Japanese Mora-Timing: A Review

Natasha Warner a Takayuki Arai b

aMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and
bDepartment of Electrical and Electronics Engineering, Sophia University, Tokyo, Japan

Abstract

Japanese is often called a ‘mora-timed’ language, and contrasted with ‘stress-timed’ or ‘syllable-timed’ languages. The definition of what constitutes mora-timing has undergone several revisions, and a wide variety of experimental evidence both for and against mora-timing has been presented. This article reviews the hypotheses, the means of testing them, and the results of nearly 40 years of experimental work on mora-timing in Japanese, and suggests directions for future work in this area.

Introduction

Japanese is very often referred to as a ‘mora-timed’ language, and is contrasted with ‘stress-timed’ languages such as English and ‘syllable-timed’ languages such as French. The simplest definition of this typology is that these units are regularly timed in their respective languages, such that morae recur regularly and each take the same amount of time in Japanese, while stresses recur regularly in English. However, it is clear that these units are not strictly isochronous, and most tests of mora-timing focus on compensation, the idea that speakers adjust the durations of segments to make morae more regularly timed than they would otherwise be. In this article, we review the literature on experimental tests of mora-timing in Japanese. Through a comparison of methodologies, we hope to clarify what can be concluded from which types of tests. Furthermore, we hope to draw attention to areas of Japanese speech timing which have been relatively neglected.

Types of Variation

At least three factors cause mora duration to vary. First, there are inherent differences in segmental durations: high vowels are shorter than low vowels, [ε] is very short while [æ] is relatively long. Thus, in the absence of compensation, the mora /ki/ would be shorter than /ka/, and /ra/ ([ra]) shorter than /sa/.

Second, there are several types of morae in Japanese which do not consist of a consonant-vowel string, and these might not have the same duration as an average CV mora. Long vowels have two morae, so that /kookoo/ ‘high school’ has the same num-
ber of morae as /tokonoma/ ‘alcove’. Geminative obstruents (transcribed as /Q/ followed by an obstruent) contribute a mora, while single obstruents do not, since they are always in the syllable onset: /kaQta/ ‘won’ has three morae, just as /kaketa/ ‘suspended’ does, while /kata/ ‘shoulder’ has two. The mora nasal (transcribed /N/) also contributes a mora, so that /kaNdai/ ‘chewed’ has three morae. Finally, high vowels between voiceless obstruents or between a voiceless obstruent and the end of an utterance can be devoiced or deleted in Japanese, but their mora remains. Thus, /kita/ [kita] ‘North’ has two morae, just as /kata/ ‘pattern’. A word with any of these special types of morae is unlikely to have inherently the same duration as a word with the same number of CV morae.

Third, there are numerous higher level effects on duration, particularly in connected speech. Kaiki and Sagisaka [1992], in an investigation of a corpus of read speech, document effects of factors such as position of the segment in the sentence, number of morae in the ‘breath group’, and part of speech of the word containing the segment. Because most experimental tests of mora-timing control for at least some of these effects by placing test words in a set frame sentence, these effects are rarely discussed. Inherent durational differences among segments and the effect of non-CV morae are the factors most often used to test mora-timing.

The Early Non-Experimental Claim of Mora-Timing

The concept of the mora as a unit related to timing in Japanese is very old. Jinbo [1980 (1927)] defines the ‘onsetsu’ as a unit consisting of a CV sequence (a mora), and claims that such units have approximately equal duration in Japanese. He contrasts the resulting rhythm of Japanese with the stress-based rhythm of English. Trubetzkoy [1958 (1939)] classifies Japanese as having the mora rather than the syllable as its prosodic unit, although his criteria for this division are based on phonological properties rather than regular duration. Bloch [1950] states that morae in Japanese all have the same duration, or are perceived as being of equal length. He uses the word ‘syllable’ instead of ‘mora’, but his definition of the syllable in Japanese is clearly equivalent to the mora, as his treatment of non-CV morae shows. (He also uses ‘fraction’ as a non-theoretical term for the mora.) Bloch emphasizes that words with the same number of morae are perceived as having the same duration, regardless of the segmental content of the words, and regardless of what non-CV morae are present. Furthermore, he says that ‘a phrase containing twice or three times as many fractions [morae] as another is heard as lasting just twice or three times as long’ [Bloch, 1950, p. 91].

Although Bloch seems to claim that individual morae have equal (perceptual) duration (‘All these fractions are heard as having the same time value’ [1950, p. 91]), his description emphasizes the total duration of words or phrases. He even suggests that one can determine the number of morae in a word ‘by comparing its duration with that of another phrase in which the number of such fractions is known’ [Bloch, 1950, p. 92]. Bloch does not distinguish between intervals which are perceived as having the same duration and intervals which have the same physical duration, but his description of Japanese timing is phrased in terms of perception. Ladefoged [1993 (1975)] clarifies the proposal by pointing out that the mora is the relevant unit, and claims that ‘each mora takes about the same length of time to say’ [1993, p. 251]. Hattori [1980 (1961)] also discusses the importance of the mora in Japanese, and states that heavy syllables (those containing a long vowel, the mora nasal, or the beginning of a geminate obstruent) are approximately twice as long as a light CV syllable.

---

Structure of this Paper

This review of the literature on mora-timing is organized not chronologically, but by the type of mora-timing hypothesis, and the means used to test it. Thus, results from one article are not all discussed together, but this allows us to compare all the results on a particular topic more effectively. We discuss first the basic hypothesis that morae tend toward being equally timed, then the hypothesis that morae are used to normalize word duration. Then we consider several alternative definitions of what makes Japanese mora-timed, and finally turn to studies which examine timing differences through the acquisition of second languages. We will not discuss the literature on the timing of English or other languages, except as necessary for comparison with Japanese.¹

The Tendency toward Isochrony Mora-Timing Hypothesis

Because of the effects of inherent segmental duration and differences in moraic structure (non-CV morae), it is clear that morae are not completely isochronous in Japanese. Han [1962b], in an early experimental investigation of Japanese timing, states that the mora is a unit of duration in Japanese, and that all morae are of approximately (but not exactly) equal length. She says this is achieved through compensation. These ideas, the nearly equal duration of different morae and compensation, are the focus of most early work on mora-timing. Two main methods are used to test this version of the mora-timing hypothesis: the calculation of duration ratios for various portions of words, and tests for correlation between durations of adjacent segments.

