

# The role of the mora in the timing of spontaneous Japanese speech

Natasha Warner<sup>a)</sup>

Max Planck Institute for Psycholinguistics, Wundtlaan 1, PB 310, NL-6500 AH Nijmegen,  
The Netherlands and Department of Linguistics, University of Arizona, PO Box 210028, Tucson,  
Arizona 85721-0028

Takayuki Arai

Department of Electrical and Electronics Engineering, Sophia University, 7-1 Kioi-cho, Chiyoda-ku,  
Tokyo 102-8554, Japan

(Received 15 August 1999; accepted for publication 28 November 2000)

This study investigates whether the mora is used in controlling timing in Japanese speech, or is instead a structural unit in the language not involved in timing. Unlike most previous studies of mora-timing in Japanese, this article investigates timing in spontaneous speech. Predictability of word duration from number of moras is found to be much weaker than in careful speech. Furthermore, the number of moras predicts word duration only slightly better than number of segments. Syllable structure also has a significant effect on word duration. Finally, comparison of the predictability of whole words and arbitrarily truncated words shows better predictability for truncated words, which would not be possible if the truncated portion were compensating for remaining moras. The results support an accumulative model of variance with a final lengthening effect, and do not indicate the presence of any compensation related to mora-timing. It is suggested that the rhythm of Japanese derives from several factors about the structure of the language, not from durational compensation. © 2001 Acoustical Society of America.

[DOI: 10.1121/1.1344156]

PACS numbers: 43.70.Fq, 43.70.Bk [AL]

## I. INTRODUCTION

Japanese is often called mora-timed, in opposition to stress-timed languages like English and syllable-timed languages like French. This division is usually taken to mean that moras are approximately regularly timed in Japanese, such that each mora is of similar duration. (In the same way, stresses are said to be nearly regular in English, and syllables in French.) One commonly argued position has been that if variability in durations of different segments interferes with this regularity, speakers compensate to make the durations of the relevant unit more regular than they would otherwise be. However, another possibility is that factors in the phonology of the language cause the relevant unit to be more regularly timed than other units, but that there is no control mechanism to create regular timing of that unit. In this article, we present new analyses of Japanese speech timing which support the latter hypothesis. We also extend the investigation to spontaneous speech, which has rarely been used in past studies.

There are at least two types of variation which would have to be compensated for to make mora durations equal in Japanese. First, there are inherent differences in segmental durations: high vowels are shorter than low vowels, [r] is very short while [s] is relatively long, etc. Second, several types of moras in Japanese do not consist of a consonant-vowel string. Long vowels have two moras, so that /tookyo/ “Tokyo” has the same number of moras as /tokonoma/ “alcove.” Geminate obstruents and the mora nasal (/N/) con-

tribute a mora: /katta/ “bought,” /kaNda/ (place name), and /karita/ “borrowed” each have three moras. High vowels in a voiceless environment can be devoiced or deleted in Japanese, but their mora remains, so that /kita/ [k̟ita] “North” has two moras. A word with any of these special mora types is unlikely to have the same duration as a word with only CV moras.

### A. Mora-timing as durational compensation

Early experimental work on Japanese duration claimed that speakers adjust the duration of the segments within a mora to make mora durations more similar to each other than they would be based on inherent variability. We will call this the “tendency toward isochrony” version of the mora-timing hypothesis. Under this theory, if one segment of a CV mora is shorter than average (e.g., [r]), or longer than average (e.g., [s]), the other segment of that mora will be lengthened, or shortened, in order to maintain near-constant mora duration. This compensation might not be complete. This theory also predicts that non-CV moras such as the mora nasal or the beginning of a geminate obstruent will have duration similar to a CV mora. Some form of this position is supported by Han (1962), Homma (1981), Sagisaka and Tohkura (1984), and Sagisaka (1992, 1999), while Beckman (1982) and Hoequist (1983a, b) find contradictory evidence. For a full review of the literature on Japanese mora-timing, please see Warner and Arai (2001).

Based on the finding (Port *et al.*, 1980; Sagisaka and Tohkura, 1984) of negative correlations across mora boundaries (in VC strings) as well as within moras (CV strings),

<sup>a)</sup>Electronic mail: Natasha.Warner@mpi.nl

Port *et al.* (1987) propose an important revision to the concept of mora-timing. They suggest that compensation takes place not only within the mora, but also across several moras, and that compensation serves not to normalize the duration of individual moras, but rather to make the duration of larger units (perhaps the word) predictable from the number of moras in them. This leads to different predictions: under the tendency toward isochrony theory, all moras are expected to have nearly equal duration. However, Port *et al.* (1987) propose that if a segment, for example the consonant of a word-medial CV mora, is shorter than average, nearby segments both in the same and other moras will be lengthened to compensate for it. Thus if one mora is shorter than average, nearby moras would be longer than average to compensate for it, making mora durations less similar to each other, not more similar. Port *et al.* (1987) refer to this relationship as “anti-compensation.”

This modification to the theory focuses attention on the duration of larger units, such as words. However, this theory does not claim that all moras within a word compensate for each other, or that the domain of compensation is exactly the word. Evidence adduced for this version of mora-timing emphasizes the relationship between duration of entire words (as a convenient unit larger than the mora) and the number of moras in them. Several studies demonstrate strong positive correlations between word duration and the number of moras in a word in Japanese (Port *et al.*, 1987; Bradlow *et al.*, 1995; Sato, 1995; Kondo, 1999), although Otake (1989) shows similar effects for Spanish (syllable-timed) and English (stress-timed). One difficulty is that it is not clear how strong these correlations must be in order to demonstrate mora-timing. Since all moras have segmental content, it is not surprising that this correlation is strong. A strong correlation between word duration and number of moras does not, of itself, demonstrate compensation. [See Warner and Arai (2001) for further details.]

## **B. The mora as a nontemporal unit and accumulation of variability**

Alternatively, the mora could play a structural role in Japanese as an abstract unit which does not control timing. Factors in the phonology of Japanese could result in the mora being more regularly timed than the syllable or the foot in Japanese speech, without any compensation to bring moras closer to regular timing than they inherently would be. Factors which would have this effect include the lack of an effect of pitch accent on duration, the lack of reduction of unaccented vowels, the prevalence of CV syllables, and the lack of nondurational cues to the geminate/singleton consonant and long/short vowel distinctions. In all of these points, Japanese differs strongly from stress-timed English. It is clear that the mora exists as a phonological and psycholinguistic unit in Japanese (see Warner and Arai, 2001), but its role need not involve temporal compensation, even if the relatively regular timing of the mora brought about by phonological factors gives Japanese a distinctive rhythm.

Dauer (1983, 1987) offers several phonological correlates of stress- and syllable-timing, and proposes that these factors combine to create a rhythm in which stresses or syl-

lables seem to be the prominent feature of timing in a language, without any temporal compensation. She does not discuss mora-timing, but Otake (1990) and Beckman (1992) extend her proposals to Japanese. They emphasize the role of statistical predominance of CV syllables and of the pitch accent system, respectively, in creating mora-rhythm. For example, Dauer points out that stress-timed languages allow fewer syllables within a stress-foot than syllable-timed languages do. Beckman (1992) points out that Japanese even allows entire phrases with no pitch accent at all, thus going further than the syllable-timed languages in this respect. She also points out the importance of the lack of an effect of pitch accent on duration. Work by Ramus *et al.* (1999) is also relevant in pointing out acoustic correlates of rhythm class other than regular timing.

The phonological length distinctions may also contribute to the impression of regularly timed moras in Japanese. Duration is effectively the only difference between Japanese long and short vowels (e.g., /kooki/ “school flag” versus /koki/ “exhalation”), unlike English (e.g., “beet” versus “bit”). Thus, this duration difference is very large in Japanese. The same applies to geminate and singleton obstruents in Japanese (e.g., /katta/ “bought” versus /kata/ “form”), where closure duration is nearly the only difference. The large durational difference between such phonologically short and long segments does not indicate any compensation, but it makes moras more regular than in a language with smaller durational differences, such as English.

