The Scope and Effects of Preboundary Prosodic Lengthening in Japanese[*]

Michael A. Shepherd
University of Southern California
mshepher@usc.edu

The acoustic duration of phonemically short and long Japanese vowels with three degrees of removal from an intonation phrase boundary--immediately preboundary, one mora preboundary, and two moras preboundary--was measured. Immediately preboundary vowels showed significant acoustic lengthening relative to like controls, with short vowels showing proportionally greater lengthening than long vowels. However, vowels even one mora preboundary were generally not significantly different from controls. Interpreted in light of the π-gesture model of prosodic influence (Byrd and Saltzman 2003; Byrd et al. 2006), this suggests that the temporal scope of preboundary prosodic lengthening in Japanese is approximately one mora and, thus, only the latter portion of a preboundary long vowel overlaps the π-gesture. This difference in the degree of lengthening of short versus long vowels also significantly reduced the long-to-short ratio of immediately preboundary vowels relative to controls. Moreover, while there was a vowel-to-word duration ratio that reliably classified control vowels as either short or long, applying the same ratio to immediately preboundary vowels classified all of them as long. These findings challenge the claim that relative acoustic duration in Japanese is invariant across speaking rates (Hirata 2004).

1. Introduction

1.1. Phrase-final lengthening

It has long been known that segments in close proximity to a prosodic boundary, such as an intonation phrase boundary, tend to be longer in duration than comparable segments that are not near a boundary (e.g. Oller, 1973; Klatt, 1976; Wightman *et al*. 1992). Phrase-final lengthening, as it is often called, has been observed in a variety of languages including Japanese (e.g. Takeda *et al*., 1989).

1.2. Duration and timing in Japanese

Japanese has a two-way length contrast for all five of its vowels—/i/, /e/, /a/, /o/, and /ɯ/—resulting in a ten-vowel inventory (Vance 1987). This length contrast will allow us to examine the effect of phrase-final lengthening on phonemically short vowels vis-à-vis phonemically long vowels.[1]

Japanese has long been described as a mora-timed language, meaning that speech consists of a series of units of roughly equal duration (e.g. Bloch, 1942). Port *et al*. (1987) and Han (1994) have presented experimental evidence supporting the durational uniformity of moras, although more recent work by Warner and Arai (2001) suggests that this may only be true of careful speech. Possible moras are a CV, a V (either on its own word-initially or as the second V in a CVV syllable), or a C (either as the first half of an NC cluster or a geminate consonant).

---

[*] The author gratefully acknowledges the support of the NIH (grant DC03172) and thanks Professor Dani Byrd for her helpful comments.

[1] Although the present study will only look at vowels, note that duration is also contrastive for the voiceless stops, the voiceless alveolar fricative, the voiceless affricate, and the nasals.

USC Working Papers in Linguistics **4**: 1-14

1.3. The π-gesture framework

In the prosodic gesture (π-gesture) framework of Byrd and Saltzman (Byrd and Saltzman, 2003; Byrd *et al.*, 2000; Byrd 2000; Byrd *et al.*, 2006), prosodic-boundary effects such as phrase-final lengthening are seen as resulting from localized decreases in speaking rate. The degree of slowing depends on the level of activation of the π-gesture, which in turn depends on the strength of the prosodic boundary. As a phrase boundary approaches, the π-gesture steadily increases in activation, causing greater and greater slowing of articulatory gestures that are active at the time, until reaching its maximum level of activation at the boundary itself. Activation then fades after the boundary has passed, such that local speaking rate and, as a result, gesture durations return to normal.

The π-gesture model makes a number of predictions concerning phrase-final lengthening. First, it predicts that any and all articulatory gestures that fall within the temporal scope of the π-gesture will undergo lengthening. Second, it predicts that the magnitude of lengthening will be greatest at the π-gesture's maximum level of activation, i.e., adjacent to the phrase boundary. Third, the limited temporal scope of the π-gesture predicts that lengthening will be limited to gestures relatively close to the boundary (Byrd and Saltzman, 2003).

1.4. Goals of the present study

Our primary goal is to investigate phrase-boundary lengthening of phonemically short and long Japanese vowels in light of the predictions of the π-gesture model of prosodic influence. In particular, we are interested in the possibility that, at a given distance before a phrase boundary, the temporal scope of the π-gesture might be sufficient to overlap a short vowel completely while only managing to overlap the latter portion of a long vowel in the same position relative to a phrase boundary. In this situation, we would expect the short vowel to undergo a proportionally greater degree of lengthening than the long vowel.

To make this prediction more concrete, consider the following schematized illustration. Let us imagine a short vowel that, at a given speaking rate, has a duration of 50ms and a long vowel that has a duration of 100ms at the same speaking rate. Let us further imagine a π-gesture that, at this same speaking rate, extends back 70ms from a given phrase boundary and doubles the duration of any articulatory gesture(s) it overlaps. If our short vowel were placed immediately before the boundary, it would fall completely within the temporal scope of the π-gesture, so its duration would be doubled to 100ms. On the other hand, if our long vowel were placed in the same position relative to the boundary, only its last 70ms would fall within the scope of the π-gesture. This overlapping portion would be doubled in duration (to 140ms), but its first 30ms would be unaffected.[2] The result would be a total duration of 30+140 = 170ms. Figures 1 through 3 illustrate this.