Duration Ratios

Many researchers calculate the ratio of duration of various segments or portions of words in order to investigate mora-timing. Ratios of geminate to singleton consonants, and long to short vowels, are particularly popular evidence, since strict mora-timing predicts that the long segments contribute the duration of an entire CV mora to the word. However, it is difficult to be certain what ratio would indicate mora-timing, as discussed below. Comparison between the duration of the mora nasal and non-moraic nasals is also relevant. Aside from these comparisons of single segments or geminates, investigators have also compared durations of 2 mora and 3 mora strings, for example, which may include a non-CV mora. Here, mora-timing straightforwardly predicts that the duration ratio of a 2 mora string to a 3 mora string should be 2:3, for example, regardless of the composition of those morae.

Han [1962b] finds that the ratio of duration of geminate to singleton obstruents is approximately 2.6:1, but sometimes as high as 3:1. She interprets this as meaning that the sequence /Q/ in /kaQta/ has the duration of an entire mora plus the duration of a single /Q/. For long and short vowels, she calculates a ratio between 2:1 and 3:1, depending on whether there is an onset consonant in the syllable, and what it is. Homma [1981] finds that the duration ratio of geminate to singleton stop closures (thus excluding VOT) is between 2.6:1 and 3.3:1. However, the argument that a 3:1 ratio means the geminate or long vowel is contributing the duration of a mora plus the

¹ There is, of course, a large body of work on timing in other languages, particularly in English. For examples of this work, as well as references to other work on English duration, see Lehiste [1977, 1980, and other works].
duration of a single consonant or vowel relies on the untenable assumption that single consonants and vowels are of approximately equal duration. (Campbell and Sagisaka [1991] determine that average duration is 54 ms for non-geminate consonants and 80 ms for phonemically short vowels, for example.) Furthermore, Beckman [1982] finds that the ratio of geminate to singleton consonants varies greatly depending on whether VOT is counted as part of the stop or not, but in any case it does not reach 3:1. Han's [1962b] work is based on measurements of relatively few spectrograms, because of the technical limitations of the time.

Campbell and Sagisaka [1991] find that geminate consonants are approximately three times as long as single ones, but that long vowels are only about 1.5 times as long as short ones in a corpus of read speech. Arai's [1999] results for long and short vowels are similar, with a slightly higher ratio for read speech than spontaneous speech. Thus, long vowels in more natural speech are certainly not three times as long as short vowels. Neither Campbell and Sagisaka nor Arai calculate these ratios for syllables with and without onsets consonants, as Han does, but regardless of duration of the preceding consonant, it is clear that long vowels are not long enough to contribute the duration of an entire mora.

Sato [1993] calculates a ratio of approximately 2.4:1 for moraic to non-moraic nasals. She concludes that this provides support for at least a weaker version of mora-timing because it is considerably higher than the ratios she obtains for nasals in syllable code and onset position in English and Korean, and because moraic nasals are at least somewhat longer than non-moraic nasals. Although Otake et al. [1996] do not address mora-timing, they replicate this ratio for moraic to non-moraic nasals.

Ratios for longer portions of words haver also been investigated. Because these comparisons usually involve entire morae instead of segments, they avoid the problem of assuming equal duration of consonants and vowels. Homma [1981] finds that the ratio of duration of a CVQCVC word to a CVCV word is approximately 3:2, and argues that each mora contributes approximately the same duration to a word. However, Homma's material includes voiced geminates, which are rare and not representative of the native Japanese vocabulary. Han [1994] replicates this result with real words. However, in a comparison of geminate obstruent QCV to singleton obstruent CV strings (extracted from VQCVC or VCV environments), Beckman [1982] obtains a ratio of 1.66:1 instead of the predicted 2:1.

Hoequist [1983a] calculates the ratio of Japanese CVN to CV durations, with a result of approximately 1.8:1. He points out that this is less than the predicted 2:1 ratio, and not much more than the result of 1.6:1 he finds for Spanish (a syllable-timed language). Sato's ratios for the same structures [1993] range from 1.4:1 to 1.8:1, and are lower than comparable ratios in Korean, but higher than English. Hoequist [1983a] finds a ratio of 1.7:1 for CVV to CV words in one study, but ratios nearer to 2:1 in subsequent work with reiterant speech [Hoequist, 1983b]. He also observes [Hoequist, 1983a] that CVN syllables are significantly longer than CVV syllables in Japanese, although both have two morae. Hoequist concludes that isochrony may not be the best way to describe the cross-linguistic differences in rhythm.

In sum, many different researchers have calculated ratios both of duration of phonologically long to short segments, and of longer portions of words with varying syllable structures, but the results are inconclusive and highly variable across studies. Most studies investigating ratios of long and short vowels indicate that long vowels are too short to contribute the duration of an entire mora to the word. The remaining results are too inconsistent to allow a clear conclusion. The variability across studies may reflect differences in the test words, or in the formality of the speech used (which ranges from nonsense words with non-native segments to spontaneous speech).

Tests of Correlation between Duration of Adjacent Segments

Using ratios to test mora-timing emphasizes the potential effects of non-CV morae on isochrony. However, such ratios do not test whether speakers normalize the inherent differences in duration among vowels or among consonants. The main mechanism proposed to normalize for shorter or longer than average segments is compensation between adjacent segments, particularly between the consonant and vowel of a CV mora. Therefore, many researchers have investigated whether there is a negative correlation between durations of adjacent segments, either by examining particular pairs chosen to demonstrate inherent variation, or by calculating correlations over all CV morae in a corpus.

Port et al. [1980] investigate vowel duration before and after several consonants in Japanese and Arabic. They determine that in Japanese, both the preceding and following vowel are longer when the intervening consonant is inherently short ([r]) than when it is long ([l]). In Arabic, only the duration of the preceding vowel varies inversely with that of the consonant, and the differences between consonants are not as clear for all conditions. They conclude that there is durational compensation in Japanese, but little in Arabic. However, the lack of compensation in Arabic is not very clear from their results, and they test compensation only for a few consonants. Otake [1988, 1989a] carries out similar comparisons in Japanese, Arabic, Spanish, and Chinese. He concludes that there is evidence for compensation in all of these languages, and that such compensation therefore cannot be used to argue for mora-timing. However, not all of the consonants Otake tests show compensation. The results of both Port et al. [1980] and Otake [1988, 1989a] seem to show for all of these languages that there is an inverse relationship between vowel and consonant duration for only some of the consonants they test.

