WORD LEVEL TIMING IN SPONTANEOUS JAPANESE SPEECH

Takayuki Arai* and Natasha Warner[†]

**Sophia University, Tokyo, Japan,*

[†]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

ABSTRACT

This study provides evidence against the hypothesis that Japanese has word level mora-timing. Unlike previous studies which used careful speech, this paper evaluates timing in a corpus of spontaneous Japanese speech from 11 speakers.

Correlations between word duration and number of moras in the word are shown to be much lower than in careful speech studies. Furthermore, if there were durational normalization at the level of the word, then there should be some unit across which this normalization takes place. However, we show that there is no consistency across speakers as to whether the lexical word or the prosodic word serves as such a unit. Analyses of arbitrarily truncated words further confirm that a linear accumulative model of variance can explain the data, but a model with mora compensation cannot.

1. INTRODUCTION

Japanese is often said to be a “mora-timed” language, and many researchers have offered evidence both for and against this claim. The early mora-timing hypothesis claimed that in Japanese, all moras tend toward equal length. This isochronous mora proposal acknowledges that all moras cannot possibly have identical duration, but claims that speakers adjust segmental durations in order to make all moras as close as possible to equal.

There are two types of segmentally induced variation which such compensation would have to adjust for, shown in (1). Because /i/ is inherently rather short, and /a/ rather long, in the absence of compensation, one would expect the mora /ki/ to be shorter than the mora /ka/. Similarly, one would expect /ma/ (with the rather long consonant /m/) to be longer than /ra/ (with the very short flap). In addition to these inherent differences in segment durations, Japanese has several types of moras which do not consist of a CV string, as shown in (1b). Long vowels, geminates, and mora nasals each contribute an entire mora to the word (so that /tookyoo/ has four moras and /katta/ and /kaNda/ each have three), but these types of moras might not have the same duration as a typical CV mora. Furthermore, a mora containing a devoiced or even deleted vowel is still an entire mora, but the duration of the vowel is very brief. Both these types of variation lead to moras not having uniform duration.

(1a) Inherent differences in segmental duration

[ki]	vs.	[ka]
[ma]	vs.	[ɾa]

(1b) Special moras

long vowels	/tookyoo/	‘Tokyo’
geminates	/katta/	‘bought’
mora nasals	/kaNda/	‘place name’
devoiced vowels	/kita/ [ki̥ta]	‘North’

Investigations of segments and mora duration in Japanese [1, 2, 3, 4] have found evidence of compensation either between segments of the same mora or between nearby segments in different moras. For example, there is a significant negative correlation between the duration of a consonant and the duration of the following vowel [4]. Furthermore, word-initial /s/ or /i/ is longer when a medial consonant is geminate than when it is single [2]. Several researchers have shown a variety of such effects and concluded that Japanese speakers normalize the duration of moras.

However, Beckman [5] failed to replicate several such effects, and explained others as universals (for example vowels being longer before voiced than voiceless stops). Otake [6] replicated some such effects for Spanish and English, not mora-timed languages. The duration of a mora in Japanese is also not well predicted by the duration of the preceding mora [7].

Port et al. [8, 9] redefined the mora-timing hypothesis by proposing that Japanese speakers attempt to normalize segment lengths to make the duration of an entire word predictable from the number of moras in the word, not to make all moras isochronous (see also [3]). Port et al. [9] showed that the duration of a Japanese word is linearly related to the number of moras it contains, with correlations as high as r=.99.

All of these studies (except [4 and 7]) used very careful speech, with target words usually produced in focus position. Some used nonsense words, including some phonotactically unlikely items. In the current study, we analyze the relationship between word duration and number of moras for a corpus of spontaneous speech in order to determine whether there is any evidence of mora-timing in natural speech.

2. METHODS

The material consists of 50 seconds of spontaneous speech from each of 11 native speakers of Japanese, a subset of the corpus in [7, 10]. The speech was collected over the telephone. After answering several questions which were presented in Japanese by a recorded voice, speakers were asked to talk on any subject they wished for approximately a minute. The first author labeled this speech at the mora level and transcribed it.