FIGURE 1 – The π-gesture

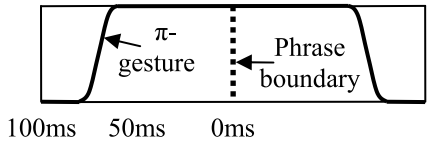
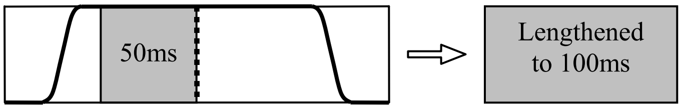


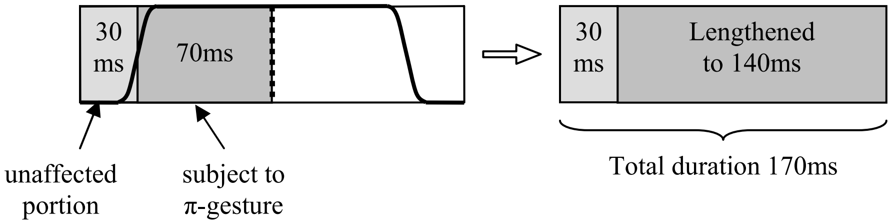
FIGURE 2 – Effect on a 50ms vowel



---

[2] We could potentially distinguish between the portion of the long vowel, if any, that does not overlap the π-gesture at all and the portion that overlaps the onset of the π-gesture, during which activation is gradually increasing. If the long vowel overlapped either of these portions, it would undergo proportionally less lengthening than the short vowel, giving rise to the predicted result.

FIGURE 3 – Effect on a 100ms vowel



Thus, long vowels might be found to undergo a greater degree of phrase-boundary lengthening than short vowels in absolute terms (a 70ms gain by the long vowel versus a 50ms gain by the short vowel in the above illustration). Crucially, however, we make the general prediction that short vowels will undergo a greater degree of lengthening in proportional terms (a 100% increase for the short vowel versus a 70% increase for the long vowel in above illustration).

If this prediction is borne out, it will also affect the relative duration ratio of long vowels to short vowels. Hirata (2004) has presented evidence that the relative duration of short and long vowels in Japanese remains constant across speaking rates. This is true despite large changes in absolute duration such that phonemically short vowels produced at a slow speaking rate are often longer than phonemically long vowels produced at a fast speaking rate. Similar findings have been reported for the short and long vowels of Arabic (Port *et al*., 1980), Icelandic (Pind, 1999), and Thai (Svastikula, 1986). Hirata (2004) cites relative duration as an invariant cue to phonemic length category in Japanese. In the above illustration, however, the long-to-short ratio changes from 2:1 when the vowels are not adjacent to a phrase boundary to 1.7:1 when they are adjacent to a boundary. Thus, if our prediction is borne out, it will present a challenge to Hirata's claim.

Another aspect of relational acoustic invariance that Hirata (2004) discusses is the duration ratio of a vowel to the word in which it occurs. This is shown to classify a vowel as phonemically short or long with 94.3% accuracy provided the number of moras in the rest of the word is known. Phrase boundary lengthening is likely to alter this ratio as well, however, since boundary-adjacent gestures have been shown to be more strongly affected by lengthening than those removed from the boundary (e.g. Byrd *et al*. 2006). Thus, if a boundary-adjacent short vowel underwent a sufficient degree of prosodic lengthening, its vowel-to-word duration ratio could come to resemble that of a non-boundary-adjacent long vowel. Note that this possibility is independent of whether short and long vowels undergo proportionally different degrees of final lengthening.

Hypothetical scenarios aside, Berkovits (1993) details a very real case of how final lengthening can affect the relative duration of segments. The duration ratio of a non-sentence-final coda fricative to the immediately preceding vowel in Hebrew is approximately 0.5:1 for voiced fricatives and 1:1 for voiceless fricatives. These differences in the consonant-to-vowel duration ratios of voiced and voiceless coda fricatives result from a combination of the tendency of voiced obstruents to be shorter than voiceless ones and the 'voicing effect,' i.e., the fact that vowels preceding voiced coda consonants tend to be longer in duration than vowels preceding voiceless ones.

What Berkovits (1993) found is that these ratios change in sentence-final position to 1:1 for voiced fricatives and 2:1 for voiceless fricatives due to proportionally greater lengthening of the coda fricatives relative to the preceding vowels. In the π-gesture framework, this is expected since the coda fricatives in question were directly adjacent to the phrase boundary whereas the preceding vowels were one segment away. As a result, the fricatives would have overlapped the π-gesture at its highest level of activation, receiving its maximal slowing/lengthening effect. The pre-fricative vowels, on the other hand, being further back from the boundary, would have been subject to a less highly activated portion of the π-gesture. Moreover, depending on the temporal scope of the π-gesture, the vowels may have only partially overlapped it. These pre-fricative vowels would, therefore, have been slowed/lengthened to a relatively lesser extent than the fricatives that followed them. The net result of maximal π-gesture lengthening of the coda fricatives combined with a lesser degree of lengthening of the preceding vowels would be exactly what Berkovits found, namely, a marked increase in the duration of the coda consonants relative to the preceding vowels at the phrase boundary.

Although no perceptual study was done, the importance of this ratio as a cue to voicing in English (e.g. Denes, 1955; Port and Dalby, 1982) led Berkovits (1993) to suggest that listeners probably need to take sentence position as well as the consonant-to-vowel duration ratio into account in determining whether a coda consonant is voiced.

Our experiment will allow us to determine whether phrase-final lengthening in Japanese can have a similar effect on the duration ratio of long vowels to short vowels and/or the duration ratio of vowels to the words they occur in. We will measure short and long vowels with three degrees of removal from an intonation phrase boundary: immediately preboundary, one mora preboundary, and two moras preboundary. This is necessary given our inability to know, a priori, exactly how far back from the phrase boundary the π-gesture will extend in Japanese.