Minagawa-Kawai [1999] calculates correlations between duration of a consonant and several other durations, including preceding and following vowel, for all combinations of three consonants and two vowels. She finds strong negative correlations between duration of the consonant and the following vowel, and weak negative correlations between the consonant and the preceding vowel. This is true of Japanese, Chinese, and Korean, but the correlations are stronger and the slopes of the regression lines steeper for Japanese than for the other two languages. Sato [1993] examines the correlation between duration of a syllable-final nasal and the following voiceless or voiced stop, in order to determine whether the well-known tendency for vowels to be longer before voiced than voiceless stops is also true of nasals. She determines that there is a negative correlation between nasal and following stop (pooled across both voiceless and voiced stops) in English, Korean, and Japanese, but that it is strongest in Japanese.
Sagisaka and Tohkura [1984] use the full set of CV morae in a corpus, and observe a negative correlation between consonant and following vowel duration, and a weaker but significant negative correlation between consonant and preceding vowel duration in Japanese. Campbell and Sagisaka [1991] also find a negative correlation between durations of consonants and following vowels (as does Hiki [1967]), but Campbell and Sagisaka report a positive correlation between durations of consonants and preceding vowels.

Beckman [1982], however, observes a negative correlation only for some consonant/following vowel pairs in her data. Furthermore, negative correlations between adjacent segments may be unreliable evidence for compensation as a linguistic mechanism. Measurement error (either actual experimenter error in placing a boundary, or a discrepancy between the psychologically important boundary and the acoustically salient boundary on the spectrogram) will automatically lead to negative correlations [Beckman, 1982; Ohala, 1975], even if there is no compensation. Beckman [1982] also points out the universal nature of some apparent compensation effects: vowels are longer before voiced than voiceless obstruents in most languages, and this effect contributes to the negative correlation without implying mora-timing.

Han [1962a] addresses a special case of compensation by claiming that consonants are longer before a devoiced vowel than before a voiced one (the long-standing issue of whether such vowels are devoiced or entirely deleted is not relevant to this point, since devoiced vowels are certainly shorter than voiced ones.) Beckman [1982], however, finds that only half of her test words show even slightly longer consonant durations before voiceless vowels (although she does not provide a statistical test). Han [1994] compares /s/ before devoiced vowels to other segments which are inherently relatively long (e.g. /h, k/) before voiced vowels, and shows that /s/ before devoiced vowels is longer. However, she does not compare any segment before both devoiced and voiced vowels directly (although this would be possible in Japanese), and provides no statistical tests on this point. Since it is unclear what to count as part of the devoiced vowel and what as part of the consonant in a CV mora, any comparison of the consonant alone is questionable.

All of these authors, except Beckman and Otake, argue from these negative correlations that Japanese is mora-timed. Even in Beckman’s data, several CV pairs do have significant negative correlations. In addition to the problems about universality and measurement error, since it is accepted that compensation does not achieve perfect isochrony, it is unclear how strong or consistent a negative correlation must be to demonstrate compensation. Furthermore, if languages of other timing categories also use compensation within the units relevant to those languages, one would predict negative correlations in those languages, too. Port et al. [1980], Otake [1988, 1989a], Sato [1993], and Minagawa-Kawai [1999] demonstrate at least similar effects in other languages which are not expected to be mora-timed. Thus, clarification of the cross-linguistic predictions regarding compensation is necessary.

The Problem of Correlation Across Mora Boundaries

Much of the research on correlations in Japanese shows negative correlations not only between the durations of the consonant and vowel in a mora, but also between adjacent segments which are not in the same mora (consonants and their preceding vowels, or the mora nasal and its following consonant [Sato, 1993]). The tendency toward isochrony version of the mora-timing hypothesis clearly predicts negative correlations between segments within a mora [Han, 1962b]. Although it is not clearly stated in the literature, tendency toward isochrony would logically predict no correlation (either positive or negative) between durations of adjacent segments in different morae.

Taking an example, if some CV mora contains an inherently short consonant (e.g. /s/), the vowel in that mora should be made longer to compensate for it. But if this compensation is not sufficient, so that the entire mora is shorter than average, then presumably other morae in the word should be shortened in order to bring their duration closer to that of the short mora. There is no clear prediction of which segments would be shortened in order to achieve this, but if nearby segments in other morae are shortened, then the duration of these nearby segments would correlate positively with the shorter than average consonant, and negatively with the compensatorily lengthened vowel. That is, there would be a positive correlation between surrounding segments and the segment which diverges from average, and a negative correlation between surrounding segments and the other segment in the mora with a non-average segment. If, within CV morae, the vowel and the consonant are equally likely to be the segment diverging from average duration, one would expect there to be on the average no correlation at all between durations of segments across mora boundaries.

An alternative possibility for the tendency toward isochrony hypothesis is that compensation stops at the border of the mora: if compensation within a CV mora is insufficient to normalize mora duration, there might be no further compensation across mora boundaries. This would also lead to no duration correlation between consonants and their preceding vowels.

However, Campbell and Sagisaka [1991] find a positive correlation between consonants and their preceding vowels, and all of the other authors who examine correlations across mora boundaries show negative correlations. Although not all authors report significance of these correlations, all who do report statistical tests find significant correlations even across mora boundaries. Thus, the results regarding correlation of segment durations across mora boundaries (particularly between consonants and their preceding vowels) are not consistent with the hypothesis that speakers make mora duration more regular than it would inherently be.

