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`

Acknowledgements: Thanks to an anonymous reviewer, Martin Cooke, Donna Erickson, Josef Fruehwald, Shin-ichiro Sano, Jason Shaw, Helen Stickney, and Andy Wedel for comments and analytical help on this work. This research is supported by JSPS grants #15F15715, #26284059 and #17K13448. Remaining errors are mine.

Journal of the Acoustical Society of America (2017) 142,
`http://dx.doi.org/10.1121/1.4994674`

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 CSJ). The results, which are based on more than 200,000 CV-mora tokens, show that there is a negative correlation between the onset consonant and the following vowel in terms of their duration. The statistical significance of this negative correlation is assessed by a traditional correlation analysis as well as a bootstrap resampling analysis, which both show that it is unlikely that the observed compensation effect occurred by chance. The compensation is not perfect, however, suggesting that it is a stochastic tendency rather than an absolute principle. This paper closes with discussion of potential factors that may interact with the durational compensation effect.

Keywords: Japanese; vowel duration; compensation; mora timing; the CSJ

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

1 Introduction

One of the phonetic characteristics of Japanese is a durational compensation effect within CV-moras, which is sometimes taken to be evidence for mora-timing—a CV unit functions as a synchronous rhythmic unit in Japanese (see Otake 2015 for a recent review). 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 after 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 durations of these consonants are 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 and 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 effect may or may not constitute evidence for the mora-timing nature of Japanese. See also Beckman (1982) for a critical evaluation of the notion of mora-timing in Japanese.

The current study aims to expand the scope of the previous studies in various aspects. First, the current 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 there is no doubt that read-speech obtained in the lab offers critical data sets for phonetic theorization and modeling, it is important and interesting to confirm a particular pattern using more naturalistic speech (see Xu 2010 for relevant discussion). In particular, 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 effect 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 were tested by the studies reviewed above (see also Sagisaka and Tohkura 1984 who tested a large set of consonants). 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 effect. Moreover, the current paper deploys a bootstrapping resampling method to estimate the statistical likelihood of the observed compensation effect.

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, 2018; Shaw and Kawahara, 2017). Using the entirety of the CSJ-RDB, this study analyzed natural speech produced by 201 speakers. The CSJ contains several speech styles, including, but not limited to, Academic Presentation Style and Spontaneous Presentation Style. The former is from real academic presentations; the latter is solicited monologue, in which speakers were given a few topics as prompts and spoke in front of a few listeners. The gender of the speakers in the corpus is more or less balanced, although there are slightly more male speakers than female speakers. The CSJ-RDB contains a hand-coded annotation tier, in which duration of each sound is specified. Further details of the CSJ can be found at `http://pj.ninjal.ac.jp/corpus_center/csj/en/`. The details of the segmentation procedure can be found in the document which is downloadable at `http://pj.ninjal.ac.jp/corpus_center/csj/k-report-f/06.pdf` (this document is written in Japanese: Shaw and Kawahara 2017 offer a translation of the segmentation 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. The duration of the following burst interval was added to the duration of <cl> in order to estimate the duration of the entire stop. If a

<cl> interval is preceded by a "Q" interval, it means that the stop consonant is a long consonant (a.k.a. geminates)—these were systematically excluded from the current analysis. Based on these procedures, the duration profiles of /p/, /t/, /k/, /b/, /d/, /g/ were calculated. /t/ is affricated as /t͡ɕ/ before /i/ in native Japanese words (annotated in the CSJ as "c"), and as /t͡s/ before /u/ (Vance, 1987, 2008). Stops and affricates were treated as different categories, however, because their distributions are not complementary in contemporary Japanese: /t͡ɕ/ can appear before vowels other than /u/ (Pintér, 2015). The current study also targeted nasals (/m/, /n/) and continuants (/ɸ/, /s/, /z/, /h/, /r/, /w/, /y/, where /ɸ/ is a bilabial fricative, shown as "f" and /y/ is a palatal glide, not a front rounded vowel—these are conventions used in the CSJ). Their non-geminate versions were extracted together with the following vowel duration.