2. Method

2.1. Stimuli and subjects

Eight sentences were used to test boundary effects on preboundary short and long vowels. The target vowels were the final (short) vowel in *dōkyo* 'roommate' and the final (long) vowel in minimally different *dōkyō* 'townmate.' Note that, in addition to forming a segmental minimal pair and being similar enough semantically to be placed in the same sentential contexts, *dōkyo* and *dōkyō* have the same pitch accent pattern (both are unaccented). Note also that, since the sentences were written in conventional Japanese orthography, *dōkyo* and *dōkyō* were written in Japanese kanji as 同居 and 同郷, respectively. The pre-target context was identical in all eight sentences, and the following contexts were matched as closely as possible in number and type of segments. The eight stimulus sentences and their English translations are given in Table 1.

The recordings were evaluated using the J_ToBI guidelines for Japanese intonation (Venditti, 2005). All sentences were realized with the expected intonation. The sentences representing the immediately preboundary condition were realized with an intonation phrase boundary (marked by a boundary tone, a break index 3, and resetting of the pitch range) after the target word (*dōkyo* or *dōkyō*). The sentences representing the one mora preboundary condition were realized with an intonation phrase boundary (marked by a boundary tone, a break index 3, and resetting of the pitch range) after the /-da/ that immediately followed the target word. The sentences representing the two moras preboundary condition were realized with an intonation phrase boundary (marked by a boundary tone, a break index 3, and resetting of the pitch range) after the /-desu/ that immediately followed the target word. The control sentences were realized as a single intonation phrase, i.e. with no medial intonation phrase boundaries.

TABLE 1 – Stimuli

#	Vowel	Prosodic Condition	Sentence and translation
1a	short	immediately preboundary	Watashitachi-wa dōky**o**. Desukara nakama desu. 'We are roommates. Therefore we are friends.'
1b	long	immediately preboundary	Watashitachi-wa dōky**ō**. Desukara nakama desu. 'We are townmates. Therefore we are friends.'
2a	short	one mora preboundary	Watashitachi-wa dōky**o**-da. Sugoi nakama-da. 'We are roommates. We are good friends.'
2b	long	one mora preboundary	Watashitachi-wa dōky**ō**-da. Sugoi nakama-da. 'We are townmates. We are good friends.'
3a	short	two moras preboundary	Watashitachi-wa dōky**o** desu. Kore-wa kazoku desu. 'We are roommates. This is family.'
3b	long	two moras preboundary	Watashitachi-wa dōky**ō** desu. Kore-wa kazoku desu. 'We are townmates. This is family.'
4a	short	control (no boundary)	Watashitachi-wa dōky**o** desukara nakama desu. 'We are roommates and friends.'
4b	long	control (no boundary)	Watashitachi-wa dōky**ō** desukara nakama desu. 'We are townmates and friends.'

Four subjects, all of them native speakers of the Tokyo dialect of Japanese, participated in the experiment. Subjects will be referred to as Subject J, Subject K, Subject M, and Subject N. The stimuli were pseudo-randomized in 17 lists of 10 sentences, such that adjacent items were never (1) identical, (2) the paired short-vowel and long-vowel sentences in a single prosodic condition, or (3) a sentence from the immediately preboundary condition and one of the controls.[3] The first and last of the 17 lists, as well as the first and last sentence on each list, were included as fillers with the goal of avoiding a possible list intonation effect in the test sentences. All eight stimulus sentences appeared once in the second through ninth positions of every list. Moreover, two of the eight stimulus sentences appeared a second time in the first and last positions. These additional sentences were chosen pseudorandomly following the same three criteria listed above. All four subjects read each of the 17 randomization lists once in the same order. All were instructed (in English) to read in a casual, conversational style, "as if talking with a close friend."

2.2. Data collection

Subjects were recorded in a quiet room using an iMac G4 with an external microphone. Subject N was recorded using Macquirer. Subjects J, K, and M were recorded using Praat. All were recorded at a sampling rate of 44.1 kHz.

2.3. Data analysis

The duration of the target vowel was defined as the time from the onset of voicing following the release of the immediately preceding /k/ to the amplitude drop accompanying the stop closure for the following /d/ (or the offset of voicing in the immediately preboundary condition). The duration of the word containing the target vowel was defined as the time from the amplitude drop accompanying the stop closure for the word-initial /d/ to the end of the target vowel. These measurements were made from waveforms and spectrograms using Audacity and Wavesurfer, respectively.

17 of Subject N's 120 stimulus sentences were lost due to data collection error.[4] Also, 15 of Subject J's stimulus sentences, as well as nine of Subject M's were deemed unmeasurable due to lenition of the stop closure following the target vowel.[5] Lastly, in four of Subject K's stimulus sentences and one of Subject M's, it was determined that the speaker had mistakenly replaced a short vowel target word with a long vowel target word or vice versa.[6] Such tokens stood out initially because their duration was roughly double (or half) the particular subject's mean for that condition. The suspicion of a word misreading was independently confirmed by a native Japanese speaker who did not participate in the experiment. All excluded and unavailable tokens (a total of 46 out of 480) were replaced with filler tokens of the same sentences by the same speakers that had been recorded during the same session. Specifically, replacements were preferentially list-initial tokens from the 15 test lists (the second through sixteenth). When additional replacements were needed, non-final tokens were taken from the first and last filler lists (a balanced number from each). List-final tokens were used only when no other replacement was available. It should be noted, however, that such tokens showed no evidence of having been read with a list intonation and were avoided only as a precaution.

