

The Stress Foot as a Unit of Planned Timing: Evidence from Shortening in the Prosodic Phrase

Heejin Kim and Jennifer Cole

Department of Linguistics
University of Illinois at Urbana-Champaign, U.S.A.

{hkim17, jscole}@uiuc.edu

Abstract

This study investigates whether the stress foot is a planned timing unit in American English, by examining the durational characteristics of the foot in three different prosodic contexts - i) within an intermediate phrase, ii) across an intermediate phrase and iii) across an intonational phrase. The results show that as the number of syllables in a foot increases, the duration of the foot increases, but the mean duration of syllables is reduced. Our examination of the internal structure of the foot reveals that there is a consistent shortening of stressed syllables within an intermediate phrase. These findings indicate that the stress foot within the intermediate phrase is a timing unit where durational shortening occurs in compensation for an increase in syllable count within the foot.

1. Introduction

Speech rhythm has been extensively studied since Pike [1] and Abercrombie [2], who described rhythm in terms of isochronous units such as the stress foot and syllable. Subsequent research focuses on durational measurements of the alleged isochronous units in different languages. A weak version of isochrony is supported in [3] and [4], which report a tendency for each syllable in a foot to be shortened as the number of component syllables increases. Several studies argue for perceptual isochrony, claiming that listeners hear stresses as being more regularly spaced than they really are [5]. However, subsequent studies in the production and perception of rhythm have failed to find supporting evidence for strict isochrony [6, 7, 8]. In light of the inconsistent experimental evidence for physical isochrony, researchers point to phonological factors such as syllable structure as a possible determinant for rhythmic differences among languages [9, 10, 11]. Their main finding is that languages known as stressed-timed show greater variability in the duration of intervocalic intervals and a low proportion of vocalic to non-vocalic intervals compared to syllable-timed languages.

In summary, experimental studies of speech rhythm present no consistent evidence for isochronous units, but establish a relation between the rhythmic classification of a language and its syllable structure. We note that these findings do not necessarily deny the existence of a timing unit where durational regularization takes place. Although the evidence fails to provide strong support for the notion of an isochronous unit, it remains an open question whether there is any linguistic unit that serves more generally as a basis for systematic durational regularization.

The study reported here investigates the issue of speech timing as durational regularization in speech from the Boston University Radio News corpus [12]. We examine the duration and

the internal structure of the foot as a function of the number of component syllables in three different prosodic contexts: within an intermediate phrase, across an intermediate phrase boundary, and across an intonational phrase boundary. Our study shows that i) as the number of syllables in a foot increases, the duration of the foot increases, but ii) the duration of the stressed syllable in the foot is shortened and iii) this shortening effect occurs only within an intermediate phrase. Based on these findings, we claim that the foot is the planned timing unit within the intermediate phrase.

2. The foot as a temporal unit in English

2.1. Variability of the foot

While prior research does not justify the claim that the foot is the unit of isochrony in English, studies on interference of prosodic properties between first language and second language suggest that the foot has status as a temporal unit in English. For example, Bond and Fokes [13] find that in their English productions, native speakers of Malaysian and Japanese (syllable-timed languages) do not compress base words in proportion to the number of suffixed syllables, as do native speakers of Thai (a stressed-timed language). Ericksson [14] also offers evidence from regression analyses for the foot as an underlying rhythmic unit in various languages. Specifically, foot duration (I) is found to be a simple linear function ($I = a + bn$) of the number of syllables (n) contained in a foot, and stress-timed languages are found to have a larger constant term (a), compared to syllable-timed languages. However, as noted in [15], the source of the extra duration in the constant term for stress-timed languages is not yet clearly understood. Further investigation is needed to determine which, if any, syllables undergo shortening as a function of foot length. We examine these issues in our study, which measures the duration of the foot and also the duration of stressed syllables and unstressed syllable as a function of the number of syllables in the foot.

2.2. Effects of prosodic phrasing

Rakerd et al. [16] examine the interaction between syntactic boundaries (e.g. noun-phrase and verb-phrase) and foot-level shortening. Comparison of stressed syllable durations in different boundary conditions shows that stressed syllables are shortened by the same duration across a syntactic boundary as within a phrase. Based on this finding, they claim that foot-level shortening is independent of boundary lengthening effects. Although it is known that there is a strong correspondence between syntactic boundaries and prosodic phrasing, prosodic boundaries are not a simple mirror of syntactic boundaries. For exam-

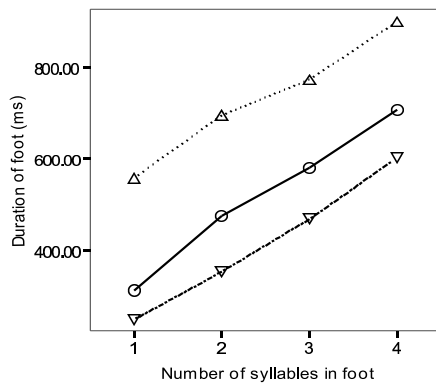


Figure 1: Foot duration against the number of syllables in a foot. ∇ -within-ip, \circ -across-ip, Δ -across-IP.