Phonologically palatalized consonants were treated as separate categories from their plain counterparts, because they are contrastive; for example, "b" and "by" were treated as separate phonemes. On the other hand, phonetic palatalization due to the following /i/, was abstracted away in the current analysis; for example, "b" and "bj" (phonetically palatalized /b/) were collapsed into one category, /b/—this was necessary because, for example, "bj" appears before /i/ and "b" appears elsewhere.

As for the analysis of vowels, all the intervals labeled as "a", "i", "u", "e", and "o" following the target consonants were extracted. Phonologically long vowels—those that are followed by an interval labelled with "H" in the CSJ—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, 2018; Port et al., 1987). This means that any vocalic intervals followed by "Q" (coda obstruent) or "N" (coda nasal) were eliminated from the analysis.

After these processes, consonants that occurred less than 100 times were excluded from the following analysis, as their duration estimates may not be accurate. Those included phonologically palatalized voiced stops and palatalized nasal consonants. The Ns of the remaining CV-moras were as follows: /pV/ = 426, /tV/ = 26,811, /cV/ = 3,161, /kV/ = 26,667, /kyV/ = 119, /bV/ = 3,345, /dV/ = 16,248, /gV/ = 11,302, /sV/ = 26,422, /syV/ = 1,506, /zV/ = 4,736, /zyV/ = 1,006, /hV/ = 3,123, /fV/ = 596, /mV/ = 12,816, /nV/ = 32,392, /rV/ = 20.203, /ryV/ = 177, /wV/ = 8,431, and /yV/ = 2,012.[1] The total N is 201,614.

To normalize the effect of speaking rate that is likely to differ across speakers, the duration data was normalized for each speaker using the following formula:

$$\text{norm}_{ij} = \frac{\text{raw}_{ij} - \text{min}_j}{\text{max}_j - \text{min}_j} \tag{1}$$

where j represents each speaker, and i represents each token. In this normalization method, the denominator defines "the duration range" that a particular speaker uses, which reflects his/her speaking rate. The numerator defines the distance between a particular token and its minimum duration. This way of normalization has an advantage over z-transformation in that we do not need to deal with negative numbers; in fact, this method has been used by other linguistic work in order to wash away inter-speaker variability (e.g. Kawahara and Shinya 2008; Truckenbrodt 2004).

3 Result

Figure 1 illustrates the combined duration of each type of consonant and the following vowel duration in terms of a median value. Median values are arguably more appropriate than mean values to use in the case at hand, because the distributions of these values are right skewed. The skewed distributions can be seen in Figure 2, which contains illustrative histograms showing the distribution of consonantal durations of /g/, /p/, and /m/ (see also Kawahara 2018; Shaw and Kawahara 2017 for vowel duration analyses of the CSJ-RDB, which show the same pattern of skew). With this in mind, though, both median and mean values were analyzed in the statistical analyses; actual median values and mean values are provided in Tables 1 and 2 in the Appendix.

---

[1] /pV/ is severely underrepresented, compared to other voiceless stops, because Japanese lost /p/ in its history, and singleton /p/ appears only in recent loanwords (Frellesvig, 2010; Ito and Mester, 2008).