Individual two-factor ANOVAs were conducted for each subject testing the effect of prosodic condition and phonemic length category (short or long) on vowel duration. Six difference of means tests (planned comparisons) were then run for each subject. These compared duration in the control condition of each phonemic length category with duration in the corresponding immediately preboundary, one mora preboundary, and two moras preboundary conditions, respectively, in order to determine whether prosodic lengthening occurred. Next, in order to determine whether short vowels underwent proportionally greater

---

[3] This last criterion was added because the immediately preboundary sentences were segmentally identical to the control sentences, differing only in the presence or absence of a phrase boundary.

[4] These consisted of three tokens of 1a, three of 2a, one of 2b, two of 3a, three of 3b, two of 4a, and three of 4b.

[5] For Subject J, these consisted of three tokens of 2a, two of 2b, four of 3a, four of 3b, one of 4a, and one of 4b. For Subject M, these consisted of three of 2a, two of 2b, one of 3a, two of 4a, and one of 4b.

[6] For Subject K, these were of one token of 2a, one of 2b, and two of 4a. For Subject M it was one token of 4a.

lengthening than long vowels, unpaired t-tests were run on the duration ratios of vowels in the immediately preboundary condition to corresponding controls.

In order to evaluate Hirata's (2004) claim regarding relational invariance, a repeated measures ANOVA was run on the mean duration ratios of the phonemically long vowels to phonemically short vowels pooled for all four subjects. This was followed by a difference of means test comparing the long to short ratio in the immediately preboundary condition with the corresponding control ratio. Criterial significance was set at $p < 0.05$. All and only statistically significant results are reported.

3. Results

3.1. Preboundary prosodic lengthening

The mean vowel durations and standard deviations for each of the four subjects are presented graphically in Figures 4 through 7, respectively.

Figure 4 – Subject J mean vowel durations (error bars indicate SD)

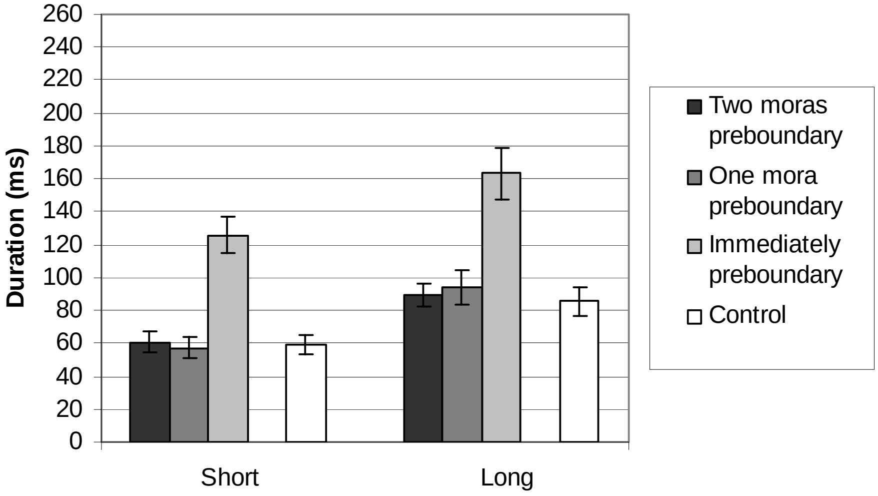


Figure 5 – Subject K mean vowel durations (error bars indicate SD)

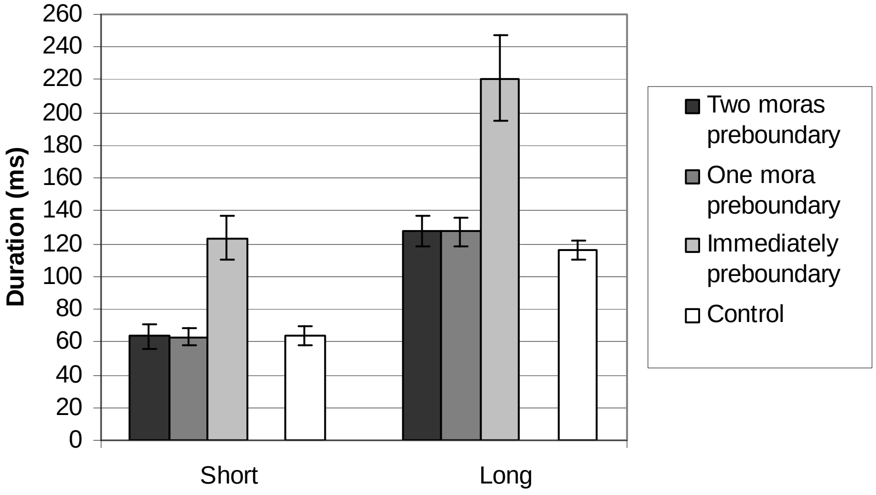


Figure 6 – Subject M mean vowel durations (error bars indicate SD)