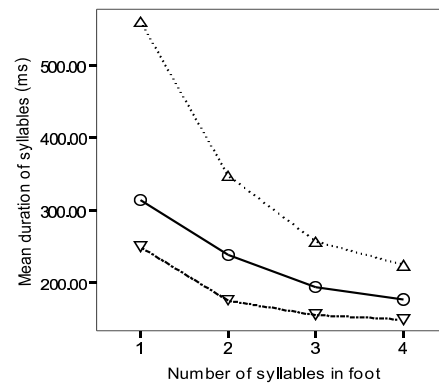


Figure 2: Mean syllable duration against the number of syllables in a foot. ∇ -within-ip, \circ -across-ip, Δ -across-IP.

ple, Fach [17] shows that due to rhythmic and other factors, only 65% to 84% of syntactic boundaries are coded in prosodic structure. This finding suggests that prosodic phrasing might be more directly related to temporal regularization than syntactic factors. Our study investigates whether the position of the foot in the prosodic phrase structure interacts with foot-level timing.

3. Method

3.1. Materials

The speech used in this study is the lab news portion of the Boston University Radio News corpus. This portion of the corpus includes a manual prosodic transcription which marks prosodic phrasing and pitch accent, following the ToBI labeling standard [18]. In addition, this corpus includes segment-level transcription, where primary and secondary lexical stresses are marked. Lexical stress for content words is correctly annotated in this corpus, but the assignment of lexical stress on function words is found to be unreliable. We corrected the stress feature of function words with manual re-coding, adopting Selkirk's [19] stress assignment principles. Monosyllabic function words are not assigned stress unless they are pitch-accented, while bisyllabic and other polysyllabic function words are assigned a primary stress. The analysis reported in this study was carried out on twelve minutes of speech from the female speaker F2B.

3.2. Measurements

In this study the stress foot is equated with the interval from one primary or secondary stress up to the following primary or secondary stress. The stress foot defined in this way may cross a word boundary, and in this sense our notion of stress foot differs from the notion typically invoked in phonological discussions of word-level stress. Durational measures are taken of each stress foot and its component stressed and unstressed syllables, and each foot is coded for its prosodic context: i) within an intermediate phrase, ii) crossing an intermediate phrase boundary, or iii) across an intonational phrase (henceforth, referred as within-ip, across ip-and across-IP, respectively).

The interval of the foot was measured from the vowel onset of the stressed syllable up to the vowel onset of the next stressed syllable. The choice of the vowel onset as the left edge of the foot, rather than the syllable onset (which may include the initial consonant or consonant cluster) is based on the understanding of the syllable 'p-center' as the locus of syllable timing.

Foot duration was automatically extracted from the sound files based on the lexical stress labels in the segment-level transcription. The number of syllables contained in each foot was also extracted, through the identification of each labeled vowel as a syllable nucleus.

Variation in foot duration was analyzed as a function of the number of syllables in a foot, for each prosodic context. In addition, the mean syllable duration for each foot was measured by dividing foot duration by the number of syllables in the foot. The strength of the relationship between foot duration and the number of syllables within the foot was assessed through Kendall's tau-b correlation analysis for each prosodic condition. The internal structure of the foot was examined through measurements of rhyme duration in both stressed and unstressed syllables, and these durations were analyzed as a function of the syllable count within the foot and as a function of the prosodic context of the foot. To control boundary effects such as domain-initial and final lengthening, the rhyme duration measurement is taken only from syllables that are in the medial position of an intermediate phrase.

4. Duration of the foot

4.1. Hypothesis

According to the isochrony hypothesis, as the number of syllables in the foot increases, the duration of the foot is expected to remain fairly constant with at most a weak increase, and mean syllable duration is expected to decrease. Since physical isochrony has been rejected in many previous studies, our main interest is to see how foot duration differs depending on the prosodic context. If foot duration and the number of component syllables are strongly correlated, we expect a higher correlation coefficient between these two factors.