If “mora-timing” reflects phonological facts about the language, and speakers do not employ temporal compensation to regularize the timing of moras, then the inherent variation in segment durations, and thus mora durations, would remain. Without compensation, each mora in a larger unit (such as a word) would contribute its variability to the variability of the whole. That is, the hypothesis that the mora is a structural unit which only indirectly affects timing (without compensation) predicts that the variance of individual smaller units, in this case moras, accumulates in larger units, instead of canceling out. One prediction of this hypothesis is that the duration of words containing many moras will be more variable than the duration of words containing fewer moras, since longer words contain more sources of variability (Ohala, 1975). This prediction is the basis of several analyses we present later to test for compensation versus only a structural role for the mora.

## **C. Goals of the current study**

Nearly all of the past work on Japanese mora-timing has used very careful speech. [Exceptions are work by Sagisaka and his colleagues (Sagisaka, 1992, 1999; Sagisaka and Tohkura, 1984), as well as by Campbell (1992) and Arai and Greenberg (1997), of which the last two do not support mora-timing.] While it is desirable to study timing under controlled conditions, it is also desirable to confirm findings from controlled speech under natural conditions if this can be done in a reliable way. Because so much work has been done on timing in controlled Japanese speech, extension to natural materials should now be possible.

Even if compensatory mora-timing exists in careful speech, speakers may not be able to maintain it in the face of the many other effects on duration present in connected speech, such as final lengthening, focus (on words other than the target item), etc. These factors might have a legitimate influence on any mora-timing, rather than simply adding noise. Since target words in the focus position of predictable frame sentences are not what speakers normally produce or listeners normally perceive, it is important to determine whether patterns which exist in that environment are also present in speech which is more typical of daily life. This is particularly important for mora-timing because Japanese listeners use the mora to help locate word boundaries (Otake *et al.*, 1993, 1996; Cutler and Otake, 1994), which must be done for connected speech. It is important to know whether it is the temporal role of the mora or something else that leads to its usefulness for segmentation. If mora-timing compensation does not exist in connected speech, then mora-timing could not be the source of the mora's usefulness for segmentation.

Of course, in studying connected speech, one must control for effects which might mask the effect under investigation, such as overall speech rate variation. We control for interspeaker rate variation by making only within-subjects comparisons. Our analyses also control for intraspeaker overall rate variation within the monologue, since nearly all of our analyses<sup>1</sup> compare different measurements from the same words. This also controls for any factors which affect the duration of a specific word, such as syntactic position in the sentence, length of the sentence, part of speech, focus, or slowing at a high level prosodic boundary. As for factors which affect units smaller than the word, we test for word-final lengthening, and other effects on sub-word-level units (such as the mora) are precisely the subject of investigation. We also exercised considerable caution in the selection of the words to test (as described later) to avoid including words which might disrupt compensation effects for known reasons. Finally, because any data on natural speech is expected to be more noisy than similar data on careful speech, null effects in natural speech would not argue strongly for any conclusion, but significant effects should not be discounted.

One goal of this study is to test past findings from careful speech in natural speech. A second goal of the study is to present new tests, particularly the analyses of truncated words, for which the compensation hypothesis makes clearer predictions than a strong correlation, since it is not clear how strong a correlation is necessary to demonstrate compensation. Some type of compensatory mora-timing has been widely accepted in the recent literature (Han, 1994; Bradlow *et al.*, 1995; Sato, 1995; Itô and Mester, 1995, p. 834; Tsujimura, 1996, p. 66; Kubozono, 1999; Kondo, 1999; Minagawa-Kawai, 1999), so these new tests are warranted. Furthermore, previous evidence for a structural, nontemporal role for the mora has been limited to the inconsistency of evidence for compensation (Beckman, 1992) or to the finding of supposed mora-timing effects in non-mora-timed languages (Otake, 1989). The truncation analyses presented in this article, however, make clear and contradictory predic-

tions for structural versus temporal, compensatory, function of the mora, and thus avoid the limitations of past evidence for both hypotheses.

Specific predictions will be discussed for each test below. In general, we predict that timing relationships will be more variable in spontaneous speech than in the careful speech of previous studies. We also predict that analyses of duration will show evidence of each mora contributing variance to the variance of the whole, rather than of compensation among moras, supporting the structural rather than temporal role of the mora.

## II. GENERAL METHODS

### A. The corpus

The materials for this study consist of 50 s of spontaneous speech, collected over the telephone, from each of 11 volunteer native speakers of Japanese [a subset of the larger corpus described by Muthusamy *et al.* (1992) and Arai and Greenberg (1997)]. Most speakers were living in the United States when the data was recorded. A recorded voice (in Japanese) asked speakers to talk on any subject they wished for approximately a minute. Speakers who read a text aloud instead of speaking spontaneously were excluded. For the current study, five native speakers of Japanese (three with phonetics training, including the second author) evaluated the corpus used by Arai and Greenberg (1997) for foreign accents. (Although the speakers stated that Japanese was their native language, some had been living in the United States for many years.) Only speakers who all five evaluators judged as having no foreign accent were included. Speakers of nonstandard dialects were not excluded, since mora-timing is not claimed to differ in most nonstandard dialects. One evaluator, an expert in phonetics and Japanese dialectology, found that only 3 of the remaining 11 speakers spoke nonstandard dialects (speakers A, E, and J).<sup>2</sup>

The corpus was labeled at the mora level by the second author, using standard criteria for the boundaries of segments, and based on spectrograms, waveforms, and listening. Filled and unfilled pauses and nonspeech noises were also labeled. In the case of word-initial voiceless stops, if the voiceless stop was immediately preceded by the final segment of the preceding word, with no pause, the intervening silence was included as the closure of the voiceless stop. However, if an initial voiceless stop followed a pause, the silence was included in the pause and the duration of the stop was counted from the burst. For vowels followed by a voiceless segment, the end of voicing was counted as the end of the vowel. Periods of creaky voicing, such as at the beginning of a word-initial vowel or end of a final one, were included in the vowel. For a vowel followed by a voiced obstruent, the boundary was placed at the sudden drop in amplitude and cessation of formants. For vowel-nasal or nasal-vowel transitions, the sudden change in formant frequencies and amplitudes was identified as the boundary. The onset of /r/ (the only liquid in Japanese, realized usually as [r], but sometimes as [l] or [d]) was identified as the beginning of low amplitude voicing with a gap in the formant structure. In vowel-glide sequences, the beginning of a re-

TABLE I. Examples of included types of items, with the environment in which they were produced. The exemplified word is shown in bold. Transcriptions are phonemic.

1a. atasi-wa <b>daigaku</b> saNneN pause de pause I-topic college 3rd. year <b>nihoN-no</b> minami-no pause Japan-gen. South-gen. <b>gakoo-made-wa</b> nizikaN kakari pause school-to-topic 2.hours take	“I’m a third year college (student)” “of the South of Japan” “It takes 2 hours to get to school”
1b. kazaNbai-ga <b>hurimasu</b> pause volcanic.ash-subj. fall-polite-non-past toku-ni zyanru-wa <b>toimaseN</b> especially genre-topic care-polite-non-past-negative tookyoo-kara <b>kimasita</b> Tokyo-from come-polite-past	“volcanic ash falls” “I don’t really care about the genre” “I came from Tokyo”
1c. pause maa <b>tomokaku</b> pause well anyway <b>sorekara</b> pause tikaku-ni and.then nearby	“well, anyway” “and then...nearby”
1d. tookyoo-ni <b>umarete</b> tookyoo-ni <b>sodatte</b> Tokyo-in born-and Tokyo-in grow.up-and amerikaziN bakari-kana-to <b>omottara</b> Americans only-wonder-quot. think-and	“(was) born and raised in Tokyo and” “I thought it was just Americans, and”
1e. daitokai pause <b>no</b> seekatu-ni nareteru big.city gen. lifestyle-to used.to	“(I’m) used to big city life”

gion of reduced amplitude (or in the absence of such, the beginning of change in the formant frequencies) was identified as the beginning of the glide. For vowel–vowel sequences, the point of most rapid change in the formant frequencies or  $F_0$  was chosen as the boundary. Listening, in combination with spectrographic cues, played a greater role in the identification of vowel–vowel boundaries than for other transitions. The speech was labeled at a phonetic level, reflecting the phenomena typical of fast, spontaneous speech.

