showed that degrees of durational compensation are larger for Japanese than for Korean or Chinese. See also Otake (1988), Otake (1989), and Sagisaka and Tohkura (1984) for similar results; see Warner and Arai (2001) for a critical review of these studies, in particular, about how the observed compensation effects may or may not speak for mora-timing nature to Japanese. See also Beckman (1982).

The current study aims to expand the scope of the previous studies in various aspects. First, this study addresses the question of whether this durational compensation within a CV mora occurs in natural speech in addition to read-speech in the lab. While read-speech in the lab offers a critical data set for phonetic theorization and modeling, it is important and interesting to confirm a particular pattern using more naturalistic speech. Especially, the studies by Port et al. (1980, 1987) used only small sets of stimuli, which are mixtures of real words and nonce words. Addressing the compensation effects with more realistic Japanese words is warranted. Second, by using a large corpus, this study tests all types of consonants in Japanese, beyond those that are tested by the studies reviewed above (see also Sagisaka and Tohkura 1984). Third, Port et al. (1980, 1987) tested only /a/ and /u/, whereas Minagawa-Kawai (1999) tested only /a/ and /i/. The current study, by using a large corpus, takes into account all the types of vowels that appear in Japanese. Finally, by testing a large number of tokens, the current study statistically examines the robustness of this compensation effects. Moreover, the current paper deploys a bootstrapping resampling method to assess the statistical likelihood of the observed compensation effects.

2 Method

The empirical analysis is based on the Corpus of Spontaneous Japanese (the CSJ: Maekawa et al. 2000; Maekawa 2003, 2015). Its core, annotated portion—the CSJ-RDB—consists of more than 1,000,000 segmental intervals, with each interval annotated with its duration. More specifically, it contains more than 300,000 vowel tokens, which allows us to perform various types of analyses with a large number of data points (Kawahara, 2017; Kawahara and Shaw, 2017). The CSJ-RDB consists of natural speech produced by 70 speakers. The CSJ contains several speech styles, including, but not limited to, Academic Presentation Style and Spontaneous Presentation Style. The former is real academic presentations; the latter is solicited monologue, in which speakers were given a few topics as prompts. The gender of the speakers 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). The CSJ-RDB contains a hand-coded annotation tier, in which duration of each sound is specified. Further details of the CSJ corpus can be found at `http://pj.ninjal.ac.jp/corpus_center/csj/en/`. The details of the segment procedure can be found in the document which is downloadable at `http://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/06.pdf` (written in Japanese: Kawahara and Shaw (2017) offer a translation of the segmental procedure between a glide and a vowel.)

Given the CSJ-RDB textfile, for oral stops, based on the annotation, all of the intervals that are annotated as "<cl>" (for closure), were extracted. If a <cl> interval is preceded by a "Q" interval, it means that that stop consonant is a long consonant, which was systematically excluded from the current analysis. Based on these procedures, the duration profiles of /p, t, k, b, d, g/ were extracted. /t/ and (some of) /d/ are affricated in Japanese (Vance, 1987, 2008). In the CSJ-RDB, affricates are coded as different from stops, which were excluded because the phonemic status of affricates in Japanese is not very clear.

The current study also targeted nasals (/m, n/) and continuants (/s, z, h, r, w, y/, where /y/ is a palatal glide, not a front rounded vowel, a convention that is used in the CSJ). Their non-geminate versions were extracted together with the following vowel duration. Phonological secondary palatalization, as well as phonetic palatalization due to the following /i/, were abstracted away from in the current analysis; for example, "b" and "by" (phonologically palatalized) and "bj" (phonetically palatalized) were all collapsed into one category, /b/. This choice is to be conservative: it would not be very surprising /bV/ and /byV/ show comparable total CV-mora duration. Increasing the number of consonant types, with repetitions of arguably non-independent samples, would increase the Type I error. For the same reason, "h", "hj" and "F" (the last label represents a bilabial fricative, an allophone of /h/ before /u/) were also collapsed.