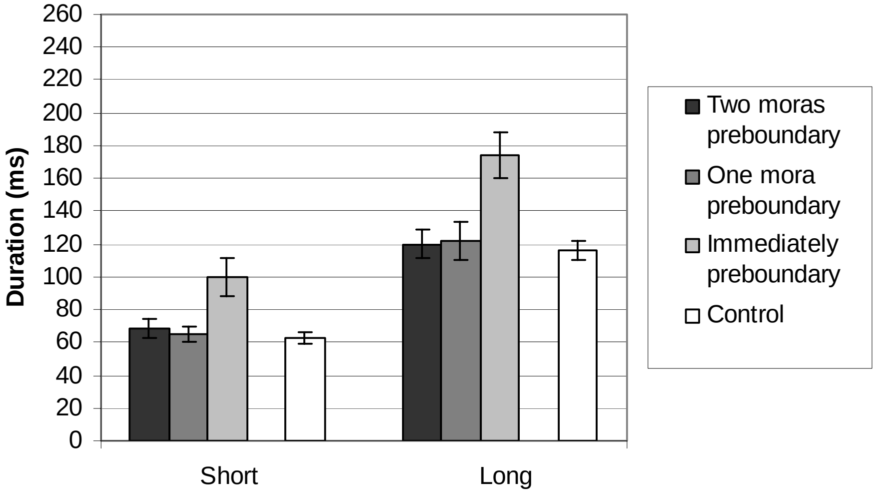
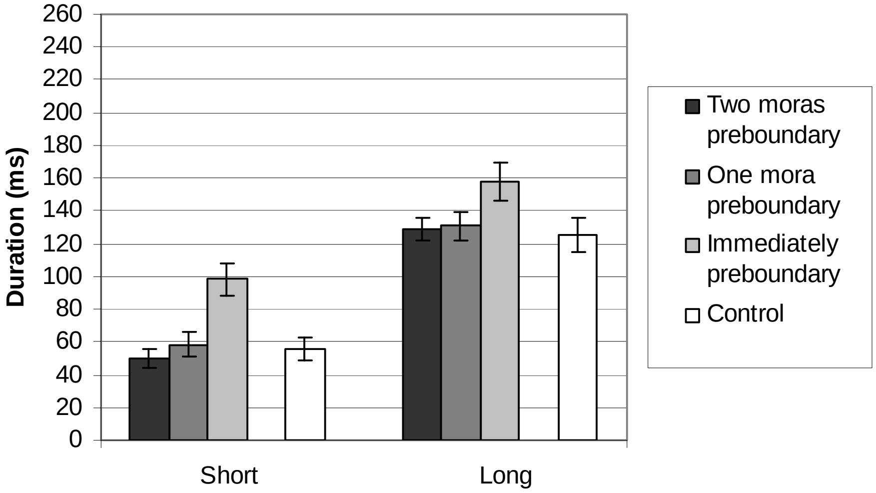


Figure 7 – Subject N mean vowel durations (error bars indicate SD)



Most striking, for both short and long vowels, is the difference between the immediately preboundary and control conditions. The means durations and standard deviations for these two conditions, as well as the differences and ratios between them, are given in Table 2. The mean durations and standard deviations for all conditions are given in Table 8 in the Appendix.

TABLE 2 – Means and standard deviations for the immediately preboundary and control conditions, and differences and ratios between them

	Immediately preboundary mean (SD) ms	Control mean (SD) ms	Difference (ms)	Ratio immediately preboundary:control
Subject J				
Short	125.8 (11.3)	59.4 (5.5)	66.5	2.1:1 (.04)
Long	163.2 (15.4)	85.7 (8.6)	77.5	1.9:1 (.03)
Subject K				
Short	123.5 (13.7)	64.0 (5.4)	59.5	1.9:1 (.05)
Long	220.7 (26.1)	136.3 (11.0)	84.4	1.6:1 (.04)
Subject M				
Short	99.6 (11.7)	62.9 (3.7)	36.8	1.6:1 (.03)
Long	174.1 (13.8)	115.6 (5.9)	58.4	1.5:1 (.01)
Subject N				
Short	98.3 (10.0)	55.7 (7.0)	43.2	1.8:1 (.03)
Long	158.1 (11.9)	125.6 (10.7)	32.5	1.3:1 (.01)

Consistent with our predictions, short vowels underwent a proportionally greater degree of lengthening than long vowels for all four subjects, as evidenced by the ratios in rightmost column. Thus for Subject J, for instance, the 2.1:1 ratio of immediately preboundary short vowels to control short vowels indicates that immediately preboundary short vowels had 2.1 times the duration of control short vowels. Compare this with Subject J's 1.9:1 ratio of immediately preboundary long vowels to control long vowels, which indicates that immediately preboundary long vowels had 1.9 times the duration of control long vowels. Note that this is true despite the fact that long vowels lengthened more than short vowels in absolute terms (except for Subject N), as shown by the difference values in the second column from the right.[7]

Individual two-factor ANOVAs (prosodic condition × phonemic length category) were run for each subject. These showed a significant main effect of prosodic condition on vowel duration for all four subjects.[8] In order to then determine which of the three test conditions in each phonemic length category were driving these main effects, six planned difference of means tests were run for each subject. These compared duration in the control condition of each phonemic length category with duration in the corresponding immediately preboundary, one mora preboundary, and two moras preboundary conditions, respectively. The tests revealed a significant duration difference between the control and immediately preboundary conditions for both short and long vowels for all four subjects. Additionally, for Subject J, there was a significant difference between the control and one mora preboundary conditions for the long vowels. The exact values are given in Table 3.

---

[7] It could be argued that this is due to a ceiling effect on vowel duration, that is, that long vowels undergo proportionally less lengthening than short vowels simply because lengthening beyond a certain duration is impossible for independent reasons. This seems unlikely, however, given that the longest vowel we measured was 264ms (an immediately preboundary long vowel produced by Subject K) while Hirata (2004) measured unstressed vowels well over 400ms in duration and stressed vowels over 500ms.

[8] There was also a significant main effect of phonemic length category on vowel duration ($p < .0001$ for all subjects). This indicates, unsurprisingly, that phonemically long vowels were longer than phonemically short vowels.