Campbell [1999] further documents negative correlations across mora boundaries, particularly between vowels and following moras. He also finds that vowels are lengthened before the mora nasal. Using a linear regression analysis, he finds that syllable structure (with syllables categorized as CV, CVN, CyV, CVV, etc.) is a better predictor of syllable duration than number of segments in the syllable is. However, Campbell’s definition of the syllable is unusual, including geminate obstruents in the syllable following them rather than the one preceding them. This allows for QCVVYN (where QC is a geminate obstruent) as the most complex possible syllable. Japanese phonology makes it clear that geminate obstruents form the coda of the preceding syllable as well as the onset of the following one. CVQ functions as a heavy syllable for the pitch accent processes, and since syllable onsets do not contribute to syllable weight, this argues for Q being part of the preceding rather than the following syllable [Pierrehumbert and Beckman, 1988; Poser, 1984; Vance, 1987]. Furthermore, if the entire QC geminate obstruent sequence were the onset of the following syllable, as Campbell assumes, geminate obstruents should also be able to occur word initially (e.g. *Otaka should be possible as well as /kaQta/), but they do not. It is not clear what effect this non-standard definition has on Campbell’s results.
Campbell [1999] also determines that vowel duration cannot be predicted well without knowledge of the syllable type in which the vowel is located (even if information such as identity of the vowel and its preceding consonant, and presence or absence of mora nasal, is available). Based on these results, he argues that the syllable is a crucial unit in determining durations in Japanese, and that there is durational compensation within the syllable. (This does not mean that syllables are in any way equally timed, though: rather, syllable structure is the primary factor determining duration.) However, Campbell does not deny the status of Japanese as mora-timed, he simply adds a role for the syllable as well.

Mora-Timing at a Higher Level

A New Definition of Mora-Timing

The finding of negative correlations across mora boundaries leads to an important revision of the mora-timing hypothesis [Port et al., 1987], aside from Campbell's [1999] syllable model. Port et al. define mora-timing not as a tendency of all morae toward equal duration, but rather as predictability of word duration from the number of morae in the word (a partial return to Bloch's [1950] original formulation). Under this new definition, if one mora is shorter than average, the others are expected to be longer in order to maintain constant word duration, not shorter in order to match it, in a relationship Port et al. term 'anti-compensation'. This version of the hypothesis predicts negative correlations both within and across mora boundaries, which is consistent with most of the findings reviewed above. Homma [1981] and Port et al. [1980] also suggest that mora-timing might involve compensation at the word level, but do not make these predictions as clear.

Port et al. [1987] test this hypothesis by investigating the durations of words with various numbers of morae, and show a strong positive linear relationship between number of morae in the word and total word duration. They point out that the correlation of word duration with number of syllables would be far weaker. They find that segmental content of the word (both inherent duration differences and non-CV morae) has relatively little effect on total word duration. However, in one of their experiments, it appears that the words with both a geminate obstruent and a devoiced vowel are somewhat shorter than other words, particularly in fast speech. In another experiment, they find that CVCCV and CVVCV words are significantly shorter than CVCCV words, although they are much longer than CVCCV words.

Port et al. [1987] show negative correlations even between non-adjacent segments in a word, and conclude that speakers adjust the durations of many surrounding segments in order to normalize for inherent differences and keep the total word duration constant for a given number of morae. They plot word duration with the durations of each component segment to show that duration for words of the same number of morae

is nearly constant despite differences in the durations of individual segments. This method is widely adopted in the later literature, but it is not clear how similar total word durations must be in order to conclude that there is word level (or higher level) compensation.

Further Tests of High Level Mora-Timing

Bradlow et al. [1995] support the high level mora-timing hypothesis by replicating the effect of a strong positive linear relationship between whole word duration and number of morae in the word, and showing that total word duration is even constant under contrastive and broad focus. Sugito [1989] replicates the finding of a strong positive linear relationship between word duration and number of morae for Tokyo Japanese, and extends it to Osaka Japanese as well.

Sato [1996] compares Japanese, English, and Korean words with medial geminate and single nasals (e.g. /minā/ 'everyone' and /minā/ 'everyone (emphatic)' in Japanese, /sane/ 'inside the company' and /same/ 'inside the mountain' in Korean, and 'hoe maker' and 'home maker' in English). One should note that the phonological status of geminate nasals in English is different from the other two languages. She finds that in all three languages, the geminate nasal is longer than the single nasal, but that only in Japanese is the duration of the entire word with a geminate substantially longer than the duration of the word with a single nasal. Although the durations of three-mora words such as /minā/ are less than 1.5 times as long as the two-mora words (i.e. /minā/) even in Japanese, she argues for high level mora-timing based on the effect on word duration in Japanese and lack of one in English and Korean. Sato [1998] finds similar results for geminate obstruents and long vowels in the same three languages, although differences in duration between long and short vowels are less clear in Korean and English than differences between geminate and singleton consonants.

These findings by Bradlow et al. [1995], Sugito [1989], and Sato [1996, 1998] all support the claim of Port et al. [1987] that the mora is involved in Japanese timing at some higher level, such as the word. However, Otake [1998b] replicates several of the mora-timing effects of Port et al. [1980, 1987] even in English and Spanish, although these languages are syllable- and stress-timed, respectively. He uses words with only CV syllables in order to simplify the cross-linguistic comparison, so he tests only compensation for inherent variation, not for non-CV morae. He finds that the degree of variability in total word duration is similar in all three languages. He also shows that total word duration has the same strong positive linear correlation in all three languages. Otake concludes that supposed mora-timing effects are also present in stress- and syllable-timed languages. Otake's materials look more like Japanese (nonsense) words than English or Spanish nonsense words, though, and speakers may not have produced them with natural English or Spanish timing. Although using only CV syllables facilitates direct comparison among the languages, it may also give an unrepresentative picture of English and Spanish.

Warner and Arai [submitted] test the high level mora-timing hypothesis in spontaneous speech, and find a significant effect of non-CV mora types on word duration. They also show that the strength of the correlation between word duration and number of morae varies considerably among speakers, and that the correlation of word duration with number of segments (a non-prosodic unit) is nearly as good as with number of morae, although the correlation with number of syllables is weaker. They compare the correlation of whole word duration with number of morae to the same correlations for
arbitrarily truncated words (in which the compensation relationship should be disturbed), and show that correlations are stronger in the truncated words than the intact words. They conclude that Japanese does not use the mora as a unit of high level durational compensation.

Han [1994] investigates several effects at the segmental level which bear on the high level mora-timing hypothesis. She finds that segments are longer before geminate than singleton obstruents (though the difference is not significant in all tests). If geminate obstruents are too short to contribute a full mora’s duration, this is the expected result under the high level mora-timing hypothesis: a geminate makes the word somewhat shorter, so surrounding segments, such as the preceding one, are lengthened. (This might explain the discrepancy between Beckman’s [1982] and Homma’s [1981] and Han’s [1994] ratios for geminates, since Beckman did not include the preceding vowel in the calculation.)