As for the analysis of vowels, all the intervals labeled as "a", "i", "u", "e", and "o", following the target consonants were extracted. However, phonologically long vowels—those that are followed by an interval with "H"—were excluded, as their frequencies are incomparably smaller than those of phonologically short vowels (less than 10%). Vowels in closed syllables were also excluded, as we know from the previous work that vowels get longer in closed syllables than in open syllables (Han, 1994; Hirata, 2007; Idemaru and Guion, 2008; Kawahara, 2006; Port et al., 1987). The remaining Ns are as follows: /p/=523, /t/=26,196, /k/=27,754, /b/=3,288, /d/=15,673, /g/=10,994, /s/=26,434, /z/=5,949, /h/=5,672, /m/=12,807, /n/=31,938, /r/=17,154, /w/=7,856, and /y/=7,102. (/p/ is severely underrepresented because Japanese lost /p/ in its history, and singleton /p/ appears only in recent loanwords: Ito and Mester 2008.) The total N is 199,340.

3 Result

Figure 1 illustrates the combined duration of each type of consonant and the following vowel's duration in terms of median. Median values are more appropriate than means to use for the case at hand, because the distributions of these values are right skewed, as shown in Figure 2, an illustrative boxplot showing the right-skewed distribution of consonantal and vowel durations in /pV/, /gV/, /sV/ moras (see also Kawahara 2017; Kawahara and Shaw 2017 for vowel duration analyses of the CSJ-RDB, which show the same pattern of skew). Actual median values and mean values are provided in Tables 1 and 2 in the Appendix.

[xxx Figure 1 here xxx]

[xxx Figure 2 here xxx]

First, focusing on the behavior of consonants, voiced obstruents are generally shorter than voiceless obstruents, which has been found in a previous acoustic experiment (Kawahara, 2006), as well as in cross-linguistic patterns (e.g. Diehl and Kluender 1989; Kingston and Diehl 1994; Lisker 1957; Ohala 1983). Within oral stops and nasal stops, there is a general tendency in which the more front the place of articulation, the longer the oral stop—compare, e.g. /t/ vs /k/, /b/ vs. /d, g/, /m/ vs. /n/ (Homma, 1981; Kawahara and Shaw, 2017). Third, we also observe that fricatives are in general longer than oral stops, again a tendency that holds cross-linguistically, including Japanese (Kawahara, 2015; Lehiste, 1970; Sagisaka and Tohkura, 1984). /r/, which is a flap in Japanese (see Arai 2013 for detail), is short, being around 30 ms in terms of median.

[xxx Figure 3 here xxx]

Now moving on to the correlation between vowel duration and consonant duration, we observe that there is a negative correlation between them ($r = -0.58, t(12) = 2.45, p < .05$), in such a way that vowels are shorter after longer consonants, as shown visually by the scatterplot in Figure 3. For example, /s/ is the longest consonant of all, and the following vowel is the shortest. /g/ is the shortest consonant of all, and the following vowel is the longest. A comparison between /m/ and /n/ illustrates the compensation effect clearly—/m/ is longer than /n/, but the following vowel is shorter after /m/ than after /n/, and the result is that /mV/ and /nV/ show comparable duration profiles.

However, the compensation effect is not perfect. For example, /p/ and /t/ show comparable duration profiles, but the following vowels are longer after /t/ than after /p/. Similarly, /d/ and /g/ show comparable duration, but the vowels are longer after /g/ than after /d/. Although /r/ is a short consonant, the following vowel does not get as long as it can get. /y/ behaves similarly: the following vowel could have become longer (e.g. as long as post-/g/ vowels) so that the entire /yV/ mora is more comparable in duration with the moras with other onset consonants.

In order to more rigorously assess the statistical significance of the durational compensation—beyond a correlation analysis between consonant duration and vowel duration—a bootstrap method was deployed (Efron and Tibshirani, 1993). The standard deviation across the 14 consonantal conditions serves as the measure of the degree to which the entire CV mora duration is kept constant. The actual standard deviation is 12.17 ms across the 14 different conditions. In the bootstrap method, first one consonant interval and one vocalic interval was randomly sampled and their duration was combined. This process was reiterated 14 times without replacement to create 14 CV combinations, and the standard deviation of these samples was calculated. This process was reiterated 50,000 times to obtain 95% and 99% confidence intervals. The whole process was automated by using R (R Development Core Team, 1993–). The results are 12.80 ms - 22.08 ms (95%) and 11.17 ms - 22.92 ms (99%). Since the observed standard deviation is outside of the 95% confidence interval, but within the 99% confidence interval, this result suggest that the durational compensation effect is significant at the $p < .05$ level.