Table 3 – ANOVA results (part 1) and difference of means results

Subject	Main effect of prosodic condition	Difference of means
J	$F(3,112) = 418.100$ $p < .0001$	Short control vs. short immediately preboundary $F = 374.381, p < .0001$ Long control vs. long immediately preboundary $F = 508.851, p < .0001$ Long control vs. long one mora preboundary $F = 5.604, p < .02$
K	$F(3,112) = 266.994$ $p < .0001$	Short control vs. short immediately preboundary $F = 166.635, p < .0001$ Long control vs. long immediately preboundary $F = 335.997, p < .0001$
M	$F(3,112) = 182.777$ $p < .0001$	Short control vs. short immediately preboundary $F = 123.693, p < .0001$ Long control vs. long immediately preboundary $F = 312.405, p < .0001$
N	$F(3,112) = 135.113$ $p < .0001$	Short control vs. short immediately preboundary $F = 182.689, p < .0001$ Long control vs. long immediately preboundary $F = 103.212, p < .0001$

These results indicate that immediately preboundary vowels (as well as one mora preboundary long vowels for Subject J) underwent significant lengthening relative to controls. Moreover, the fact that preboundary prosodic lengthening was (largely) limited to the immediately preboundary condition suggests that the π-gesture associated with intonation phrase boundaries in Japanese has a preboundary temporal scope of approximately one mora. A π-gesture with this preboundary scope would be sufficient to overlap an entire immediately preboundary short vowel (and preceding consonant) but would only overlap the latter portion of an immediately preboundary long vowel. As was illustrated in the introduction, the π-gesture model (Byrd and Saltzman, 2003; Byrd *et al.*, 2006) predicts that this differential degree of overlap should result in a greater proportional degree of lengthening of immediately preboundary short vowels than immediately preboundary long vowels. The ratios presented in Table 2 show that this was indeed the case for all four subjects. Moreover, there was a significant interaction between prosodic context and phonemic length category for Subjects K, M, and N, which suggests that degree of final lengthening across prosodic conditions varied as a function of phonemic length category.[9] The exact values are given in Table 4.

Table 4 – ANOVA results (part 2)

Subject	Interaction between prosodic context and phonemic length category
K	$F(3,112) = 11.383, p < .0001$
M	$F(3,112) = 10.302, p < .0001$
N	$F(3,112) = 6.850, p < .0005$

What these interactions cannot tell us, however, is whether the proportionally greater degree of lengthening of immediately preboundary short vowels relative to immediately preboundary long vowels was significant. In order to answer this question, we first derived the duration ratios of each of a given subject's 15 immediately preboundary token to the mean of all 15 of its corresponding control tokens.[10] This was done

---

[9] The interaction for Subject J narrowly failed to reach significance (F(3,112) = 2.579, p = .0572).

[10] The immediately preboundary condition was chosen for comparison because it is the only condition that consistently exhibited significant phrasal lengthening.

by individually dividing each subject's 15 immediately preboundary short vowel tokens by the mean of that subject's 15 short control tokens and, likewise, each subject's 15 immediately preboundary long vowel tokens by the mean of that subject's 15 long control tokens. Individual unpaired t-tests were then run comparing the resulting ratios from each subject's short vowels with the ratios from the same subject's long vowels. The differences between the immediately preboundary-to-control ratios for phonemically short versus phonemically long vowels were significant for Subjects J, K, and N.[11] These results show that, consistent with our predictions, the proportional degree of lengthening of phonemically short vowels was significantly greater than that of phonemically long vowels for these three subjects. The t-test results are given in Table 5.

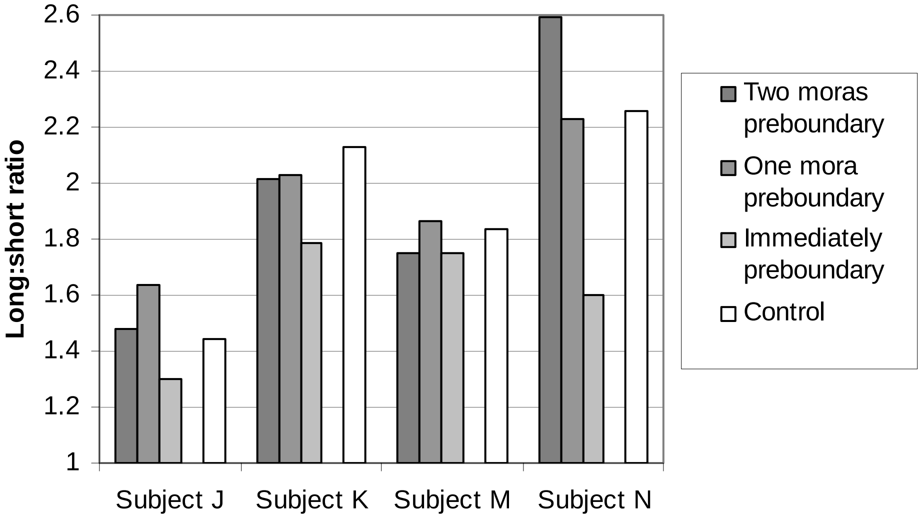
Table 5 – T-test results

Subject	Interaction between prosodic context and phonemic length category
J	$t(28) = 3.194, p < .005$
K	$t(28) = 4.162, p < .0005$
N	$t(28) = 9.864, p < .0001$

3.2. Effects on relative duration

This difference in the degree of phrase-final lengthening of phonemically short vowels versus phonemically long vowels also affects the duration ratios between them. This is noteworthy because the ratio of long vowels to short vowels in Japanese has been cited (Hirata 2004) as a source of invariance across speaking rates, allowing reliable identification of the phonetic length category of vowels despite large changes in their absolute duration. As Figure 8 illustrates, however, at least in the immediately preboundary condition, which showed significant differences in the degree of lengthening of short vowels relative to long vowels, the long-to-short ratio is smaller than in the control condition for all four subjects. The exact values are given in Table 6.

Figure 8 – Duration ratios of phonemically long vowels to phonemically short vowels



---

[11] The difference between the long- and short-vowel immediately preboundary-to-control ratios for Subject M, while in the correct direction, failed to reach significance (t(28)=1.395, p=.1738).