Only those words in the corpus for which word boundaries are clear were analyzed. It is somewhat difficult to determine what constitutes a “word” in Japanese because of the large variety of final particles, suffixes, and particle-like constructions, and Port et al. [9] did not define explicitly what unit they predicted the temporal compensation to involve. Therefore, in this study, units such as a noun plus its following particle or a verb plus its inflectional suffixes were included as single words. Adverbs or conjunctions in isolation were also included. However, tightly bound constructions where prosodic boundaries might not coincide with syntactic boundaries were excluded, as were interjections, nouns followed by any form of the copula, and nominalized verbs. Nouns followed by three or more particles were also excluded because of the uncertain prosody of long strings of particles. In general, only cases in which word boundaries are clear were included in order to provide a fair test of the word level mora-timing hypothesis. All potential cases of noun compounds were treated as single words. In evaluating Port's [9] hypothesis, using too small a unit would obscure the effect, but using too large a unit would not introduce any error.

The duration of each remaining word was calculated from the mora durations which had been measured for previous studies [7]. Words with more than 7 moras were excluded from all analyses, in order to increase comparability with [9]. This also prevents an inordinately strong effect of the few words with many moras on the correlations.

3. RESULTS

3.1. Correlation of word duration with number of moras

The correlation between word duration (without particles) and number of moras in the word was calculated for each speaker. Final particles are very common in Japanese (2). The data in [9] were produced with a following quotative particle /to/, but this particle was excluded from the analyses. In order to replicate this procedure, we also initially excluded particles.

(2)	/yuuhaN-ga/	'dinner-subject'
	/puuru-o/	'pool-object'
	/kagosimasinai-de-wa/	'Kagoshima.city.center-at-topic'

The correlations between word duration and number of moras (Table 1) vary among speakers from r=.701 to r=.931. Figure 1 shows this relationship for Speaker D. Obviously, this is far below the correlation of r=.999 (or r=.998 in fast speech) reported in [9]. (The correlations in the current study are also not weak. Words with more moras are of course longer, but that fact alone does not support the mora-timing hypothesis.) It is perhaps not surprising that spontaneous speech is far more variable in timing than careful speech. However, it is interesting to note how widely the speakers vary in the predictability of word duration from number of moras.

3.2. Word duration with particles included vs. excluded

Although Port et al. [9] excluded particles from the word unit, particles are almost certainly part of the prosodic word in Japanese. They must form part of the same accentual phrase as the preceding word, and can never form an accentual phrase by themselves [11]. Durational compensation is a prosodic process, so one would expect it to take place at the level of the prosodic, not lexical, word. We therefore calculated the duration/number of moras correlation with the particles included and excluded, for each speaker. For example, in the particle excluded correlation, for /yuuhaN-ga/ 'dinner-subject,' only the duration of /yuuhaN/ (with 4 moras) was used. In the particle included correlation, the duration of /yuuhaN-ga/ (with 5 moras) was used.

If the prosodic word is the unit of durational compensation, speakers should have a stronger correlation between word duration and number of moras with the particle than without it. Correlations for all speakers are shown in Table 1.

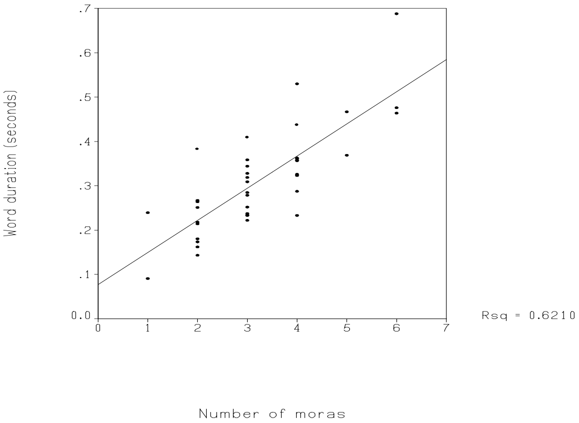

Figure 1. Correlation between word duration and number of moras for speaker D.