However, Han [1994] also finds, with the same degree of statistical reliability, that the final segment of words containing a geminate obstruent is shorter than in words without a geminate, not longer. This supports the tendency toward isochrony hypothesis and contradicts high level mora-timing. These opposing results cannot both be considered as support for mora-timing. Campbell and Sagisaka [1991] and Campbell [1999] similarly show that the vowel before a geminate consonant is lengthened, and the vowel after it shortened. They also show that both the consonant and vowel before a moraic nasal are lengthened, while the consonant immediately after a moraic nasal is shortened. All of these effects are significant, and again, they support opposing versions of the mora-timing hypothesis.

Han [1994] also determines that the medial consonant is significantly longer in a VCV(C) word than in a CVCCV(C) word. In comparing VCV to CVCCV words, she follows up earlier work by Torii [cited in Hattori, 1980 (1961)], who found that the first vowel is longer in a VCV word than in a CVCCV word. Both results support the high level mora-timing hypothesis: a word ‘lacking’ the initial consonant is shorter than usual, so other segments, such as the medial consonant or first vowel, are lengthened. However, Han’s data still show a large difference between the duration of words with and without an onset consonant, indicating that any such compensation is weak.

The clearest difference between the predictions of the tendency toward isochrony hypothesis and the high level mora-timing hypothesis concerns the correlation between durations of whole morae, not segments. If morae tend toward isochrony within inherent limits, then there should be a positive correlation between the durations of adjacent morae (within whatever unit is used for such normalization, perhaps the word). However, the high level mora-timing hypothesis predicts negative correlation of adjacent morae. Few researchers have directly tested this difference in predictions. Arii and Greenberg [1997] plot durations of morae against the duration of the preceding mora (without reporting correlation coefficients, though). One might expect some noisiness in this relationship because they did not take word boundaries into account, but their graph reveals no relationship between current and preceding mora durations, either negative or positive.

Cross-Linguistic Comparisons

Looking for Japanese Effects in Stress- or Syllable-Timed Languages

In several studies discussed above, researchers establish a particular effect, claimed to show mora-timing, for Japanese, and also show that the same effect is not present or weaker in non-mora-timed languages such as English, Spanish, Korean, Chinese, or Arabic. Discussions of these results were presented above to facilitate comparison with results of other studies using the same method to test only Japanese. In sum, Otake [1988, 1989a, b] replicates effects related to high-level mora-timing and compensation in English and Spanish, Chinese, and Arabic. Port et al. [1980], Sato [1993] and Minagawa-Kawai [1999] find that relationships which are claimed as evidence for mora-timing in Japanese are weaker or restricted to certain environments in non-mora-timed languages.

Looking for English Effects in Japanese

Another approach to investigating mora-timing has been to test Japanese for some durational effect which is well known in English, and claimed to be related to stress-timing. The most common such effect to test has been syllable compression. It is well known that the average duration of a syllable in English is lower in words containing more syllables. This is often claimed as evidence of stress-timing: syllables are shorter in a stress foot with many syllables, because the duration of the stress foot must stay roughly constant. Neither version of the mora-timing hypothesis would predict syllables or morae in Japanese to be shorter in words containing more of them, since it is the mora which is the unit of normalization.

Hoequist [1983b] examines the average duration of a syllable in words with various numbers of syllables in English, Spanish, and Japanese. He finds no syllable compression effect for Japanese or Spanish, but his results for English are also unclear. Sato [1995] investigates average syllable duration and total word duration for English, Japanese, and Korean. (Her study uses exclusively CV syllables in all three languages, so that the syllable is identical to the mora in Japanese and Korean.) She finds that average syllable duration decreases with increasing number of syllables in the word strongly for English, very slightly for Japanese, and immediately for Korean. She concludes that Japanese is mora-timed and English stress-timed, with Korean having some features of both.

However, Beckman [1992] points out that cross-linguistic comparison of average syllable duration does not justify this conclusion. Stressed syllables in English are much longer than unstressed syllables, and in a word with two syllables, half the syllables are stressed, whereas a one-foot word with three syllables has one third stressed syllables. Therefore, average syllable duration in English would decrease in longer words simply because duration is one cue for stress in English, regardless of any compression to keep stress foot duration constant. Indeed, Sato shows greater average syllable duration in four-syllable than three-syllable English words, possibly because the four-syllable words may contain two stresses. Duration is not a cue for Japanese pitch accent (and Hoequist [1983a, b] finds little or no effect of pitch accent on duration in Japanese), so one would not expect such an effect on average syllable duration in Japanese. Thus, the ‘compression effect’, or the lack of it, can be used to show that English and Japanese timing are different, but not to argue for mora- or stress-timing in the sense of compensation. Hoequist [1983b] avoids this problem by calculating aver-
age syllable durations for English separately for stressed and unstressed syllables, but his results do not show clear compression. Beckman [1992] also points out that most studies of average syllable duration, or of any compression effect, include a confound with phrase-final lengthening.

Arai and Greenberg [1997] propose that if English is stress-timed, it should have something approximating an alternating pattern of longer and shorter syllables, whereas Japanese should not. They plot the duration of each syllable against the duration of the preceding syllable for corpora of naturally produced English and Japanese speech. They observe that neither language has such an alternating pattern, and that the degree of scatter in syllable duration is similar for the two languages. They emphasize universal constraints on production and perception and similarities between the two languages rather than language-specific timing mechanisms. However, the proposal of alternating longer and shorter syllables for English is a simplification, since many stress feet contain more than one unstressed syllable, and two stressed syllables can also follow each other. Furthermore, Arai and Greenberg do not test statistically whether the distribution of syllable durations in the two languages differs.

Looking for Mora-Timing Effects in Other Mora Languages

Nagano-Madsen [1992] takes a different approach, examining duration in languages where the mora seems to play the same role it does in Japanese. She argues that the status of the mora is phonologically similar in Japanese, Eskimo, and Yoruba. Since extensive durational measurements exist for Japanese, but not for the other two languages, she replicates past studies which have argued for mora-timing in Japanese in the other two languages, rather than presenting new measurements on Japanese. She focuses on three topics, (1) the ratio of long to short segment duration (both for consonants and vowels), (2) the relationship of total word duration to number of morae in the word, and (3) compensation between a consonant and its preceding and following vowels.