In order to confirm the accuracy of this procedure, the first author independently relabeled all boundaries used in any analysis below for a subset of the data, namely all words with three moras (including particle, if any was produced with the word). This resulted in a total of 364 boundaries from 108 words. The time points of the boundaries placed by the two authors were subtracted, and the absolute values of these differences analyzed. The average error in boundary placement over these 364 measurements was 7.58 ms. The error for most boundaries was quite small: the median error was 4 ms, and 30.5% of boundaries had errors of less than 3.0 ms. These errors are very small relative to the average duration of a mora (136.36 ms for the words which were relabeled). Since most results in this article are based on the correlation of the duration of some larger string with the number of moras in it, these measurement errors are unlikely to affect the strength of the correlations much. Furthermore, there is no significant difference in the size of the measurement errors at the various points which were labeled in the words (beginning and end of the word, beginning of the particle, if any, and end of the first and second moras) [ $F(4,359) = 1.35, p > 0.05$ ]. Therefore, measurement error is not likely to have any systematic effect on the results.

## B. Word boundaries

Determining what constitutes a word in Japanese is not trivial. Japanese orthography does not mark word boundaries, and there are many tightly bound syntactic constructions which may often form a single prosodic unit. Japanese also has many grammatical particles, which usually form a prosodic word with the preceding lexical item. It is important to identify word boundaries correctly in order to maximize the chance of detecting any compensation effect. If compensation takes place across mora boundaries but stops at the word boundary<sup>3</sup> then identifying too small a string as a word might obscure the compensation effect. We included only words for which the boundaries are clear, exemplified in Table I. We included nouns followed by zero, one, or two particles (1a), verbs ending in the polite or nonpolite past or nonpast suffixes or the negative suffix (1b), and adverbs or conjunctions in isolation (1c). We also included verbs ending in the /-te/ suffix (continuative) if not followed by an auxiliary verb, as well as other continuative forms (1d), and particles isolated from the preceding utterance by a pause (1e).

Examples of excluded items appear in Table II. We excluded special constructions which may be grammaticalized (e.g., those with /tame, wake, no hoo, -to yuu/, 2a), nominalized verbs (both /no/ and /koto/ nominalizers, 2b), verbs derived from nouns (2c), anything followed by the copula (2d), and verbs ending in the /-te/ suffix followed by auxiliary verbs (2e). In the case of the copula, for example, there is no way to be sure whether a noun and the copula after it are two different words, or whether the copula is equivalent to an inflection. For all of these constructions, the number of lexical words may not equal the number of prosodic words. We also excluded any form followed by three or more particles or by a very long conjunction (2f), because long strings of

TABLE II. Examples of excluded items. The excluded word and the environment which necessitates its exclusion are shown in bold. Transcriptions are phonemic unless bracketed.

2a. umi-ni <b>tikai tame-ni</b> sea-to near since seekatu-ni <b>nareteru wake</b> desu-ga lifestyle-to used.to reason copula-but <b>inaka-no hoo</b> kara kita country-gen. direction from came boku-no suNderu <b>ryoo-tte yuu-no-wa</b> I-subj living dorm-quot. called-nom.-topic	“since it’s near the sea” “it’s that I’m used to the lifestyle, but” “came from the country” “as for the dorm I’m living in”
2b. syumi-wa piano-o <b>hiku-koto</b> desu hobby-topic piano-obj. play-nom. copula koNsato-ni <b>iku-no-ga</b> suki desu concert-to go-nom.-subj. like copula	“(my) hobby is playing the piano” “(I) like going to concerts”
2c. <b>saNka suru</b> yoo-ni narimasita participation do case-into became	“it became the case that I participated”
2d. <b>yuume</b> <b>datta</b> -node popular copula-past-because	“because (it) was popular”
2e. rai-to yuu mati-ni <b>suNde imasu</b> Rye-quot called town-in live-TE be	“I’m living in a town called Rye.”
2f. <b>tigai-nado-ni-mo</b> yoru difference-etc.-on-also depend maeoki-o <b>simasita-keredomo</b> introduction-obj. did-but	“it depends on differences and such too” “I did an introduction, but”
2g. nineNkaN imasu-kedo <b>maa</b> 2.year.period be-but well	“I’ve been (here) two years, but, well”
2h. ippuN kudasatte <b>arigat[o] gozaimasu</b> 1.minute give-and thank you	“Thank you for giving me one minute”
2i. <b>umare</b> pause <b>u</b> pause [m]mareta-no-wa born (interrupted twice) born-nom.-topic	“As for where (I) was born”
2j. hoka-no kata-wa mina <b>amekaziN</b> -na-node other-gen. person-topic all American-be-because	“Since everyone else is American” (amekaziN for /amerikaziN/)
2k. pause [ <b>ato:</b> ] pause yuuhaN-ga owatte-kara after dinner-subj. finish-after	“after...after dinner was over”

particles may form a separate prosodic word. Interjections (2g) and set phrases (2h) were excluded because they may have abnormal durations for pragmatic reasons. Speech errors with self-repairs were excluded (2i), as were fast speech deletions which result in a form with a different number of moras than it would have in careful speech (2j). Furthermore, utterances in which a segment (usually the final vowel) was held unnaturally long, forming a type of filled pause, were excluded (2k).

All cases which could be considered as compounds (e.g., /tenisu-kurabu/ “tennis club”) were treated as single words. This provides a fairer test of the compensation hypothesis, since compensation is more likely to cross a compound boundary than a word boundary. All words with more than seven moras (without particles) were excluded from all analyses. Port *et al.* (1987) tested words with one to seven moras, so this facilitates comparison of results, and also minimizes the possibility of the few words with many moras (e.g., /wasiNtoNdiisii-koogai/ “Washington D.C.-suburb,” 13 moras) unduly influencing the statistical correlations. These exclusions left 28 to 49 words from each of the 11

speakers, and a total of 416 words. All analyses are based on these words or a subset of them.

### III. ANALYSES

#### A. Correlation of word duration with number of moras

One of the main claims of the across-mora compensation proposal is that the duration of some larger unit such as a word is (nearly) predictable from the number of moras in it, and Port *et al.* (1987, experiment 2) showed very strong correlations ( $r \geq 0.998$ ) between the duration of whole words and the number of moras in the word. However, if the mora is not used for temporal compensation, one would still expect a strong correlation between word duration and number of moras, since moras do contribute segmental material.

We calculated this correlation for each speaker in our corpus separately. Port *et al.* (1987) excluded particles from word duration, so we initially also excluded particles [e.g., in /minami-no/ “South-genitive,” the duration of /minami/ (three moras) was used]. The correlations between whole

TABLE III. (a) Correlation coefficients ( $r$ ) between whole word duration and number of units in the word for each speaker, with moras, syllables, and segments (long vowels as one or two segments) as the unit. Number of words in parentheses. (b) Average word durations by number of each unit. Number of words in parentheses.