Speaker	Without particle (r)	With particle (r)
A	0.831	0.809
B	0.795	0.802
C	0.701	0.739
D	0.788	0.782
E	0.889	0.814
F	0.818	0.819
G	0.931	0.942
H	0.800	0.808
I	0.798	0.816
J	0.874	0.840
K	0.845	0.839

Table 1. Correlations of word duration with number of moras, with and without particle.

Only six of the eleven speakers show a stronger relationship between word duration and number of moras when particles are included in the word. Furthermore, the differences in strength of the relationship are small for all speakers. This shows not only that half of the speakers are not using a prosodic unit for this compensatory process, but that there is no consistent unit of temporal compensation across speakers. If the effect of word level mora-timing were as strong as Port et al. [9] claim, there should be a clear difference in the strength of the correlation depending on whether the correct unit is evaluated or not.

3.3. Final truncation

If the proposed effect of word level mora-timing exists, whatever unit it operates within, when one removes some arbitrary part of that unit, the relationship between the duration of the remaining part of that unit and the number of moras in it should be weaker than the relationship between duration and number of moras for the entire unit. This is because one has removed some of the segments which are compensating for variation in the remaining segments, or removed some of the segments which the remaining segments are compensating for. One is thus left with either too little or too much compensation.

We calculated the durations of the words used in Section 3.1 without their final two moras. (Particles were first excluded for this purpose.) For example, for the word /kagosimasi/

‘Kagoshima City,’ the duration of the string /kagosi/ was calculated. Two moras is an arbitrary portion of the word, and the final two moras do not form any particular unit. If one removed only one mora, this might not be a large enough portion of the word to disturb the compensation effect clearly. If one removed three moras or more, many words would be too short to examine. We of course excluded words with only one or two moras, and words in which the boundary between the penultimate and antepenultimate mora falls during a long vowel or geminate, such as /ikkai/ ‘one time’ and /nyuuyooku/ ‘New York.’ It is difficult to locate accurate boundaries during these segments.

We calculated the correlation of the duration of the truncated words with the number of moras in the truncated words, and compared this to the correlation between whole word duration and number of moras in the whole words, for only those words which could be used for the truncation analysis. (Exactly the same words are used in the two correlation analyses.) These correlations are shown in Table 2, and Figure 2 shows the relationship of the two correlations for one speaker.

Speaker	Truncated (r)	Entire word (r)
A	0.947	0.790
B	0.915	0.749
C	0.934	0.702
D	0.846	0.745
E	0.952	0.870
F	0.930	0.745
G	0.936	0.904
H	0.910	0.577
I	0.922	0.707
J	0.862	0.781
K	0.879	0.766

Table 2. Correlation between duration and number of moras for truncated and whole words.

For all eleven speakers, the correlation between duration and number of moras is weaker for whole words than for truncated words. This is the opposite of what Port’s [9] mora-timing hypothesis predicts. Furthermore, the difference in degree of predictability is quite large for most speakers. It appears that word duration is much more predictable from number of moras without the final two moras than with them. This is clear evidence against the word level mora-timing hypothesis.

3.4. Initial truncation

The results of the final truncation analysis could mean that there is no mora-timing in Japanese, or they could indicate that the final two moras of a word are particularly variable (perhaps because of final lengthening), but that moras otherwise tend toward the same duration. This would be a modified version of the isochronous mora hypothesis, in which the final two moras of the word are admitted to be highly variable for some reason, but compensation makes all other moras tend toward isochrony. (If all moras tended to be approximately the same length, word duration would also be predictable from number of moras.)

To test this possibility, we performed an initial truncation analysis. The procedures were identical to the final truncation analysis, but the initial two moras were removed instead of the final two. Words with the boundary between their second and third moras falling during a long vowel or geminate were excluded, as were words with fewer than three moras. Again, the same words were excluded from a new whole word correlation.