Nagano-Madsen [1992] finds that the ratio of geminate to singleton consonants and of long to short vowels is rather similar in Japanese and Eskimo. (The phonological structure of Yoruba makes it impossible to test such ratios in that language.) For consonants, Nagano-Madsen compares to Beckman’s [1982] results. For vowels, she compares to Han’s [1962a] findings. With regard to total word duration, Nagano-Madsen shows that the relationship between word duration and number of morae in Eskimo and Yoruba is very similar to the Japanese results of Port et al. [1987]. One Eskimo speaker’s results are almost identical to the Japanese results, and the other Eskimo speaker and the Yoruba speaker’s word durations increase with approximately the same slope, although the relationship is slightly less linear for them. However, Nagano-Madsen does not test statistically whether there is an effect of language on the linearity of the relationship. A common problem in studies which test the relationship of word duration to number of morae is that there is no definition of how linear the relationship must be to demonstrate mora-timing, and differences in degree of linearity among conditions are often not tested.

Nagano-Madsen’s test of compensation between a consonant and its surrounding vowels does not show similarity between Japanese and the other two languages. The relationship between consonant duration and duration of either the preceding or following vowel is not consistent in Eskimo or Yoruba. One Yoruba speaker does seem to show compensation, but only in words with a voiced stop as the consonant. Neither the voiceless stops from this speaker, nor any stops from the other Yoruba speaker or the Eskimo speakers, show compensation.

Nagano-Madsen points out that the domain and method of compensation for Japanese has been widely disputed. Since the other two tests show similarities between Japanese and the other two languages, she concludes that mora-timing is not a phenomenon specific to the phonology or phonetics of Japanese, but rather a general feature for languages in which phonetically short and long segments are better analyzed as a structural property (e.g. mora analysis) [Nagano-Madsen, 1992, p. 72]. Thus, her results do not strengthen or weaken the claim that Japanese is mora-timed, but instead clarify what phonological factors are related to mora-timing. This is related to the claim, discussed below, that the role of the mora in Japanese is structural rather than durational, although Nagano-Madsen accepts the mora as a durational unit.

Alternative Definitions of Mora-Timing

All the mora-timing hypotheses discussed above claim that speakers adjust the durations of segments or morae in order to normalize inherent variation. These hypotheses all involve compensation, whether between segments of a mora, across morae, or both. However, several researchers have suggested alternative definitions of what makes Japanese mora-timed, or of what the function of the mora in Japanese is. These proposals vary from a more explicit version of the tendency toward isochrony hypothesis to proposals which do not involve any type of isochrony.

Mora-Timing as Underlying Isochrony

Sagisaka [1991, 1999a] suggests that morae may have underlying isochrony in some sense, but that this isochrony is obscured by numerous other effects on duration. He considers the mora as an ‘internal clock’ or a basic timing unit for the planning of speech. Campbell and Sagisaka [1991] investigate strings of five consecutive CV-only morae in a read speech corpus. They test for effects of the identity of other segments on the duration of the consonant and vowel of the third mora. They find that the segments which most strongly affect the duration of the third mora vowel are the consonant of that mora and the consonant and vowel of the fourth (following) mora. The segments with the strongest effect on duration of the third mora consonant are the vowel and consonant of the fourth mora, and the vowel of the third mora.

Indeed, it appears from Campbell and Sagisaka’s [1991] results that every segment in the five-mora string has a significant effect on the duration of the segments of the third mora, although they do not emphasize the smaller effects. For example, the identity of even the first consonant in the five-mora string has a significant, though small, effect on the duration of the both the vowel and consonant of the third mora. Because Campbell and Sagisaka use a large corpus of speech, even very small effects are significant. However, they do not report the nature of the effects, except for the effects of adjacent segments. Thus, it is not clear whether the slight relationship between the identity of the initial consonant and the duration of the third mora vowel represents a negative or positive correlation, or whether the direction of the effect varies among consonants.

Kaiki and Sagisaka [1992] investigate the contributions of a wide variety of segmental and higher-level effects to segment duration, separating the contribution of each
The model that describes the relationship between the duration of vowels and consonants in Japanese, known as the Vowel Center of Gravity Model, proposes that the duration of vowels is determined by the position of the consonants preceding and following them. This model suggests that the duration of a vowel is influenced by the preceding and following consonants, with the duration being longer when the preceding consonant is voiceless and the following consonant is voiced. This model has been supported by a variety of experiments, including stop-syllable and the preceding consonant. The model also explains the phenomenon of vowel duration in Japanese, which is longer than in English, due to the influence of the preceding and following consonants. The model has been extended to include the effects of pitch and stress on vowel duration, providing a comprehensive explanation of the complex nature of vowel duration in Japanese.
from data on the motion of the lower lip and jaw during the reiterant speech, in order to clarify both universal constraints on motion and language specific factors. Using an articulatory dynamic model, they argue that the difference between mora-, syllable-, and stress-timing may lie in the overall setting of the stiffness parameter and in whether the stiffness or equilibrium parameters, or both, are involved in stress/accident distinctions. They find that pitch accent in Japanese does affect kinematic measures of jaw motion, even though it has a negligible effect on duration: high-tone vowels have less jaw lowering than low-tone vowels. This differs from their results for English, but is similar to French. Neither this proposal nor Smith's [1995] requires any type of durational compensation on the part of the speaker. Nor do these proposals imply that the durations of morae need have any particular relationship to each other.

Fujimura [1992, 1999] provides an alternative, articulatorily based view of the structure of speech and syllables through his C/D model. Although his model emphasizes articulation and is not primarily a model of acoustic duration, it may provide a useful alternative to linear ordering of segments or morae by dividing the consonants in a syllable into the onset, coda, and s-fix, within which linear order need not be specified. This could be useful in investigating cross-linguistic timing differences, although Fujimura does not make explicit predictions based on his model about how these timing differences are realized.

**Mora-Rhythm Rather Than Mora-Timing**

Extending Dauer's [1983, 1987] work on the difference between syllable- and stress-timed languages, several authors argue that what differs between Japanese and English, French, etc., is a more general rhythm, rather than timing specifically [Beckman, 1992; Cutler, 1991; Otake, 1990; Warner and Arai, submitted]. Dauer finds that stresses are timed with an equal degree of isochrony (or departure from it) in both stress- and syllable-timed languages, and suggests that structural factors, rather than isochrony, make languages of these categories different. She mentions variety of possible syllable structures, reduction of unstressed vowels, predictability of stress location, and number of syllables possible in a stress foot as some factors which influence whether a language is perceived as stress-timed or syllable-timed. Dauer does not address mora-timing, but several researchers have pointed out structural factors which could be relevant to the rhythm of Japanese. Otake [1990] calculates the percentage of syllables with various syllable structures in several types of Japanese texts, and shows that a very high proportion of syllables in Japanese texts consists of a single CV mora. The proportion of such syllables is even higher than in Spanish or French, both of which have a higher proportion of CV syllables than English, a factor which Dauer relates to the rhythm typology.