(a) Speaker	Moras	Syllables	Segments (VV=1)	Segments (VV=2)
A (34)	0.832	0.595	0.724	0.738
B (42)	0.795	0.598	0.795	0.769
C (31)	0.701	0.596	0.713	0.716
D (39)	0.759	0.567	0.708	0.699
E (49)	0.884	0.729	0.799	0.885
F (46)	0.818	0.669	0.783	0.806
G (31)	0.931	0.746	0.847	0.884
H (32)	0.800	0.812	0.823	0.831
I (28)	0.798	0.554	0.645	0.754
J (48)	0.874	0.819	0.907	0.915
K (36)	0.845	0.651	0.807	0.799

Average word duration (ms)				
(b) No.	Moras	Syllables	Segments (VV=1)	Segments (VV=2)
1	195.85 (10)	206.89 (24)	151.27 (5)	154.89 (1)
2	252.71 (106)	334.35 (199)	214.72 (15)	185.01 (13)
3	369.93 (133)	449.33 (126)	253.62 (35)	246.07 (35)
4	484.41 (120)	518.60 (52)	294.64 (95)	273.60 (79)
5	596.53 (30)	719.63 (11)	389.07 (76)	366.01 (72)
6	649.89 (12)	783.08 (4)	429.73 (91)	425.02 (96)
7	872.13 (5)		522.10 (39)	480.45 (50)
8			545.84 (36)	549.79 (40)
9			657.92 (10)	623.05 (13)
10			657.13 (8)	646.78 (9)
11			773.14 (1)	754.69 (3)
12			741.29 (3)	682.53 (2)
13			883.63 (2)	875.35 (3)

word duration and number of moras in the word, as well as average word durations, are shown in Table III (along with results from Sec. III B). The correlation coefficients vary from  $r=0.701$  to  $r=0.931$ . These correlations are far weaker than those of Port *et al.* (1987). (In that study, word durations were averaged across speakers for each of the 28 utterances before calculating the correlations, minimizing variability. Because ours is a corpus study, one cannot average across speakers' productions of the same utterances, so a direct comparison is not possible.) Although this analysis cannot establish compensation, it does show that predictability of word durations from number of moras is much weaker in spontaneous speech than in careful speech, and that the strength of the relationship between number of moras and word duration varies greatly across speakers.

### B. Correlation of word duration with other units

Port *et al.* (1987) point out that the correlation of word duration with number of syllables for their data would be far worse than the correlation with number of moras, and claim this as evidence for mora-timing. We calculated the correlation of word duration with both the number of syllables and the number of segments in the word (without particles), and compared these results to the correlations with number of moras. The compensation hypothesis claims that moras are used to determine word duration, so it predicts a stronger correlation of word duration with number of moras than with any other unit. However, the hypothesis that the mora only

seems to be regularly timed because of phonological, structural factors predicts that the correlations of word duration with number of moras and number of segments will be about equally strong. This is because there is very little variability in how many segments occur within a mora. This hypothesis also predicts that the correlation with number of syllables will be weaker, because of the large durational difference between one- and two-mora syllables discussed earlier.

For this analysis, the palatalized noncoronal consonants /py, by, ky, gy, hy/ were counted as two segments, while the palatalized coronal segments /sy/ [ʃ], /ty/ [tʃ], /zy/ [dʒ] (which undergo assimilation and are phonetically a single segment) were counted as single segments.<sup>4</sup> Since there is no consensus as to whether long vowels are one or two segments, we calculated the correlations in two ways, with long vowels (e.g., /oo/ in /gakkoo/ "school") counted as one and as two segments. In both cases, sequences of nonidentical vowels (e.g., /ai/ in /ikkai/ "once") and geminate obstruents<sup>5</sup> were counted as two segments. For the syllable count, /(C)ai/ and /(C)oi/ (e.g., /ai/ in /ikkai/ "once," /oi/ in /sugoi/ "great"), as well as syllables containing a long vowel, the mora nasal, or the beginning of a geminate obstruent, were counted as single syllables. /(C)au/ did not occur, and other nonidentical vowel sequences (e.g., /ia/ in /azia/ "Asia," /oa/ in /huroa/ "floor") were counted as two syllables (Vance, 1987).

The results of these analyses appear in Table III. Correlation with number of syllables is significantly weaker than

with number of moras [ $F(1,10)=43.46, p<0.0001$ ]. The correlation with number of segments is significantly weaker than that with number of moras if long vowels are counted as a single segment [ $F(1,10)=6.10, p<0.04$ ], but the difference is not significant if long vowels are counted as two segments [ $F(1,10)=3.08, p>0.10$ ]. Even if long vowels are counted as single segments, the correlation with number of segments is only weaker than that with number of moras for 7 of the 11 speakers.

This shows that a strong correlation between word duration and the number of some unit is not sufficient evidence for that being a unit of timing normalization for the language. It has never been claimed that Japanese (or any other language we know of) is “segment-timed,” and yet segments are, by one count, statistically equivalent to moras in the strength of their relationship to word duration. If one were to calculate, in some arbitrary language, the correlation of word duration with the number of each potential sub-word-level unit (segments, moras, syllables, feet), one of these would be the best predictor of word duration. Which that is may reflect syllable structure constraints, and does not mean that speakers normalize the duration of words to adjust for variability according to this unit.

### C. Particles and the domain of compensation

One important issue for Japanese word duration is grammatical particles, which clearly form part of a prosodic word with the preceding lexical item. Thus, although the genitive particle /no/ in /nihoN-no/ “Japan’s” is not part of the lexical item /nihoN/ “Japan,” it is part of the word for prosodic purposes. If there is compensation, the correlation of duration of some larger unit with number of moras should be strongest if the larger unit is exactly the domain within which compensation occurs. For example, if compensation takes place within the prosodic word /nihoN-no/, but the duration of only the lexical word /nihoN/ is measured, compensation between the final moras of /nihoN/ and the particle /no/ will be disturbed.

Since the proposed across-mora compensation effect is a prosodic process, its domain is more likely to be the prosodic than the lexical word. [Itô and Mester (1995) assume this.] Port *et al.* (1987) do not test the domain of across-mora compensation, but if compensation stops at the boundary of the prosodic word, their theory would predict that the correlation between word duration and number of moras would be stronger when particles are included in the word. (That is, in the string /nihoN-no minami/ “South of Japan,” if there is compensation within /nihoN-no/ and within /minami/ but not between /no/ and /mi/, correlations would be stronger if the duration of /nihoN-no/ is measured than if that of /nihoN/ alone is.) If the domain of compensation is the lexical word (i.e., compensation within /nihoN/ but not between /nihoN/ and /no/), their theory would predict a stronger correlation of number of moras with lexical word duration (particles excluded) than with prosodic word duration (particles included). If there is compensation even across word boundaries (i.e., compensation between /no/ and /mi/ as well as within /nihoN-no/ and /minami/), their theory would predict equally strong correlations whether particles are included in

TABLE IV. (a) Correlations ( $r$ ) between word duration and number of moras, with particles included and excluded. Only words which were produced with a following particle are included in the analysis. Number of words in parentheses. (b) Average durations for these words, by number of moras. Number of words in parentheses. (Since particles can have more than one mora, number of words in the two columns does not match.)

(a) Speaker ( $n$ )	Particles excluded	Particles included
A (21)	0.916	0.785
B (28)	0.818	0.748
C (14)	0.736	0.782
D (23)	0.795	0.618
E (27)	0.924	0.740
F (19)	0.872	0.710
G (16)	0.973	0.936
H (16)	0.872	0.752
I (14)	0.800	0.790
J (30)	0.902	0.853
K (17)	0.901	0.860

(b) No. of moras	Average duration without particles (ms)	Average duration with particles (ms)
1	171.68 (4)	not applicable
2	246.08 (61)	351.69 (3)
3	339.83 (79)	404.12 (54)
4	430.28 (55)	497.79 (72)
5	607.23 (17)	588.79 (57)
6	583.08 (7)	771.88 (22)
7	857.40 (2)	740.96 (11)
8	not applicable	985.13 (4)
9	not applicable	1233.72 (2)

word duration or not, since neither the prosodic nor the lexical word would be the domain of compensation.

However, the hypothesis that there is no mora-based compensation (that each mora simply contributes its variability to the variability of the larger unit, the word) predicts that the correlation between number of moras and whole word duration will be stronger without particles than with them. This is because the particles, as additional moras, contribute additional variability to the variability of whole word duration. Thus, a stronger correlation when particles are excluded than when they are included would favor either the hypothesis that the mora is a nontemporal unit (i.e., no compensation) or the hypothesis that the lexical word is the domain of compensation. Any other outcome would favor the hypothesis that there is compensation within some domain larger than the lexical word.