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

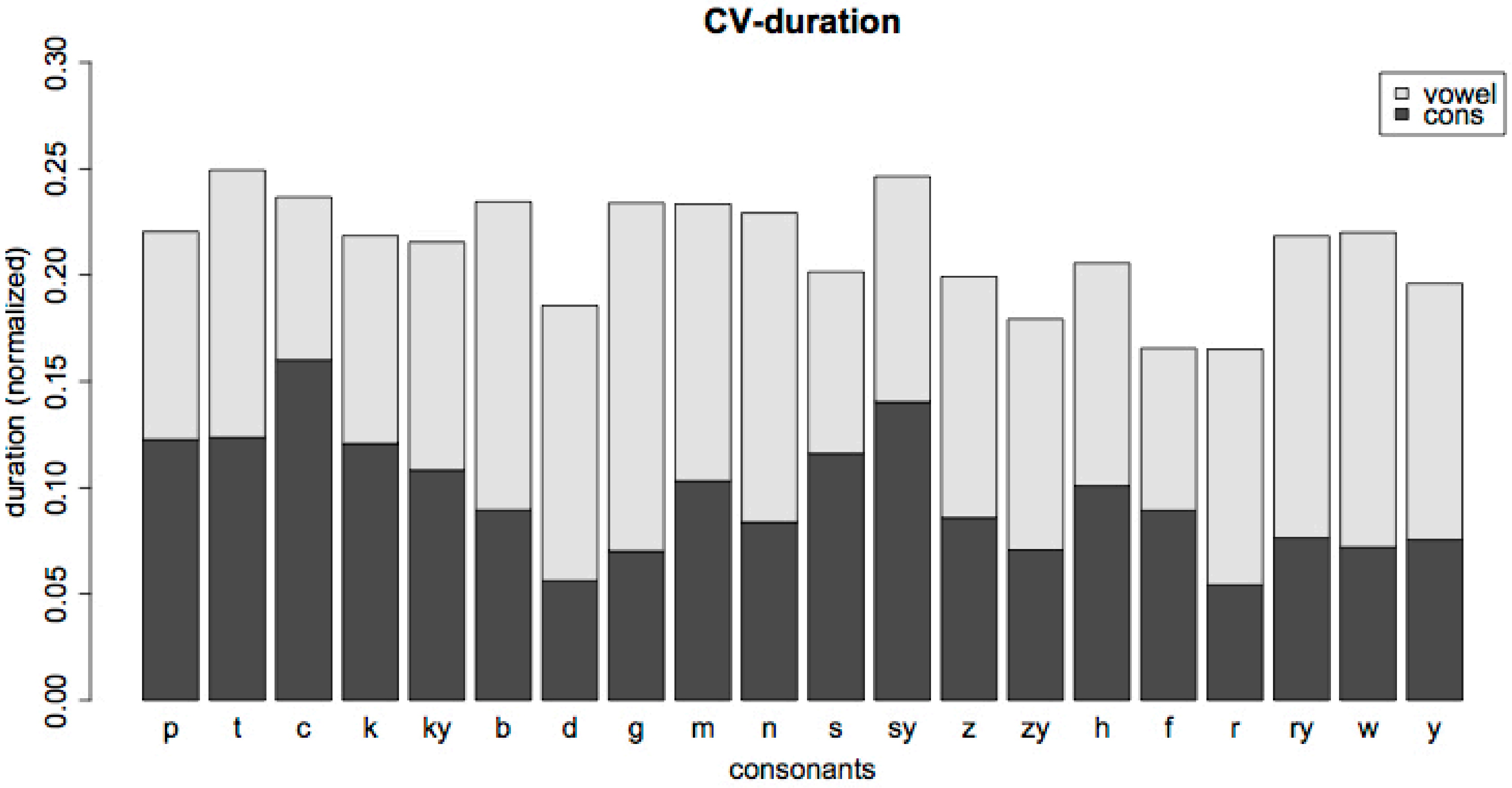