4 Summary and discussion

This paper has shown with a large-scale corpus of spoken Japanese that in Japanese, vowel duration varies in response to the duration of the preceding consonant: the shorter the consonant, the longer the vowel tends to be. The bootstrap resampling analysis has shown that Japanese adjusts the duration of a CV mora unit in such a way that its variability is lower than it could have occurred by chance. This finding supports the previous experimental findings about durational compensation with a large number of natural speech tokens.

However, we also found that duration compensation is not perfect. Vowel duration can differ between two consonants whose duration profiles are comparable; vowels sometimes do not get as long as they could have been, so that the resulting mora's duration is more similar to the duration of other moras. It therefore seems safe to conclude that durational compensation is a stochastic tendency rather than an absolute principle. It is likely the case that there are other principles at work in regulating the duration of Japanese vowels. For example, Kawahara and Shaw (2017) show that average predictability of the vowel given the preceding consonant, quantified in terms of Shannon's entropy ($H(V|C) = \sum p(v_i|C) \times -\log_2 p(v_i|C)$: Shannon 1948), can impact the duration of some vowels in Japanese. Thus, exploring the interaction of the durational compensation effect and other principles, like predictability effects, offers an interesting opportunity for future research work.

Appendix: Median and mean values

Table 1: Actual median values

	p	t	k	b	d	g	s	z	h	m	n	r	w	y
cons	46.6	47.2	40.6	26.9	19.0	18.0	70.7	46.2	53.2	57.9	45.7	28.2	37.4	43.5
vowel	57.0	62.3	53.7	72.5	68.8	78.1	47.8	61.3	53.7	67.9	76.5	61.6	78.0	50.1
total	103.7	109.4	94.4	99.4	87.8	96.2	118.5	107.5	106.9	125.8	122.2	89.8	115.3	93.6

Table 2: Mean values. $r = -0.60, t(12) = -2.60, p < .05$.

	p	t	k	b	d	g	s	z	h	m	n	r	w	y
cons	47.3	49.3	42.2	32.1	23.2	21.2	72.9	48.2	58.8	58.5	47.5	29.7	39.3	49.8
vowel	59.2	76.1	60.8	79.4	92.2	92.9	55.2	68.5	63.3	79.2	92.1	71.2	95.8	55.4
total	106.5	125.4	103.0	111.5	115.5	114.2	128.1	116.7	122.1	137.7	139.6	100.9	135.1	105.2

References

Arai, T. (2013), "On why Japanese /r/ sounds are difficult for children to acquire," *Proceedings of INTERSPEECH 2013*, pp. 2445–2449.

Beckman, M. (1982), “Segmental duration and the ‘mora’ in Japanese,” *Phonetica*, 39, 113–135.

Diehl, R., and Kluender, K. (1989), “On the objects of speech perception,” *Ecological Psychology*, 1, 121–144.

Efron, B., and Tibshirani, R. J. (1993), *An Introduction to Bootstrapping*, Boca Raton: Chapman and Hall/CRC.

Han, M. (1994), “Acoustic manifestations of mora timing in Japanese,” *Journal of the Acoustical Society of America*, 96, 73–82.

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.

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. (2008), “Lexical classes in phonology,” in *The Oxford Handbook of Japanese Linguistics*, eds. S. Miyagawa, and M. Saito, Oxford: Oxford University Press, pp. 84–106.

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. (2015), “The phonetics of *sokuon*, or obstruent geminates,” in *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. H. Kubozono, Berlin: Mouton, pp. 43–73.

Kawahara, S. (2017), “Vowel-coda interaction in spontaneous Japanese utterances,”, Ms. Keio University.

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].

Kingston, J., and Diehl, R. (1994), “Phonetic knowledge,” *Language*, 70, 419–454.

Lehiste, I. (1970), *Suprasegmentals*, Cambridge: MIT Press.

Lisker, L. (1957), “Closure duration and the intervocalic voiced-voiceless distinction in English,” *Language*, 33, 42–49.

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.

Minagawa-Kawai, Y. (1999), “Preciseness of temporal compensation in Japanese timing,” *Proceedings of ICPhS*, pp. 365–368.

Ohala, J. J. (1983), “The origin of sound patterns in vocal tract constraints,” in *The Production of Speech*, ed. P. MacNeilage, New York: Springer-Verlag, pp. 189–216.

Otake, T. (1988), “A temporal compensation effect in Arabic and Japanese,” *Bulletin of the Phonetic Society of Japan*, 189, 19–24.

Otake, T. (1989), “A cross-linguisic contrast in the temporal compensation effect,” *Bulletin of the Phonetic Society of Japan*, 191, 14–19.