We calculated the correlation between whole word duration and number of moras in the word, with particles included, for each speaker. Thus, the entire duration of /minami-no/ “South-genitive” (four moras) was used. For this analysis, only words which were produced with a particle were included. (Words such as /daigaku/, /hurimasu/, and /tomokaku/ (1a–c), which were produced with no following particle, were excluded.) We also calculated the correlation between whole word duration and number of moras with particles excluded for these same words. Results appear in Table IV. Of the 11 speakers, 10 (all but speaker C) show a stronger relationship when particles are excluded, and the difference in strength of correlation coefficients is significant [ $F(1,10)=14.17, p<0.004$ ]. Thus, word duration is more

TABLE V. Types of syllable structures into which words were classified, number of words of each type in the corpus, and average mora duration (average duration of moras of each word, then averaged across words of that syllable type). No word is counted in more than one category.

Type	No. of words	Average mora duration (ms)
Geminate obstruent	29	130.30
Mora nasal	51	123.85
Long vowel	52	108.31
Devoiced vowel (word containing a voiceless or deleted vowel in a devoicing environment)	34	130.49
Fast speech deletion (word with an underlying segment elided in a non-devoicing environment)	27	112.05
Multiple (word containing more than one of the above)	47	110.86
Other (word containing only CV or V moras, where no CVV is CV <sub>1</sub> V <sub>1</sub> )	176	132.88

predictable if particles are excluded (leaving duration of the lexical word) than if they are included. For the compensation hypothesis, this is an unlikely result, since it suggests that the domain of compensation is the lexical word rather than the prosodic word or some other unit. However, it is in accord with accumulation of variance: there are simply more moras to contribute variability if particles are included.

#### D. Effects of syllable structure—Special types of moras

The compensation hypothesis predicts that if a syllable containing the mora nasal, a long vowel, or the beginning of a geminate obstruent is not as long as two CV moras, compensation will normalize for such durational differences, keeping the total duration of the word (nearly) predictable from its number of moras regardless of the structures of its syllables. The hypothesis that the mora is a structural, non-temporal unit, however, predicts that non-CV syllables will affect total word duration. This hypothesis claims that the mora-rhythm of Japanese stems not from compensation for non-CV moras, but rather from (among other things) the statistical predominance of CV moras. Port *et al.* (1987, pp. 1577–1578, experiment 2) find little or no effect of syllable type on word duration, but they do not statistically test the apparent low durations of words in their data with geminate obstruents and a devoiced vowel. In another experiment (Port *et al.*, 1987, pp. 1582–1583, experiment 4), they find that CVVCV and CVCCV words are significantly shorter than CVCVCV words, although they are certainly longer than CVCV words. [Han (1994) shows similar results.]

To test the effect of non-CV syllables, we classified all words in the corpus by syllable type. The categories and number of words in them are shown in Table V (along with results described below). This classification is somewhat abstract: fast speech deletions were counted as a separate category, even if the deletion produced one of the other special syllable types phonetically. For example, [bok<sup>w</sup>a] or [bokya] for /boku-wa/ ‘‘I-topic’’ was counted as a fast speech deletion rather than as a devoiced (or deleted) vowel, because the environment for phonological devoicing is not present, and

these forms cannot occur in careful speech. [maintʃi] for /mainiti/ ‘‘every day’’ was classified as a fast speech deletion rather than as a mora nasal. Words containing more than one special mora type were treated as a separate category.

Words of the special syllable types are not equally distributed across the range of number of moras in the word. For example, geminate obstruents can only occur in a word with at least three moras, since they can be neither word-initial nor word-final. Furthermore, individual speakers did not produce words of each special syllable type for each possible number of moras (e.g., eight speakers produced three-mora long-vowel words, but only three produced five-mora long-vowel words). The effect of special mora type on duration thus cannot be investigated for each speaker separately.

We therefore tested the effect of syllable type on the average duration of a mora within each word, using a within subjects analysis of variance. If non-CV syllables affect word duration, this will be reflected as an effect on average mora duration (Table V). The effect of syllable type is significant [ $F(6,58)=5.260$ ,  $p<0.0003$ ]. Words with long vowels, fast speech deletions, or multiple non-CV syllable types are relatively short. While it is somewhat surprising that words with devoiced vowels are hardly shorter than words with no special moras, the fact that there are significant differences among the syllable types argues against compensatory mora-timing. There could be limited compensation which is not strong enough to equalize all differences, but it is clear that special mora types affect duration in significantly different ways.

#### E. Truncating words by removing arbitrary portions

##### 1. Final truncation

As discussed in Sec. III C, Port *et al.* (1987) do not establish whether the word is the domain of compensation, but do claim that compensation functions to make the duration of some unit such as the word predictable from its number of moras. If the domain of the across-mora compensation effect is the word (so that there is no compensation across word boundaries), then when one removes some arbitrary part of the word, the relationship between the duration of the remaining part of the word and the number of moras in it should be weaker than the relationship between duration and number of moras for the entire word. 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 for which the remaining segments are compensating.

For example, if the final mora of a word contains a segment which is inherently shorter than average, such as /r/, under Port *et al.*'s prediction, segments both in that mora and in others lengthen in compensation. If one removed the duration of the final mora, the duration of the remainder of the word would be longer than predicted by the number of moras it contains, because the compensatory adjustments would remain, without the segment /r/ that caused them. The reverse would apply if a segment of the final mora were longer than usual. If a segment in the penultimate, rather than final, mora

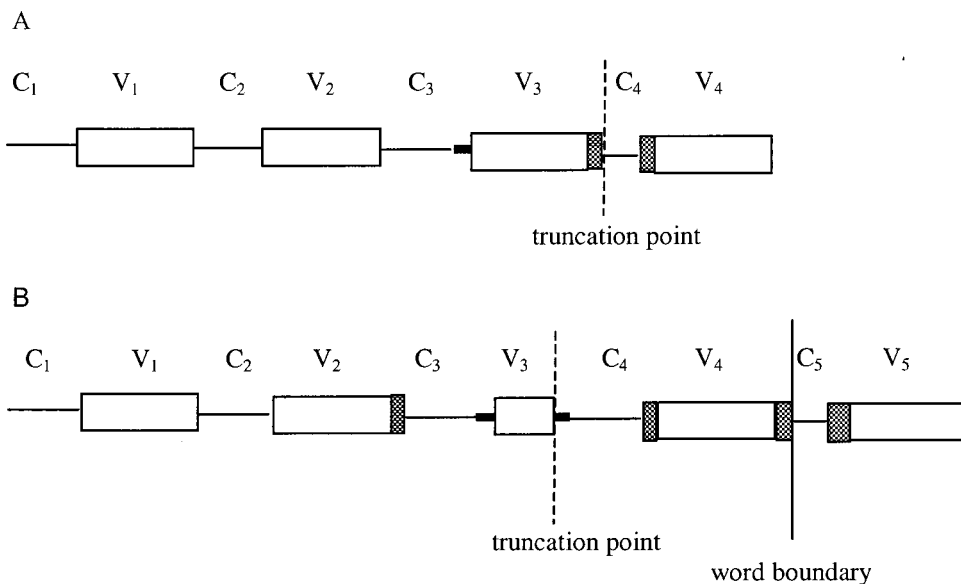


FIG. 1. (a) Durations of consonants (lines) and vowels (boxes) in a hypothetical four-mora word, where the fourth consonant is shorter than average, assuming compensation across moras within the word. Gray portions of boxes and the thick portion of the line for  $C_3$  represent compensatory lengthening to adjust for the short  $C_4$ . The noncompensatory portions of all vowels (white boxes) are of equal length, as are the noncompensatory part of  $C_1$ ,  $C_2$ , and  $C_3$  (thin lines). (b) The same, for a hypothetical four-mora word and the first mora of the following word ( $C_5V_5$ ), where  $V_3$  and  $C_5$  are shorter than average, assuming compensation both within the word and across word boundaries. Lengths of the thin lines (except  $C_5$ ) and the white boxes (except  $V_3$ ) are equal.

were longer or shorter than normal, the final mora (among others) would compensate for it, so removing the final mora would leave uncompensated variation. Thus, removing any arbitrary portion of the word would leave undercompensated or overcompensated variability, so that the duration of the remaining portion of the word would not be predicted as well by the number of moras in it as the duration of the entire word is.