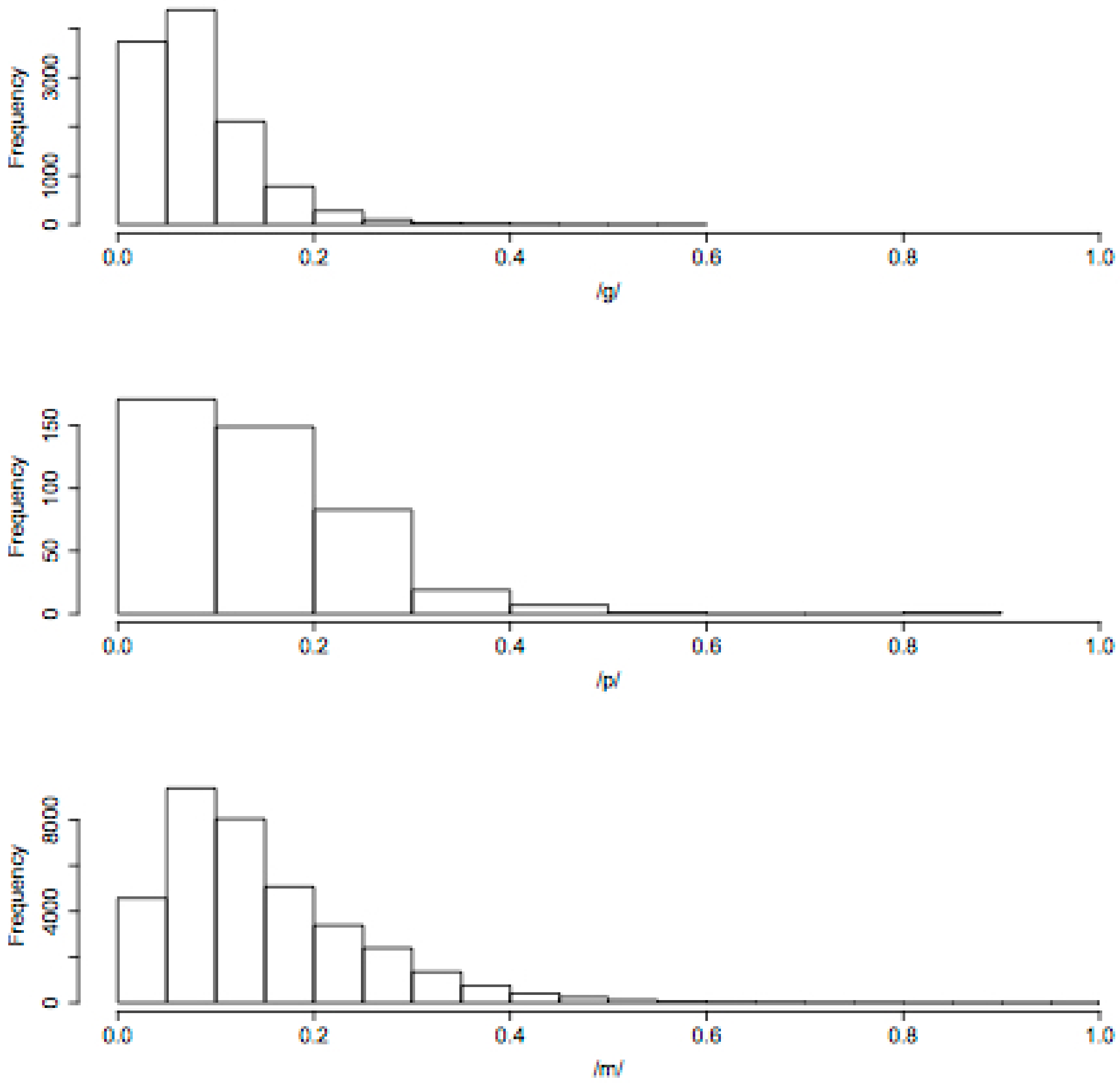
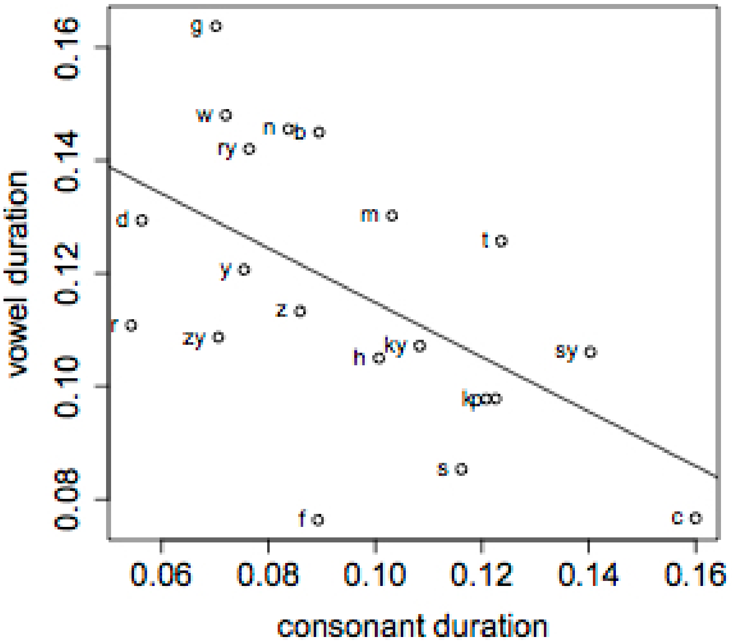