TABLE 6 – Mean long:short ratios

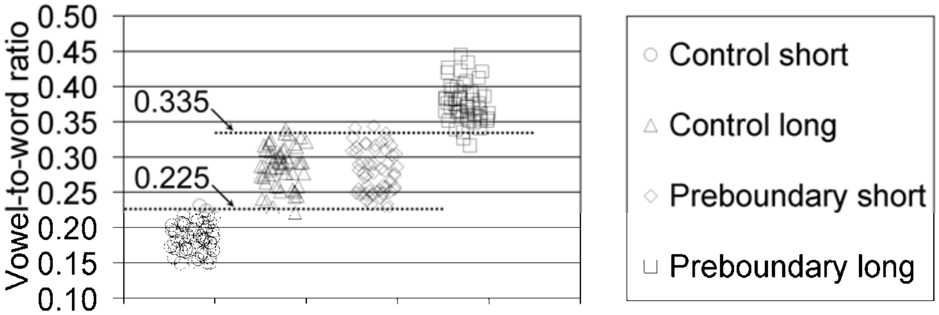
Subject	Two moras preboundary	One mora preboundary	Immediately preboundary	Control
J	1.5:1	1.6:1	1.3:1	1.4:1
K	2.0:1	2.0:1	1.8:1	2.1:1
M	1.8:1	1.9:1	1.7:1	1.8:1
N	2.6:1	2.2:1	1.6:1	2.3:1

In order to determine whether these differences are significant (if only for the long-to-short ratios of vowels in the immediately preboundary condition relative to controls), a repeated measures ANOVA was run on the long-to-short ratios of all four subjects. Also, a contrast of means test was run comparing the long-to-short ratios of the immediately preboundary and control conditions directly. The results showed that while the main effect of prosodic condition on ratio was only marginal,[12] the difference of means between the immediately preboundary and control conditions was indeed significant ($F = 5.443$, $p < .05$).

Another aspect of relational acoustic invariance that Hirata (2004) discusses is the duration ratio of a vowel to the word in which it occurs. This is shown to classify a vowel as phonemically short or long with 94.3% accuracy provided the number of moras in the rest of the word is known. For our data, this was the duration ratio of the final vowel in *dōkyo* (for short vowels) or *dōkyō* (for long vowels) to the entire word. Consistent with Hirata's finding, we found that dividing up all of the control (phrase-medial) tokens from all four subjects based on whether their vowel-to-word ratio was above or below 0.225:1 accurately classified 98.3% (118 out of 120) as phonemically long or short, respectively. However, since the scope of preboundary lengthening in our data was only about one mora, the vowel-to-word duration ratios of the immediately preboundary tokens were much higher. As a result, the same vowel-to-word ratio of 0.225:1 that accurately classified 98.3% of vowels from the control condition was found to classify all 120 immediately preboundary tokens as long. Since 60 of them were actually short, this put accuracy at 50%. While it was possible, using a ratio of 0.335:1, to accurately identify the phonemic length category of 96.7% (116 out of 120) of the immediately preboundary vowels, applying this ratio to the control condition tokens would classify all but one (out of 120) as short. In other words, there was no single (invariant) ratio that would accurately handle both control and immediately preboundary tokens.

A scatter plot of the vowel-to-word duration ratios is presented in Figure 9. The mean vowel durations, mean word durations, vowel-to-word duration ratios, and standard deviations are given in Table 7.

FIGURE 9 – Scatter plot of vowel-to-word duration ratios



---

[12] F(3,9)=3147, p=.0793.

TABLE 7 – Mean vowel durations, mean word durations, and vowel-to-word ratios

Prosodic condition	Control		Immediately preboundary	
Length category	Short	Long	Short	Long
vowel duration (SD)	60.497 (6.305)	115.815 (21.037)	111.968 (17.191)	179.019 (30.337)
word duration (SD)	328.729 (34.211)	401.495 (51.214)	400.274 (38.546)	474.621 (56.723)
vowel:word ratio (SD)	0.185:1 (0.020)	0.287:1 (0.028)	0.280:1 (0.033)	0.376:1 (0.026)

Perhaps most striking is the stark contrast between the minimal overlap of the vowel-to-word ratios of the short and long vowels within each prosodic condition (the basis of Hirata's claim) and the almost complete overlap between the vowel-to-word ratios of the control long vowels and the preboundary short vowels. As we saw above, this overlap makes it impossible to find a single (invariant) vowel-to-word ratio that can identify vowels as phonemically short or long with even a moderate degree of accuracy.

4. Discussion

Consistent with a variety of findings on phrase-final lengthening, including Takeda et al. (1989) on final lengthening in Japanese, both short and long vowels in the immediately preboundary condition were significantly longer than controls for all subjects. Moreover, the fact that lengthening was (largely) limited to the immediately preboundary condition suggests that the π-gesture associated with intonation phrase boundaries in Japanese has a preboundary temporal scope of approximately one mora. This is sufficient to overlap an entire short vowel (and preceding consonant) but only overlaps the latter portion of a long vowel. Consistent with the predictions of the π-gesture model (Byrd and Saltzman, 2003; Byrd *et al.*, 2006), this differential degree of overlap with the π-gesture resulted in a greater proportional degree of lengthening of immediately preboundary short vowels than immediately preboundary long vowels for three out of four subjects.