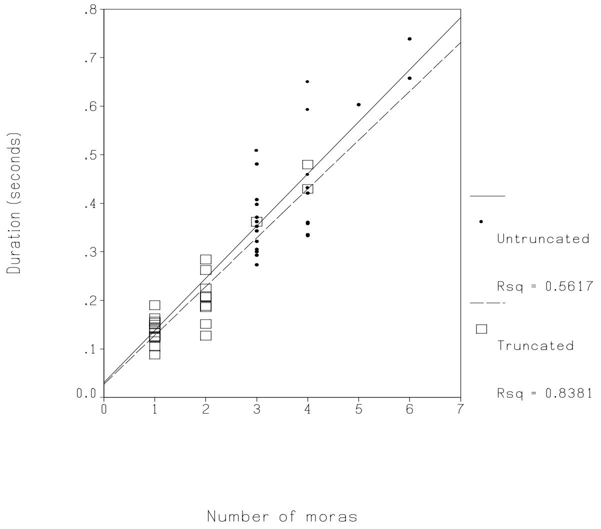

Figure 2. Final truncation for speaker B.

Port’s [9] hypothesis would predict, as with final truncation, that the relationship between initial truncated word duration and number of moras should be weaker than the relationship between whole word duration and number of moras, since part of the unit across which compensation should take place has been removed. A modified version of the isochronous mora hypothesis would also predict that under initial truncation, the correlation between number of moras and truncated duration should be weaker. This is because, if the final two moras of the word are the most variable, removing two less variable moras and leaving those most variable moras should produce more variability overall.
The results of this analysis appear in Table 3, and a graph of the relationships for one speaker is shown in Figure 3. For ten of the eleven speakers (all except Speaker G), the relationship between duration and number of moras is stronger when the initial two moras are removed than for whole words. This is very similar to the results for final truncation. This shows again that word level mora-timing does not apply in this data. When an arbitrary

Speaker	Initial truncated (r)	Entire word (r)
A	0.810	0.796
B	0.844	0.758
C	0.827	0.702
D	0.885	0.713
E	0.937	0.906
F	0.815	0.806
G	0.884	0.902
H	0.568	0.534
I	0.775	0.717
J	0.874	0.776
K	0.846	0.453

Table 3. Correlation between duration and number of moras for initial truncated and whole words.

portion of the word is removed, whether the beginning or the end, duration of the remainder is more predictable from number of moras, not less predictable. This is directly counter to the word level mora-timing hypothesis, since truncation removes part of the word which should be compensating for variation in the remainder of the word.

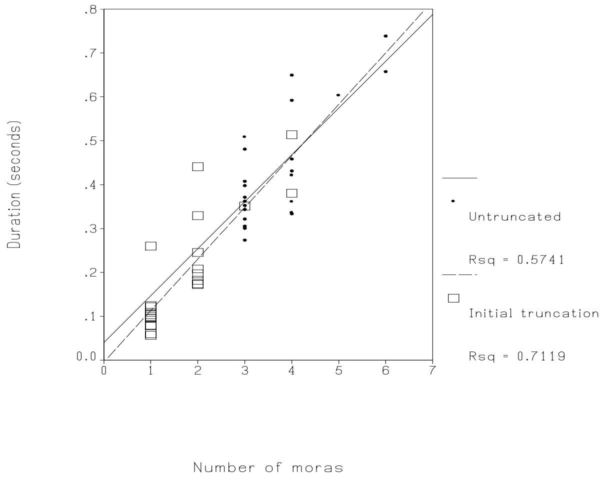

Figure 3. Initial truncation for speaker B.

The increase in duration predictability with initial truncation cannot be explained by the isochronous mora version of the mora-timing hypothesis either. It is not simply that the final two moras are the most variable in the word, with all other moras having nearly the same durations, since removing the first two moras also makes duration more predictable. One cannot suggest that the initial *and* final two moras of words are the most variable, with the moras in between them being highly isochronous: there are few words in the analysis with many moras other than the initial and final two. While our analyses are not designed to test the isochronous mora hypothesis directly, that version of mora-timing does not offer an explanation for our results.