Figure 2: The distribution of consonant duration for /gV/, /pV/ and /mV/.



First, focusing on the behavior of consonants, voiced obstruents are generally shorter than their corresponding voiceless obstruents, as has been found in previous studies on Japanese (Homma, 1981; Kawahara, 2006; Shaw and Kawahara, 2017); the same tendency is known to hold cross-linguistically (e.g. Diehl and Kluender 1989; Kingston and Diehl 1994; Lisker 1957; Ohala 1983). In the current data, this tendency holds both among stops and fricatives. Second, for both voiced stops and nasal stops, labial consonants are longer than coronal and dorsal consonants (cf. Homma 1981; Shaw and Kawahara 2017 for similar observations). Third, we observe that voiceless fricatives and affricates—in particular "c" and "sy" —are longer than other consonants, again a tendency that holds cross-linguistically, including Japanese (Kawahara, 2015; Lehiste, 1970; Sagisaka and Tohkura, 1984). Finally, /r/, which is a flap in Japanese (see Arai 2013 for detail of its various realization patterns), is short, as expected.

Figure 3: The scatterplot showing the negative correlation between consonant duration and vowel duration (based on all vowels). The linear regression line is also shown.



Now moving on to the correlation between vowel duration and consonant duration, we observe that there is a statistically significant negative correlation between them ($r = -0.56, t(18) = 2.86, p < .05$), in such a way that vowels are shorter after longer consonants, as shown visually by the scatterplot in Figure 3—this negative correlation holds in terms of means as well to a statistically significant degree ($r = -0.60, t(18) = -3.20, p < .01$). For example, in Figure 1, we can observe that "c" is the longest consonant of all, and the following vowel is the shortest. The second longest consonant "sy" has a following short vowel as well. /g/ is one of the shortest consonants, and the following vowel is the longest. Furthermore, a comparison between /m/ and /n/ illustrates the compensation effect very 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. The minimal pair of /k/ and /ky/ also shows a similar pattern: /k/ is longer than /ky/ but the following vowel is shorter after /k/ than after /ky/, the result of which is comparable CV-durations. Comparing /b/ and /g/ points to the same observation.

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, /g/ is longer than /d/, but the vowel is also longer after /g/ than after /d/—the direction that is the opposite of what is expected from the compensation effect. Although /r/ is a short consonant, the following vowel does not get as long as it could 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 becomes more comparable to the moras with other onset consonants in their duration.

In order to 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). First, the standard deviation across the 20 consonantal conditions, calculated in terms of medians, served as the measure of the degree to which the entire CV mora duration is kept constant. The actual standard deviation is 0.025 across the 20 different conditions. In the bootstrap method, first one consonant interval and one vocalic interval were randomly sampled and their duration was combined. This process was reiterated 20 times without replacement to create 20 random 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 obtained confidence intervals, based on the median values, are 0.025 - 0.047 (95%) and 0.021 - 0.051 (99%). Since the observed standard deviation coincides with the lower end of the 95% confidence interval, this result indicates that the probability of the compensation effect occurring by chance is about 5%. The same analysis was run using the mean values for the 20 CV-moras, whose observed standard deviation is 0.028. The 95% confidence interval is 0.33–0.53 and the 99% confidence interval is 0.029-0.056. Therefore, from this analysis based on means, the probability of getting the observed standard deviations based on the mean values is less than 1%. Whether we rely on means or medians, it seems safe to conclude that the compensation effect observed in the current result is unlikely to have arisen by chance.

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: generally, the shorter the consonant, the longer the vowel tends to be. The bootstrap resampling analyses have 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, reviewed in the introduction section, with a large number of natural speech tokens. This paper moreover offers the first analysis that includes all types of consonants and all types of vowels in Japanese as targets.

Although we have observed a statistically significant compensation effect, we also found that durational 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, 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.