This is demonstrated schematically in Fig. 1(a), which shows a hypothetical four-mora word in which the fourth consonant is shorter than average (perhaps /r/). The preceding and following vowels, as well as the consonant before that, are lengthened to compensate (shown as the gray boxes and thick line). If the final mora is removed, the compensatory lengthening of the penultimate consonant and vowel will remain, making the remaining portion of the word longer than would be expected based on its mora count. Such disturbances of the compensation relationship would lower the predictability of the duration of the remaining portion. However, the hypothesis that moras simply contribute variability to the variability of larger units, without compensation, predicts that the duration of an arbitrarily truncated part of a word will be more predictable from the number of moras in it than the duration of the whole word is. This is because a truncated word has fewer moras—truncation removes sources of variability by removing moras.

Port *et al.* (1987) do not rule out the possibility that compensation applies within some unit larger than the word, resulting in compensation even across word boundaries (e.g., between /no/ and /mi/ in /nihoN-no minami/ “South of Japan”). If there were substantial compensation across word boundaries, they could not have found the strong correlation between word duration and number of moras that they did. However, even if there is compensation within some unit larger than the word, so that moras spanning a word boundary also compensate for each other, the compensation hypothesis would still predict that truncation would lower the predictability of duration, as long as there is more compensation among the moras within the word than across the

word boundary. Figure 1(b) shows a hypothetical four-mora word followed by another mora (which is in a different word), with the third vowel and the fifth consonant shorter than average. The durations of the second and fourth vowels and the third and fourth consonants are lengthened to compensate for the short third vowel, and the fourth and fifth vowels are lengthened to compensate for the short fifth consonant. Unless there is at least as much compensation across the word boundary as between the third and fourth moras, removing the last mora of the word (the fourth mora) will still reduce the predictability of the duration of the first three moras. In our corpus, it is unlikely that there is on the average more compensation across word boundaries than within words, since some words precede a pause.

We tested this prediction of the across-mora compensation hypothesis by measuring the durations of words without their final two moras (e.g., the duration of /kagosi/ in the word /kagosimasi/ “Kagoshima City”). The correlation of these truncated durations with the number of moras in the remaining part of the word was then calculated. Particles were excluded before truncation. The final two moras do not form any prosodic unit, and constitute an arbitrary portion of the word. Removing only one mora might not be enough to demonstrate a clear effect (either of increased or decreased predictability), while removing more than two moras would make many words too short to examine. Words with only one or two moras were, of course, excluded from this analysis. Words in which the boundary between penultimate and antepenultimate moras falls during a geminate obstruent (e.g., /ikkai/ “one time”) or long vowel (e.g., /nyuuyooku/ “New York”) were also excluded, as it is difficult to place an accurate mora boundary within these segments. To allow accurate comparison of truncated and nontruncated words, we also calculated the usual whole word correlation for exactly those words which could be used for the truncation analysis.

As shown in Table VI, the correlations between number of moras and duration are significantly stronger for the truncated words than for the corresponding whole words

TABLE VI. (a) Correlation coefficients ( $r$ ) of final-truncated duration and whole word duration with number of moras in the truncated or whole word string. Number of words in parentheses. (b) Average durations for the same words. Number of words in parentheses.

(a) Speaker	Truncated	Entire word
A (26)	0.947	0.790
B (25)	0.915	0.749
C (20)	0.934	0.702
D (26)	0.846	0.745
E (34)	0.952	0.870
F (30)	0.930	0.745
G (21)	0.936	0.904
H (18)	0.910	0.577
I (21)	0.922	0.707
J (25)	0.862	0.781
K (21)	0.879	0.766

(b) No. of moras	Truncated (ms)	Whole word (ms)
1	118.81 (125)	not applicable
2	224.16 (103)	not applicable
3	356.14 (23)	374.68 (125)
4	416.99 (12)	484.51 (103)
5	590.22 (5)	616.22 (23)
6	not applicable	649.89 (12)
7	not applicable	872.13 (5)

[ $F(1,10) = 36.03, p < 0.0002$ ]. This result contradicts the hypothesis of across-mora compensation, and supports the hypothesis of accumulation of variance. Rather than disturbing the compensation relationships within the word and making the duration of the remainder less predictable, removal of a portion of the word makes the duration of the remainder significantly more predictable by removing sources of variability.

The compensation hypothesis proposed by Port *et al.* (1987) could only account for this result if many words had compensation across the word boundary, but less or none across the truncation point. This is unlikely, since there is no reason for durations to be particularly stable at the arbitrary truncation point, but compensation might well stop at the word boundary, and words which were followed by a pause cannot possibly have compensation across the word boundary. However, a modified version of the earlier hypothesis that there is compensation only within the mora might be able to account for this result with the addition of a final-lengthening effect. If compensation made each mora more similar in duration to other moras, but an additional final-lengthening effect, independent of mora-timing, made the last few moras of a word longer, then the last few moras of a word would be the most divergent from the average duration of a mora because they would be the only ones subject to any effect other than normalization. Thus, their exclusion would improve the predictability of the duration of the remaining string.

## 2. Initial truncation

The question of whether the final truncation results indicate traditional mora-timing (within-mora compensation) plus final lengthening or simply the lack of compensation (accumulative variance) can be addressed by removing the

TABLE VII. Same as Table VI, for initial truncation.

(a) Speaker	Truncated	Entire word
A (23)	0.810	0.796
B (23)	0.844	0.758
C (20)	0.827	0.702
D (26)	0.885	0.712
E (33)	0.937	0.902
F (32)	0.815	0.806
G (19)	0.884	0.902
H (16)	0.568	0.534
I (21)	0.775	0.717
J (22)	0.874	0.776
K (19)	0.846	0.453

(b) No. of moras	Truncated (ms)	Whole word (ms)
1	130.16 (110)	not applicable
2	261.41 (102)	not applicable
3	364.92 (28)	371.56 (110)
4	433.51 (11)	485.74 (102)
5	571.70 (3)	595.15 (28)
6	not applicable	638.34 (11)
7	not applicable	863.54 (3)

initial two moras of the word instead of the final two. If the stronger correlations under final truncation mean that moras tend toward the same duration with the exception of the lengthened final few, then removing the initial two moras should lead to weaker correlations, because one would be removing two of the least divergent moras. The across-mora compensation hypothesis of Port *et al.* makes the same prediction as with final truncation: weaker correlations when any arbitrary portion of the word is removed. The accumulation of variance hypothesis (lack of compensation) also makes the same prediction as for final truncation: correlations should be stronger when moras are removed, whichever moras those are.

For this analysis, the procedures were identical to those of the final truncation analysis, but correlations were calculated with the initial two moras removed instead of the final two (i.e., the duration of /simasi/ in /kagosimasi/ “Kagoshima City” was measured). Words in which the boundary between the second and third moras falls during a geminate obstruent or long vowel, such as /tyotto/ “a little” and /izyoo/ “above,” were excluded. These are not always the same words as must be excluded for final truncation. The correlation of (untruncated) whole word duration with number of moras was also calculated for exactly those words which could be used for the initial truncation analysis.

The choice to exclude the initial two moras is motivated by the same factors as the truncation of the final two moras, but the initial two moras of a word do usually constitute a foot in Japanese, and thus could form a prosodic unit. However, mora-timing has not to our knowledge been proposed to depend on foot structure. There is relatively little evidence of foot structure influencing duration in Japanese, and this evidence regards highly rhythmic speech (Tajima, 1998). Since the speech in our corpus is relatively fast, foot structure is unlikely to influence the predictions of the various theories regarding initial truncation.