Beckman [1992] points out several structural aspects of Japanese which may contribute to cross-linguistic rhythm differences. In addition to the restricted syllable structure, she mentions that Japanese has no phonotactic minimality constraints which apply only to accented syllables, which would increase the difference in duration between accented and unaccented syllables. Japanese does not have allophonic rules conditioned by pitch accent, as English does for stress. Beckman further notes, on the issue of how many syllables can be in a single stress foot, that Japanese can have entire phrases without any accent. The lack of an effect of pitch accent on duration is also likely to keep morae more regular than they are in English, and Japanese also lacks the reduction of unstressed vowels which helps make stressed syllables more prominent in English. Warner and Arai [submitted] emphasize the large durational difference between phonemically long and short segments in Japanese. Long and short Japanese vowels differ greatly in duration but minimally in quality, unlike languages in which vowel 'length' distinctions are cued by both duration and quality differences (e.g. English). Thus, the very large differences between duration of one- and two-mora syllables may lead listeners to think that two-mora syllables take as long as two single morae, even in the absence of isochrony or compensation.

The primary evidence for a structural, rather than temporal, basis to mora-rhythm consists of problems with the evidence for compensatory mora-timing and the existence of structural (phonological) factors which would seem likely to make morae relatively regular in duration. Warner and Arai [submitted] present analyses which test the hypothesis of compensatory mora-timing against the hypothesis that each mora simply contributes variability to the variability of the word (accumulation of variance), and find evidence in favor of accumulation of variance. They relate accumulation of variance to the hypothesis that rhythm derives from structural factors rather than compensation. This approach provides positive evidence for mora-rhythm, rather than simply pointing out the lack of evidence for compensatory mora-timing.

Still, until recently there has been no acoustic measure which could distinguish among languages with mora-rhythm, syllable-rhythm, or stress-rhythm by reflecting the structural factors which are hypothesized to create those rhythms. Ramus et al. [1999] present such a measure. They calculate the proportion of the speech stream which is vocalic and the standard deviations of vocalic and consonantal interval durations, providing a measure of how regular the durations of vowels and of the intervals between vowels are. They find that vocalic proportion and standard deviation of consonantal intervals are quite effective at distinguishing stress-, syllable-, and mora-rhythm languages. Japanese has very low consonant variability and a high vocalic proportion, English, Dutch, and Polish (primarily stress-rhythm) have the opposite, and Spanish, Italian, French, and Catalan (primarily syllable-rhythm) fall in between.

Ramus et al. relate this finding to the structurally based rhythmic distinction. For example, Japanese has the lowest consonant variability because it has so few consonant clusters. They do not claim that speakers manipulate vocalic proportion or consonantal variability in order to reach some target, but rather that these measures are reflections of the phonological factors underlying the rhythm, and thus can provide a quantitative measure of rhythm category. However, the measures Ramus et al. devise probably do not reflect all of the phonological factors which contribute to rhythm, and additional research is necessary to determine the relationship between these phonological factors and the acoustics of rhythm. In a related study, Ramus and Mehler [1999] find that French listeners can distinguish Japanese from English even if the signal is resynthesized to leave only information about syllable structures. Thus, syllable structures seem to be quite important in determining listeners' perception of language rhythm.

In sum, there are several aspects of the structure of Japanese which may contribute to the perception of relatively regular morae. However, only some of these factors involve duration at all, and none of them require either isochrony of morae or a tendency toward it. The authors who emphasize the structural role of the mora in Japanese conclude that the mora-rhythm of Japanese is created by these structural factors rather than by any normalization of duration. None of the research on structural factors in mora-rhythm disavows the importance of the mora in Japanese. Although Beckman [1982] denies any role other than an orthographic and historical one for the mora, this
is not the position she takes in more recent work [Beckman, 1992; Pierrehuubert and Beckman, 1988]. Recent psycholinguistic studies of speech perception [Cutler et al., 1996; Cutler and Otake, 1994; Kakehi et al., 1996; Kubozeno, 1996; Morais et al., 1996; Otake et al., 1993, 1996], as well as analyses of Japanese pitch accent and phonology [Haraguchi, 1977, 1996; Ito and Mester, 1995; McCawley, 1968; Pierrehuubert and Beckman, 1988; Poser, 1984, 1990], confirm the importance of the mora as a unit of the language, regardless of any role it may have in timing.

Perceptual Mora-Timing

It is possible that even if morae are not regularly timed, listeners expect morae to be timed in a certain way, or use the mora as a unit in evaluating the timing of speech. Kato [1999] and Kato et al. [1997] extensively test listeners’ evaluation of Japanese words with normal duration and with durations of some segments altered, in order to determine what types of temporal alterations listeners are sensitive to, and thus improve the quality of synthetic speech. They also perform experiments on perception of nonspeech sequences matched to the temporal structure of the speech to distinguish psychoacoustic effects from language-specific ones.

The aspect of Kato’s work most relevant to mora-timing involves temporal modification of adjacent segments. Within four-mora Japanese words consisting entirely of CV morae, Kato modified the duration of a CV or VC string. In one condition, he modified (either lengthened or shortened) only the vocal, and in another only the consonant. In another condition, he modified the durations of both the consonant and vowel in the same direction (both shorter or both longer), and in a fourth condition he modified both C and V in opposite directions, so that the modifications compensated for each other.

Kato finds that modifications to the vocal alone, the consonant alone, or the vocal and consonant in opposite directions (compensation) cause approximately the same decrement in acceptability to listeners. However, modification of both the vocal and consonant in the same direction (both longer or both shorter) causes a significantly greater drop in acceptability. Moreover, there is no difference in these results for modification of CV versus VC sequences. Kato reasons that if listeners used the mora perceptually as a timing unit, modification of both C and V in opposite directions (compensating for the change) would be acceptable if the C and V were in the same mora, but unacceptable if they were in different morae. He concludes that the mora unit is not relevant in listeners’ evaluations of timing, and explains the results instead through the degree of amplitude jump across segment boundaries, a psychoacoustic effect.