There are actually good reasons to expect that the compensation is not absolute, because there are many other linguistic factors that affect segments' duration profiles as well. The fact that we have found a significant compensation effect, in spite of there being other linguistic factors affecting segmental durations, actually provides stronger evidence for the active role of the compensation principle than otherwise. Let us consider a few—perhaps non-exhaustive—factors that may have blurred the compensation principle in the current analysis. For example, there is a collocation restriction in such a way that only /a/ can follow /w/ Vance (1987, 2008), but /a/ is the longest of all five vowels in Japanese (Campbell, 1999; Han, 1962; Kawahara et al., 2017; Kawahara, 2018; Shaw and Kawahara, 2017; Sagisaka and Tohkura, 1984). Coronal stops are also affricated before high vowels in native words (Vance, 1987, 2008), so that most of the vowels following /t/ and /d/ are non-high, which are generally longer than high vowels (although loanwords do allow coronal stops followed by high vowels: Pintér 2015). This distributional skew may explain why vowels are longer after /t/ than after /p/, despite the fact that /t/ and /p/ show comparable consonantal duration profiles; it may also explain why the following vowels are longer after /g/ than after /d/. In general, since vowels do not distribute evenly after different consonants (see, in particular, Shaw and Kawahara 2017), differences in intrinsic vowel duration would obscure the durational compensation principle.[2]

It is likely that the non-even distribution of vowels is not the only factor, because there are many factors that potentially affect segments' duration profiles, as we have known since the classic work by Klatt (1976). For example, voiced stops are sometimes spirantized intervocalically (Vance, 1987), and therefore, their duration estimates may not be always as reliable. Other factors like phrase-initial strengthening (e.g. Keating et al. 2003) and phrase-final lengthening (e.g. Wightman et al. 1992) can complicate the picture further. The effect of pitch accent on duration in Japanese is reported to be very small, but not non-existent (Hoequist, 1983*a*,*b*). Those elements that are informationally new or those elements that receive contrastive focus would be realized as

---

[2] A question still remains why intrinsic durational differences among different vowels are not overridden by the CV-mora compensation effect. More generally, modeling how different phonetic principles, which sometimes conflict with each other, interact to yield actual durational patterns is an important topic for future research (see e.g. Flemming 2001; Flemming and Cho 2017; Zsiga 2000 for concrete models.)

longer than more semantically neutral elements. Although the current analysis normalized speech rate within each speaker, there is no guarantee that speakers did not change their speech rate during the recording. In short, there are many other factors that could have blurred the compensation principle.

It is also likely the case that there are other linguistic principles at work in regulating the duration of Japanese vowels. For example, Shaw and Kawahara (2017) demonstrate that the average predictability of the vowels 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. Their conclusion is that the uncertainty associated with which vowel to produce after a particular consonant can potentially lengthen vowels' duration. Shaw and Kawahara (2017) also show that transitional probabilities, quantified in terms of Surprisal ($-\log_2 p(v|C)$), can impact the vowel duration. Shaw and Kawahara (2017) further demonstrate that /o/ is longer after palatal consonants, because speakers may need extra time to achieve the low F2 target. Finally, we need to take into consideration the fact that vowel length is contrastive in Japanese (Hirata, 2004; Hirata and Tsukada, 2009), and therefore, lengthening a vowel too much would jeopardize this length contrast. This consideration, for example, may explain why vowels do not lengthen as much after /r/.

The point of the discussion here is not to undermine the results of the current study—the real intent is that we should not expect the durational compensation to be perfect in natural speech corpora, because there are so many other linguistic factors that affect vowel and consonant duration. It is worth emphasizing, therefore, that it is all the more impressive that we observed a statistically robust compensation effect, *despite* there being other factors that could potentially have obscured it. All in all, exploring the interaction of the durational compensation effect and other principles, like predictability effects and collocation restrictions, offers an interesting opportunity for future research.

Appendix: Median and mean values

[XXX Insert Tables 1 and 2 here XXX]

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.

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.

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.

Flemming, E. (2001), “Scalar and categorical phenomena in a unified model of phonetics and phonology,” *Phonology*, 18(1), 7–44.

Flemming, E., and Cho, H. (2017), “The phonetic specification of contour tones: Evidence from Mandarin rising tone,” *Phonology*, 34(1), 1–40.