Table VII shows the correlations of duration of initial-

truncated words with number of remaining moras, as well as correlations for the same words intact. All speakers except speaker G have a stronger relationship between duration and number of moras for initial-truncated words than whole words, and this difference in the correlation coefficients is significant [ $F(1,10) = 7.03, p < 0.03$ ].<sup>6</sup> This result contradicts the across-mora compensation hypothesis, and also rules out the possibility that the final-truncation result was due to within-mora compensation (tendency toward isochrony) plus final lengthening, but supports the hypothesis that moras simply contribute variability to the variability of the whole. It is not simply that the final two moras are the most variable. Rather, removing either the final or the initial two moras decreases variability in the remainder of the word. One could extend this test by removing other arbitrary portions of words, but it appears thus far that no matter what portion of the word is removed, the duration of the remainder of the word is more predictable than duration of the whole word.

For most speakers, final truncation improves the predictability of duration by more than initial truncation does. This may indicate a final lengthening effect without any tendency toward isochrony. We tested for the presence of final lengthening in this corpus in several ways. We found that the average duration of moras located in the final two moras of the word (exclusive of particle), 128.71 ms, was significantly longer than the average duration of moras earlier in the word, 115.40 ms [ $F(1,267) = 18.33, p < 0.001$ , excluding the same words as in the final truncation analysis]. The average duration of a mora within the first two moras of the word, 116.16 ms, is significantly shorter than the average duration of later moras, 128.31 ms [ $F(1,253) = 11.43, p < 0.002$ ]. The average duration of moras in particles, which are always final (151.50 ms), is significantly longer than the average duration of moras in the word stem (115.77 ms) [ $F(1,224) = 62.03, p < 0.001$ ]. Finally, words produced without any particle have significantly higher average mora duration, 134.76 ms, than the stems of words produced with particles (115.77 ms) [ $F(1,10) = 16.68, p < 0.003$ ], indicating that when particles are present, final lengthening affects them rather than the stem.

These tests indicate that there is final lengthening in this data. We do not investigate the domain of final lengthening, and have no reason to doubt Kaiki, Sagisaka, and colleagues' finding that it is the last mora of the "breath group" that is lengthened in Japanese (Takeda *et al.*, 1989; Kaiki *et al.*, 1992a, b; Sagisaka, 1992; Kaiki and Sagisaka, 1992, 1993). Our results mean simply that the boundaries of the units to which final lengthening applies often enough coincide with word boundaries to show a significant lengthening effect at word boundaries. This clarifies the difference between the final- and initial-truncation results. Because of final lengthening, the final two moras of a word diverge from the typical duration of a mora by more than other moras in the word do. Final lengthening indicates that there is at least one form of speaker control or planning of duration: it is one way in which mora durations systematically diverge from random variance. However, final lengthening does not require any relationship to the mora, and is orthogonal to the issue of mora-timing.

#### IV. DISCUSSION

These results contradict the mora-timing hypothesis based on across-mora compensation as put forth by Port *et al.* (1987). They support the hypothesis that moras simply contribute variability to the variability of the word, without compensation. This is therefore evidence that the importance of the mora in Japanese rhythm stems from structural factors in the phonology of the language rather than durational normalization. We have shown that the relationship between word duration and number of moras is far weaker in spontaneous speech than in the careful speech of the previous literature, and that the mora is only a slightly better predictor of word duration than the segment, a nonprosodic unit (Secs. III A and B). We have also shown that word duration is more predictable within the lexical than the prosodic word, which would be an unexpected result for a prosodic compensation effect but is predicted by the hypothesis that variance accumulates (Sec. III C). There is also a significant effect of non-CV syllable structures on word duration (Sec. III D).

Finally, our analyses of word duration predictability in truncated and whole words (Sec. III E) provide very strong evidence against compensatory mora-timing, particularly against the hypothesis that compensation serves to normalize the duration of some higher unit such as the word. As discussed earlier, the only way, under this hypothesis, that truncating words could increase the predictability of their durations would be if there were compensation across word boundaries, but little or none within the word at the truncation point, which is unlikely. However, both types of truncation show an increase in predictability of duration under truncation.

All of the results are consistent, however, with the hypothesis that the mora is not used for normalizing duration, and that its importance in Japanese rhythm rather derives from phonological factors. If there is no compensation, then the simplest assumption is that each mora (or other subword unit) contributes variability to the variability of the larger unit, the word. That is, variance accumulates (Ohalo, 1975). The variance of a word with many moras is greater than the variance of a word with few moras, because there are more sources of variability. By showing that the duration of whole words is less predictable than the duration of truncated words, we have shown that the truncated moras also contribute variability to the duration of the whole. They do not reduce the variability of the whole, as they would if there were a compensation relationship between the truncated portion and the remaining portion. The comparison of correlations with and without particles also supports this finding, since removing particles is a type of truncation as well. The only systematic effect on word duration for which we find evidence is final lengthening, which is not related to mora-timing.

Our results do show that for most speakers, the mora is the unit which best predicts word duration in Japanese (among the units tested). If mora-timing were re-defined to mean only that the mora is for most speakers the best predictor of word duration in the language (a definition without compensation or isochronous moras), then one could state that Japanese is mora-timed. However, this is not one of the

definitions of mora-timing which appears in the literature.

Since the mora is clearly relevant in the segmentation, and generally the psycholinguistic processing, of Japanese, the evidence against compensatory mora-timing brings up the question of what makes the mora psycholinguistically important, if it is not a tendency toward regular timing. We conclude that the mora is *relatively* regularly timed in Japanese (in that it is more regularly timed than other prosodic units), and is a relatively good predictor of word duration, not because of compensation but because of phonological, structural factors. These include the lack of an effect of pitch accent on duration, the lack of reduction of unaccented vowels, the lack of an opposition or alternation between accented and unaccented vowels, the statistical predominance of CV syllables in Japanese, and the lack of perceptual cues other than duration difference to the phonemic length distinctions. We propose that these factors, without any durational compensation, create the mora-rhythm of Japanese, which is what listeners make use of in parsing. Thus, we do not conclude that the mora is a historical relic maintained only by orthography (as in Beckman, 1982). The mora is important in Japanese, but it is factors in the structure of the language, not durational compensation, which create its mora rhythm.

## ACKNOWLEDGMENTS

We thank Anne Cutler, Steven Greenberg, Setsuko Imatomi, Hajime Inozuka, Masahiko Komatsu, Terry Nearey, John Ohala, Toni Rietveld, and Roel Smits for discussions of this material. We are also grateful to the four volunteer Japanese speakers who judged the recordings, and to Pieter Meima and Bram de Kruijff for help with the data. Any errors are, of course, our own. Some of these results were presented at the ICPHS 1999 in San Francisco, and a preliminary version of part of this research appeared in the proceedings of that conference.

<sup>1</sup>The only analyses which do not use the same words in all conditions are the syllable structure analysis (Sec. III D) and one test for final lengthening [comparing words produced with and without particles (Sec. III E 2)].

<sup>2</sup>One speaker, A, came from Kagoshima, the dialect of which is said to have syllable-timing rather than mora-timing. This speaker's results are included here nonetheless, because they do not seem to differ from other speakers' results. In particular, speaker A shows a very weak effect of number of syllables on duration, just as other speakers do (Table III). The speaker may have used a variety of Japanese which is rhythmically similar to the standard dialect during the recording because he was not addressing members of his own dialect community. There is some evidence from pitch accent that the speaker was attempting to speak Standard Japanese in the recording. The other speakers of nonstandard dialects were not from Kagoshima.

<sup>3</sup>Port *et al.* (1987) do not rule out the possibility of compensation across word boundaries, but since the recent literature on mora-timing focuses on predictability of word duration, we chose a careful definition of the word.

<sup>4</sup>Even noncoronal palatalized segments are sometimes analyzed as single segments on phonological grounds, thus /kʲ/ instead of /ky/. We also performed this analysis with all palatalized segments counted as single segments, with similar results.