Kato’s predictions relate to the tendency toward isochrony version of the mora-timing hypothesis, in which compensation is expected to take place primarily within the mora. Under the high-level mora-timing hypothesis [Port et al., 1987], one might expect modification in opposite directions to be acceptable regardless of mora boundary location, since compensation is expected across several segments in that theory. Modification in opposite directions would cause a smaller decrement in acceptability than modification to only the C or the V, since the double modification maintains the correct total word duration. Kato’s results show that modification to both C and V in opposite directions is at least as damaging to acceptability as modification of either segment alone (and slightly but significantly more damaging than modification to just the consonant). Thus, the compensation between the consonant and vowel does not improve acceptability, and provides evidence against perceptual use of mora-timing, regardless of the domain of compensation.

Acquisition of Timing Patterns

Using a different approach to the issue of the timing typology, several authors have investigated Japanese native speakers’ acquisition of English timing or English native speakers’ acquisition of Japanese timing. Mochizuki-Sudo and Kiritani’s [1991] study in this area is particularly interesting in that it incorporates both production and perception. They test native English speakers, native Japanese speakers proficient in English, and native Japanese speakers with limited English on production of English sentences in which the number of unstressed syllables within a stress foot is varied. They measure stress foot duration and stressed vowel duration, and find that stress foot duration increases approximately linearly with increasing number of syllables for all three groups of speakers, but it increases the most for the nonproficient speakers and the least for the native English speakers (a slope difference). Furthermore, for native and proficient nonnative English speakers, the stressed vowel is shorter in stress feet with two unstressed syllables than in those with just one unstressed syllable. (The decrease does not continue in longer stress feet, though.) For nonproficient speakers, there is no such decrease. Mochizuki-Sudo and Kiritani conclude that the nonproficient speakers have not learned English foot level compression (shown by the duration of the stressed vowel), while the proficient nonnative speakers have. They interpret this as supporting the stress-timed versus mora-timed division.

In several perception experiments, these authors investigate the sensitivity of native English speakers and native Japanese speakers (not proficient in English) to modification of durations in English words, particularly modification of stressed vowel duration. They find that English listeners are more likely to accept shortening of the stressed vowel than Japanese listeners are, perhaps because of the Japanese vowel length distinction. They also present words for naturalness judgement with either just the stressed vowel lengthened, or the stressed vowel lengthened and the vowel of the following unstressed syllable compensatorily shortened. They determine that for native English listeners, modification of the stressed vowel alone decreases naturalness more than compensatory modification of the two vowels. However, for Japanese listeners, both types of modification cause an equal decrement in naturalness. They conclude that English listeners use the stress foot as a unit of timing in perception, while Japanese listeners do not.

Ueyama [1996] also investigates Japanese speakers’ production of English timing patterns. She tests for shortening of both stressed and unstressed syllables with increased number of syllables in the stress foot, and specifically separates the influences of phrase-final lengthening, prosodic boundary strength, and compression within the stress foot, in order to avoid the confound pointed out by Beckman [1992]. She observes differences between native English speakers’ and non-native speakers’ production of phrase-final lengthening, but she fails to find a consistent effect of compression with increased number of syllables in the stress foot, even for native speakers of English. Therefore, she does not draw any conclusions about Japanese speakers’ acquisition of stress-timing.

Ueyama [1999] focuses on the duration of unstressed syllables. She compares production of unstressed syllables within a larger content word to unstressed syllables which are the only syllable of a function word. She also tests stress feet with varying numbers of unstressed syllables. She shows that the status of an unstressed syllable (function word or within a longer word) has no effect for English speakers, but that
investigates Japanese children’s acquisition of the distinction between the mora nasal followed by another nasal vs. a single nasal (e.g. /haNna/ vs. /hana/). She finds that Japanese children do not begin to produce the durational difference in an adult-like way until they are 5–6 years old. Nor can children perceive the distinction accurately until at approximately age 6. This study provides interesting information about children’s acquisition of distinctions which are cued by duration, but it does not address the question of whether the mora is used in regulating timing, or how children acquire that feature of the language if it is.

**Conclusions**

Although the description of Japanese as mora-timed has been pervasive, the evidence for mora-timing in the sense of speakers compensating in order to make some interval more regular than it would inherently be has been inconsistent. There are serious methodological problems regarding the interpretation of many findings which have been used to argue for mora-timing. This has been true both of the original tendency toward isochrony hypothesis, and of the high level mora-timing hypothesis. The most serious of the problems discussed above involve duration ratios, negative correlations between adjacent segments, and positive correlations between whole word duration and number of morae. The calculation of duration ratios for long and short segments in the early literature [Han, 1962b, among others] is problematic because of assumptions which are necessary in order to determine whether a non-CV mora contributes the duration of an entire mora. This problem was later solved by calculating ratios of larger strings (e.g. CVV:CV rather than VV:V).

However, there can be no solution for the problem with negative correlations between adjacent segments: such findings can never be good evidence of temporal compensation, because these negative correlations will also appear automatically if there is any measurement error in placing boundaries (even if this is only because the clear acoustic boundary is not the psychologically important one, and not because of actual experimenter error). Negative correlations between adjacent segments have been and continue to be common as evidence for mora-timing [e.g. Minagawa-Kawai, 1999; Port et al., 1980; Sagisaka and Tokhara, 1984].

In the case of positive correlations between whole word duration and number of morae in the word, which have been very widely used to argue for mora-timing in recent years [Bradlow et al., 1995; Kondo, 1999; Port et al., 1987; Sato, 1995; Sugito, 1989], the problem is that it is not clear how strong such correlations must be in order to provide evidence for mora-timing. As Warner and Arai [submitted] point out, even if there is no compensation, a strong correlation would be expected, since all morae do have segmental content. Furthermore, within the studies using these positive correlations, there is no consistent criterion for how linear the increase in word duration with increasing number of morae must be, so that similarly variable data can be counted as evidence for or against mora-timing in different studies. (For example, compare the non-native Japanese speakers of Kondo’s [1999] study to the fast speech condition of the study by Port et al. [1987].) Furthermore, in some cases, an effect presented as evidence for mora-timing does not show isochrony or compensation, but simply that duration is not affected by Japanese pitch accent (e.g. Sato’s [1995] work on compression in words with many syllables). However, the research on the subject of Japanese mora-