4. GENERAL DISCUSSION

This study shows that word duration in Japanese is far more variable in spontaneous speech than previous studies with read speech would lead us to predict. Furthermore, it offers strong evidence against the concept of word level mora-timing in Japanese. The results discussed above (particularly in 3.3 and 3.4) cannot be explained by mora-timing, but can be explained by a linear accumulative model of variance [12]. In such a model, each unit contributes variability. The more units are concatenated, the more variability the whole has. This explains the results of the two truncation analyses quite well: when one removes two moras of the word, one is removing two of the units which contribute variability to the whole, and thus reducing the overall variability of the remaining portion of the word. This interpretation of the results is consistent with the idea that speakers do not attempt to normalize the duration of moras, either to make them more similar to each other or to control the duration of the whole word.

This result does not mean that the mora does not exist in Japanese. A wide variety of research has demonstrated the crucial role of the mora in Japanese phonology [13], and psycholinguistic work has shown its importance in the processing of Japanese speech [14, 15]. The mora is also the tone bearing unit for the pitch accent system [16]. Furthermore, the extremely long durational differences in the long/short vowel or geminate/single consonant distinctions in Japanese may contribute to the perception of mora-timing. However, none of these things mean that speakers of Japanese intentionally manipulate the duration of moras in order to normalize the duration of words. The mora is important in Japanese phonetics, phonology, and psycholinguistic processing, but our data does not support it as a unit of temporal normalization.

ACKNOWLEDGMENTS

We thank John Ohala, Anne Cutler, and Antje Meyer for discussions of this material, and Pieter Meima for help in analyzing the data. Any errors are, of course, our own.

REFERENCES

[1] Han, M.S. 1962. The feature of duration in Japanese. *Onsei no Kenkyuu*, 10, 65-80.

[2] Han, M.S. 1994. Acoustic manifestations of mora timing in Japanese. *JASA*, 96, 73-82.

[3] Homma, Y. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, 9, 273-281.

[4] Sagisaka, Y. and Y. Tohkura. 1984. Phoneme duration control for speech synthesis by rule. *IEICE Trans*, J67-A:7, 629-636.

[5] Beckman, M. 1982. Segment duration and the ’mora’ in Japanese. *Phonetica*, 39, 113-135.

[6] Otake, T. 1989. Counter evidence for mora timing. In *Proceedings of the 16th Lacus Forum*. 313-322.

[7] Arai, T. and S. Greenberg. 1997. The temporal properties of spoken Japanese are similar to those of English. In *Proceedings of the 5th European Conference on Speech Communication and Technology*, 2, 1011-1014, Rhodes, September 1997.

[8] Port, R.F., S. Al-Ani, and S. Maeda. 1980. Temporal compensation and universal phonetics. *Phonetica*, 37, 235-252.

[9] Port, R.F., J. Dalby, and M. O’Dell. 1986. Evidence for mora timing in Japanese. *JASA*, 81, 1574-1585.

[10] Muthusamy, Y.K., R.A. Cole, and B.T. Oshika. 1992. The OGI multi-language telephone speech corpus. *Proceedings of the International Conference on Spoken Language Processing.* 895-898.

[11] Poser, W.J. 1984. The phonetics and phonology of tone and intonation in Japanese. MIT dissertation.

[12] Ohala, J.J. 1975. The temporal regulation of speech. In Fant, G., and M.A.A. Tatham (eds.), *Auditory Analysis and Perception of Speech*. 431-453. London: Academic Press.

[13] Kubozono, H. 1996. Speech segmentation and phonological structure. In Otake, T., and A. Cutler (eds.), *Phonological Structure and Language Processing*. 77-94. Berlin: Mouton de Gruyter.

[14] Otake, T., G. Hatano, A. Cutler, and J. Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, 258-278.

[15] Cutler, A., and T. Otake. 1994. Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33, 824-844.

[16] Pierrehumbert, J.B., and M.E. Beckman. 1988. *Japanese tone structure.* Linguistic Inquiry, Monograph 15. Cambridge: MIT Press.