Our findings also challenge Hirata's (2004) claims concerning relational acoustic invariance in Japanese. First, we found that the duration ratio of phonemically long vowels to phonemically short vowels was significantly smaller in the immediately preboundary condition than in the control condition. This resulted from the fact that, as predicted, phonemically short vowels underwent a significantly greater degree of proportional lengthening than phonemically long vowels. We also found that a vowel-to-word duration ratio that reliably classified the control (phrase-medial) tokens as either short or long would classify all of the immediately preboundary tokens as long. Likewise, a ratio that reliably classified the immediately preboundary tokens would classify all but one (out of 120) of the control tokens as short. In other words, there was no single (invariant) ratio that could accurately handle both control and immediately preboundary tokens. Thus, our results suggest, consistent with what Berkovits (1993) proposed for VC ratios in Hebrew, that listeners cannot depend on duration ratios alone in identifying phonemic categories. Rather, it is necessary to take prosodic context into account.

As a final note, some researchers (e.g. Cho and Jun, 2000; Cho and Keating, 2001; Hacopian, 2003; Keating *et al.*, 2004; Cho *et al.*, in press) have discussed prosodic boundary effects in terms of listener-oriented goals such as featural enhancement. That is to say that, for these authors, the fact that gestures tend to "get larger, longer, and further apart" at phrase edges (Byrd and Saltzman, 2003, p. 159) is intended to aid the listener in perceiving contrasts. However, given that duration is the primary cue to the phonemic length category of Japanese vowels (Fujisaki *et al.*, 1975), prosodic lengthening (particularly in different proportions for short vowels versus long vowels) is certainly not doing the listener any favors. On the contrary, as we have seen throughout this paper, phrasal lengthening wreaks havoc on what would otherwise be very reliable cues to the length category of vowels. Moreover, we see no evidence that prosodic boundary effects in Japanese, in which duration is contrastive for both consonants and vowels, are different from prosodic boundary effects in English, in which duration is not generally contrastive. These facts suggest to us that prosodic boundary effects are not listener oriented, but rather are implementation strategies employed by the speaker in articulating speech.

Appendix

TABLE 8 – Mean durations and standard deviations

	Control mean (SD) ms	Two Moras Preboundary mean (SD) ms	One Mora Preboundary mean (SD) ms	Immediately Preboundary mean (SD) ms
Subject J (*n* = 120)				
Short	59.4 (5.5)	60.5 (6.5)	57.4 (5.9)	125.8 (11.3)
Long	85.7 (8.6)	89.2 (7.3)	93.9 (10.4)	163.2 (15.4)
Subject K (*n* = 120)				
Short	64.0 (5.4)	63.4 (7.2)	62.9 (5.4)	123.5 (13.7)
Long	136.3 (11.0)	127.7 (9.8)	127.5 (8.7)	220.7 (26.1)
Subject M (*n* = 120)				
Short	62.9 (3.7)	68.5 (5.6)	65.3 (4.5)	99.6 (11.7)
Long	115.6 (5.9)	120.1 (9.2)	121.8 (12.0)	174.1 (13.8)
Subject N (*n* = 120)				
Short	55.7 (7.0)	49.9 (5.4)	58.6 (7.1)	98.3 (10.0)
Long	125.6 (10.7)	129.2 (7.1)	130.6 (8.8)	158.1 (11.9)

References

Berkovits, R. (1993). Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36, 89–98.

Bloch, B. (1942). Studies in colloquial Japanese IV: Phonemics. *Language* 26, 86–125.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica* 57, 3–16.

Byrd, D., Kaun, A., Narayanan, S. & Saltzman, E. (2000). Phrasal signatures in articulation. In Broe, M. B. & Pierrehumbert, J. B. (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, 70–87. London: Cambridge University Press.

Byrd, D., Kirvokapić, J. & Lee, S. (2006). How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America* 120, 1589–1599.

Byrd, D. & Saltzman, J. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149–180.

Cho, T. & Jun, S.-A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. *UCLA Working Papers in Phonetics* 99, 57–70.

Cho, T. & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29, 155–190.

Cho, T., McQueen, J. M. & Cox, E. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics* 35, 210–243.

Denes, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America* 27, 761–764.

Fujisaki, H., Nakamura, K. & Imoto, T. (1975). Auditory perception of duration of speech and non-speech stimuli. In Fant, G. & Tatham, M. A. A. (eds.), *Auditory Analysis and Perception of Speech*, 197–219. London: Academic Press.

Hacopian, N. (2003). A three-way VOT contrast in final position: Data from Armenian. *Journal of the International Phonetic Association* 33, 51–80.

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

Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics* 32, 565–589.

Keating, P., Cho, T., Fougeron, C. & Hsu, C. (2004). Domain-initial articulatory strengthening in four languages. In Local, J., Ogden, R. & Temple, R. (eds.), *Papers in Laboratory Phonology VI: Phonetic Interpretation*, 143–161. London: Cambridge University Press.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 59, 1208–1221.

Oller, K. D. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 53, 1235–1247.

Pind, J. (1999). Speech segment durations and quantity in Icelandic. *Journal of the Acoustical Society of America* 106, 1045–1053.

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

Port, R. F. & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception and Psychophysics* 32, 141–152.

Port, R. F., Dalby, J. & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81, 1574–1585.

Svastikula, M. L. K. (1986). *A Perceptual and Acoustic Study of the Effects of Speech Rate on Distinctive Vowel Length in Thai*. PhD Dissertation, University of Connecticut.

Takeda, K., Sagisaka, Y. & Kuwabara, H. (1989). On sentence-level factors governing segmental duration in Japanese. *Journal of the Acoustical Society of America* 86, 2081–2087.

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

Venditti, J. J. (2005). The J_ToBI model of Japanese intonation. In Jun, S.-A. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 172–200. Oxford: Oxford University Press.

Warner, N. & Arai, T. (2001). The role of the mora in the timing of spontaneous Japanese speech. *Journal of the Acoustical Society of America* 109, 1144–1156.

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