Frellesvig, B. (2010), *A History of the Japanese Language*, Cambridge: Cambridge University Press.

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.

Hoequist, J. C. (1983*a*), “Durational correlates of linguistic rhythm categories,” *Phonetica*, 40, 32–62.

Hoequist, J. C. (1983*b*), “Syllable duration in stress-, syllable- and mora-timed languages,” *Phonetica*, 40, 203–237.

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. (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. (2018), “Vowel-coda interaction in spontaneous Japanese utterances,” *Acoustical Science and Technology*, .

Kawahara, S., Erickson, D., and Suemitsu, A. (2017), “The phonetics of jaw displacement in Japanese vowels,” *Acoustical Science and Technology*, .

Kawahara, S., and Shinya, T. (2008), “The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase and Utterance,” *Phonetica*, 65(1-2), 62–105.

Keating, P. A., Cho, T., Fougeron, C., and Hsu, C.-S. (2003), “Domain-initial strengthening in four languages,” in *Papers in Laboratory Phonology VI: Phonetic interpretation*, Cambridge: Cambridge University Press, pp. 145–163.

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

Klatt, D. (1976), “Linguistic uses in segmental duration in English: Acoustic and perceptual evidence,” *Journal of the Acoustical Society of America*, 44, 401–407.

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.

Otake, T. (2015), “Mora and mora timeing,” in *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. H. Kubozono, Berlin: Mouton, pp. 493–524.

Pintér, G. (2015), “The emergence of new consonant contrasts,” in *The Handbook of Japanese Language and Linguistics: Phonetics and Phonology*, ed. H. Kubozono, Berlin: Mouton, pp. 121–165.

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.

Shaw, J., and Kawahara, S. (2017), “Effects of Surprisal and Entropy on vowel duration in Japanese,”, Ms. Keio University [Revision submitted to Language and Speech].

Truckenbrodt, H. (2004), “Final lowering in non-final position,” *Journal of Phonetics*, 32, 313–348.

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.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. (1992), “Segmental durations in the vicinity of prosodic phrase boundaries,” *Journal of the Acoustical Society of America*, 91, 1707–1717.

Xu, Y. (2010), “In defense of lab speech,” *Journal of Phonetics*, 38(3), 329–336.

Zsiga, E. (2000), “Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian,” *Journal of Phonetics*, 28, 69–102.

Table 1: Actual median values

	p	t	c	k	ky	b	d	g	m	n
vowel	0.123	0.124	0.160	0.121	0.108	0.089	0.056	0.070	0.103	0.084
cons	0.098	0.126	0.077	0.098	0.107	0.145	0.129	0.164	0.130	0.146
total	0.220	0.249	0.237	0.219	0.215	0.234	0.186	0.234	0.233	0.229

	s	sy	z	zy	h	f	r	ry	w	y
vowel	0.116	0.140	0.086	0.071	0.101	0.089	0.054	0.076	0.072	0.075
cons	0.085	0.106	0.113	0.109	0.105	0.076	0.111	0.142	0.148	0.121
total	0.201	0.246	0.199	0.179	0.206	0.166	0.165	0.218	0.220	0.196

Table 2: Actual mean values

	p	t	c	k	ky	b	d	g	m	n
vowel	0.146	0.142	0.180	0.141	0.143	0.105	0.069	0.082	0.114	0.094
cons	0.126	0.160	0.098	0.121	0.125	0.168	0.174	0.200	0.156	0.176
total	0.271	0.303	0.278	0.263	0.268	0.273	0.243	0.282	0.271	0.270

	s	sy	z	zy	h	f	r	ry	w	y
vowel	0.140	0.159	0.098	0.084	0.121	0.110	0.061	0.082	0.080	0.089
cons	0.107	0.128	0.136	0.133	0.125	0.093	0.140	0.154	0.196	0.143
total	0.246	0.287	0.234	0.217	0.246	0.203	0.201	0.236	0.277	0.232

Figure captions:

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

Figure 2: The distribution of consonant duration for /ɡV/, /pV/ and /mV/.

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