<sup>5</sup>Geminate obstruents are best analyzed as two segments: they make up the coda of one syllable and the onset of the next. The fact that they cannot occur word-initially or word-finally supports this analysis. There are reasons based on the pitch accent system to count long vowels as two segments, as well (Vance, 1987), but the situation is less clear than for geminate obstruents.

<sup>6</sup>Since the closure portion of a post-pausal word-initial voiceless stop cannot be measured, the closure duration of word-initial voiceless stops was in-

cluded in the total word duration only if the stop followed some other speech sound rather than a pause. This unavoidable difference in measurements could increase the variability of the word-initial mora. Therefore, we also performed the initial truncation analysis excluding any voiceless stop-initial words which follow a pause (those for which closure duration could not be measured). This leaves rather few words (as few as 11 for one speaker), so the correlations are less reliable. Still, 9 of the 11 speakers have a stronger correlation for initial truncated words than for the same words intact, although the difference is not quite significant [ $F(1,10) = 4.45, p = 0.061$ ].

- Arai, T., and Greenberg, S. (1997). "The Temporal Properties of Japanese are Similar to those of English," in Proceedings of the 5th European Conference on Speech Communication and Technology, Vol. 2, pp. 1011–1014.
- Beckman, M. (1982). "Segment Duration and the 'Mora' in Japanese," *Phonetica* 39, 113–135.
- Beckman, M. (1992). "Evidence for Speech Rhythms across Languages," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 458–463.
- Bradlow, A., Port, R. F., and Tajima, K. (1995). "The Combined Effects of Prosodic Variation on Japanese Mora Timing," in Proceedings of the International Congress of Phonetic Sciences, Vol. 4, pp. 344–347.
- Campbell, N. (1992). "Segmental Elasticity and Timing in Japanese Speech," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 403–418.
- Cutler, A., and Otake, T. (1994). "Mora or Phoneme? Further Evidence for Language-specific Listening," *J. Memory Lang.* 33, 824–844.
- Dauer, R. M. (1983). "Stress-timing and Syllable-timing Reanalyzed," *J. Phonetics* 11, 51–62.
- Dauer, R. M. (1987). "Phonetic and Phonological Components of Language Rhythm," in Proceedings of the Eleventh International Congress of Phonetic Sciences, Vol. 5, pp. 447–450.
- Han, M. S. (1962). "The Feature of Duration in Japanese," *Onsei no Kenkyu* 10, 65–80.
- Han, M. S. (1994). "Acoustic manifestations of mora timing in Japanese," *J. Acoust. Soc. Am.* 96, 73–82.
- Hoequist, Jr., C. (1983a). "Durational Correlates of Linguistic Rhythm Categories," *Phonetica* 40, 19–31.
- Hoequist, Jr., C. (1983b). "Syllable Duration in Stress-, Syllable-, and Mora-timed Languages," *Phonetica* 40, 203–237.
- Homma, Y. (1981). "Durational Relationship Between Japanese Stops and Vowels," *J. Phonetics* 9, 273–281.
- Itô, J., and Mester, R. A. (1995). "Japanese Phonology," in *A Handbook of Phonological Theory*, edited by J. A. Goldsmith (Blackwell, Cambridge), pp. 817–838.
- Kaiki, N., and Sagisaka, Y. (1992). "The Control of Segmental Duration in Speech Synthesis Using Statistical Methods," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 391–402.
- Kaiki, N., and Sagisaka, Y. (1993). "Prosodic Characteristics of Japanese Conversational Speech," *IEICE Trans. Fundamentals* E76-A, 1927–1933.
- Kaiki, N., Takeda, K., and Sagisaka, Y. (1992a). "Gengojoohoo o Riyoo shita Boin Keizoku Jikanchoo no Seigyo [Vowel duration control using linguistic information]," *IEICE Trans.* J75-A, 467–473.
- Kaiki, N., Takeda, K., and Sagisaka, Y. (1992b). "Linguistic Properties in the Control of Segmental Duration for Speech Synthesis," in *Talking Machines: Theories, Models, and Designs*, edited by G. Bailly, C. Benoit, and T. R. Sawallis (Elsevier Science, Amsterdam), pp. 255–263.
- Kondo, M. (1999). "Manifestation of Lexical Accent and Timing Strategy in English Speakers' Japanese," in Proceedings of the Fourteenth International Congress of Phonetic Sciences, pp. 1467–1470.
- Kubozono, H. (1999). "Mora and Syllable," in *The Handbook of Japanese Linguistics*, edited by N. Tsujimura (Blackwell, Malden, MA), pp. 31–61.
- Minagawa-Kawai, Y. (1999). "Preciseness of Temporal Compensation in Japanese Mora Timing," in Proceedings of the Fourteenth International Congress of Phonetic Sciences, pp. 365–368.
- Muthusamy, Y. K., Cole, R. A., and Oshika, B. T. (1992). "The OGI Multi-language Telephone Speech Corpus," in Proceedings of the Third International Conference on Spoken Language Processing, pp. 895–898.

- Ohala, J. J. (1975). "The Temporal Regulation of Speech," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, London), pp. 431–453.
- Otake, T. (1989). "Counter Evidence for Mora Timing," in Proceedings of the 16th Lacus Forum, pp. 313–322.
- Otake, T. (1990). "Gengo no Rizumu to Onsetsu Koozoo" [Rhythmic Structure of Japanese and Syllable Structure], IEICE Technical Report 89, pp. 55–61.
- Otake, T., Hatano, G., and Yoneyama, K. (1996). "Speech Segmentation by Japanese Listeners," in *Phonological Structure and Language Processing*, edited by T. Otake and A. Cutler (Mouton de Gruyter, Berlin), pp. 183–201.
- Otake, T., Hatano, G., Cutler, A., and Mehler, J. (1993). "Mora or Syllable? Speech Segmentation in Japanese," *J. Memory Lang.* **32**, 258–278.
- Port, R. F., Al-Ani, S., and Maeda, S. (1980). "Temporal Compensation and Universal Phonetics," *Phonetica* **37**, 235–252.
- Port, R. F., Dalby, J., and O'Dell, M. (1987). "Evidence for Mora Timing in Japanese," *J. Acoust. Soc. Am.* **81**, 1574–1585.
- Ramus, F., Nespors, M., and Mehler, J. (1999). "Correlates of linguistic rhythm in the speech signal," *Cognition* **73**, 265–292.
- Sagisaka, Y. (1992). "On the Modeling of Segmental Duration Control," in *Speech Perception, Production, and Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (Ohmsha, Tokyo), pp. 451–455.
- Sagisaka, Y. (1999). "Koopasu Beesu Onsei Goosei: Onsei Kagaku Chishiki ni Motozuku Goosei Shisutemu Koochiku Gijutsu no Shin Paradimu" (Corpus-based speech synthesis—A new paradigm for synthesis system building based on the knowledge in speech science), in Proceedings of the Meeting of the Acoustical Society of Japan, September–October, pp. 197–200.
- Sagisaka, Y., and Tohkura, Y. (1984). "Phoneme Duration Control for Speech Synthesis by Rule," *IEICE Trans.* **J67-A:7**, 629–636.
- Sato, Y. (1995). "The Mora Timing in Japanese: A Positive Linear Correlation between the Syllable Count and Word Duration," *Onsei Gakkai Kaihoo* **209**, 40–53.
- Tajima, K. (1998). "Speech Rhythm in English and Japanese," unpublished dissertation, Indiana University, Bloomington.
- Takeda, K., Sagisaka, Y., and Kuwabara, H. (1989). "On sentence-level factors governing segmental duration in Japanese," *J. Acoust. Soc. Am.* **86**, 2081–2087.
- Tsujimura, N. (1996). *An Introduction to Japanese Linguistics* (Blackwell, Malden, MA).
- Vance, T. J. (1987). *An Introduction to Japanese Phonology* (State Univ. of New York, Albany, NY).
- Warner, N., and Arai, T. (2001). "Japanese Mora-timing: A Review," *Phonetica* **58**, 1–25.