Port, R., Al-Ani, S., and Maeda, S. (1980), “Temporal compensation and universal phonetics,” *Phonetica*, 37, 235–252.

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.

R Development Core Team (1993–), *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria.

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.

Shannon, C. (1948), A mathematical theory of communication,, MA Thesis, MIT.

Vance, T. (1987), *An Introduction to Japanese Phonology*, New York: SUNY Press.

Vance, T. (2008), *The Sounds of Japanese*, Cambridge: Cambridge University Press.

Warner, N., and Arai, T. (2001), “Japanese Mora-Timing: A Review,” *Phonetica*, 58, 1–25.

Figure 1: Duration of CV units with different onset consonants, based on median.

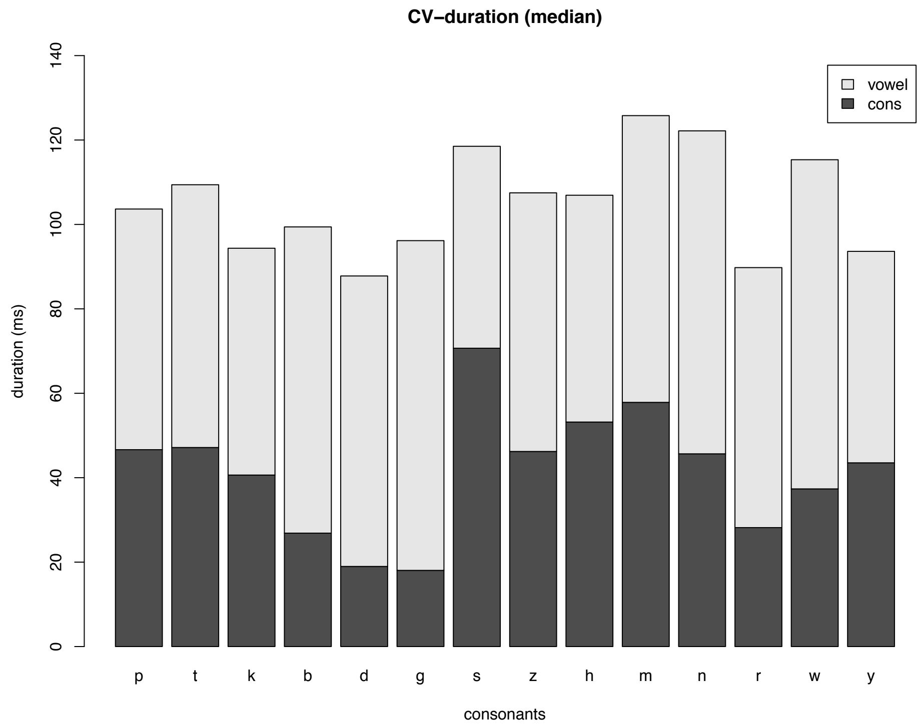

Figure 2: The distribution of consonant duration and vowel duration for /pV/, /gV/ and /sV/.

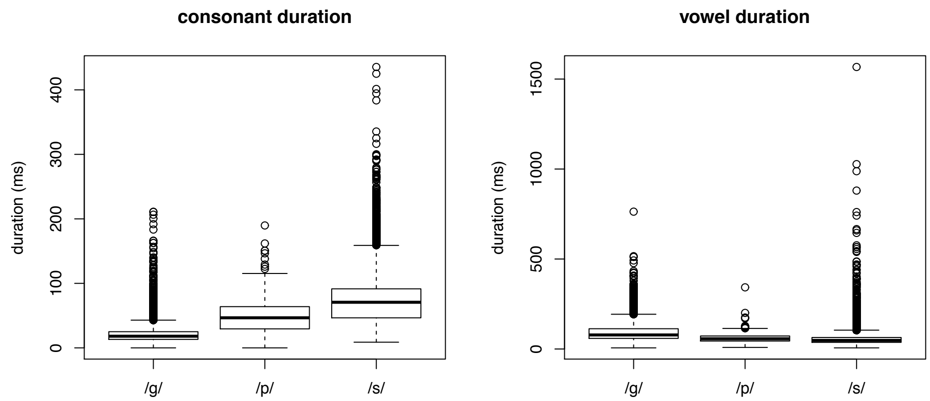

Figure 3: The scatterplot showing the negative correlation between consonant duration and vowel duration. The linear regression line is also shown.