4.2. Results

The mean foot duration is plotted against the number of component syllables in Figure 1. Each line represents foot duration in a different prosodic context. Figure 1 shows that for all prosodic contexts, foot duration increases as a function of the number of component syllables. A 1-way ANOVA was performed for each prosodic condition, comparing foot duration with the number of component syllables as a factor. For each prosodic condition, foot duration differed significantly at $p < .001$. (within-ip: $F(3, 934) = 489.785$, across-ip: $F(3, 133) = 55.222$, across-IP:

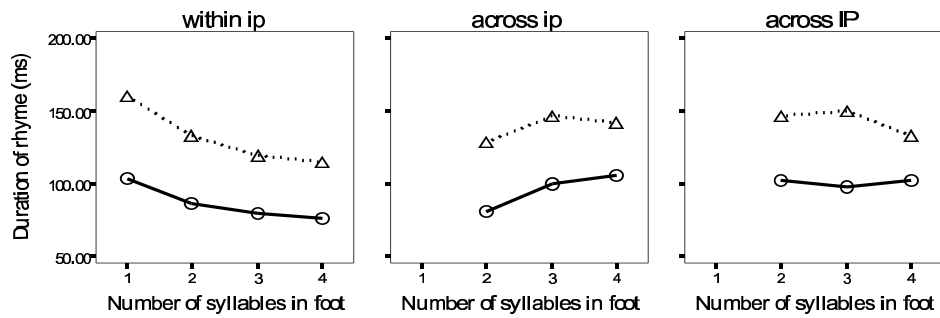


Figure 3: Mean duration of the rhyme of stressed syllables against the number of syllables in a foot. ○-when rime consists of one segment, △-when rime consists of two segments.

$F(3, 352) = 30.927$

Correlation coefficients between the foot duration and the number of syllables in the foot were significant for all prosodic conditions at $p < .001$, but the correlation is strongest for the within-ip condition (.629), while the across-ip condition showed a weaker correlation (.598) and the across-IP condition showed the weakest correlation (.315).

In Figure 2, the mean syllable duration is plotted against the number of component syllables. It is shown that the mean duration of syllables within a foot decreases as the number of component syllables increases.

4.3. Discussion

The observed increase in foot duration as a function of an increase in the number of component syllables suggests that there is no general robust pattern of shortening that holds foot duration constant in the prosodic contexts considered here. However, we also observe that the mean duration of syllables decreases as a function of an increased number of syllables in a foot, which may be indicative of a shortening effect. It is important to note that as the number of component syllables increases, the additional syllables are all unstressed. Given the general observation of longer segments in stressed syllables compared to unstressed syllables in English, a decrease of mean syllable duration is expected with the addition of unstressed syllables, even without any additional shortening effect in the foot. In order to examine the source of the observed decrease in mean syllable duration, we must look inside the foot. Specifically, we must investigate whether the duration of stressed syllables, unstressed syllables or both decreases as a function of the number of component syllables.

We note that foot duration values are different depending on prosodic condition. The duration of the foot is shortest within-ip, and increases in the order within-ip < across-ip < across-IP. In addition, the correlation coefficient is highest for the within-ip condition and lowest for the across-IP condition. One might argue that longer duration of the foot in across-ip and across-IP conditions is due to final lengthening and the presence of pauses at phrase boundaries. Similarly, the lower correlation coefficient in across-ip and across-IP conditions can be interpreted as reflecting variability in the lengthening and pause phenomena. In order to investigate the difference between prosodic contexts in terms of foot duration and its correlation with syllable count, we compare the durational characteristics of stressed and unstressed syllables as a function of increased syllable count in the foot, for different prosodic contexts.

In summary, the results shown in this section demonstrate positive correlations between syllable count and foot duration. The next section looks at durations of stressed and unstressed syllables within the foot for evidence of shortening and for possible effects of prosodic context on shortening.

5. Internal structure of foot

5.1. Hypotheses

Evidence for a foot-level shortening effect will be in the observation of decreased duration of stressed or unstressed syllables as a function of an increase in the number of syllables in the foot. We formulate two hypotheses about the effect of prosodic context on syllable duration.

Hypothesis 1: longer foot duration in across-ip and across-IP conditions is due to final lengthening and the occurrence of pause at boundary positions, and there is no interaction between foot-level shortening and the prosodic context. Under this hypothesis, we expect the effect of the syllable count on syllable duration to be the same for each prosodic condition.

Hypothesis 2: longer foot duration in the across-ip and across-IP conditions is due to the interaction between a shortening effect and the prosodic context. In other words, the extra duration in across-ip and across-IP conditions reflects either a weaker shortening effect or no foot-level shortening compared to the within-ip condition. Under this hypothesis, adding more syllables in the foot after a prosodic phrase boundary should not affect the duration of the syllables in a foot.

5.2. Results

The mean duration of the rhyme of stressed syllables is plotted against the number of component syllables in Figure 3. Each panel represents a different prosodic condition. Since the duration of the rhyme differs depending on the number of segments in the rhyme, rhyme duration is plotted separately for rhymes with one and two segments. Figure 3 clearly shows that stressed syllables are shortened only in the within-ip condition. The results of 1-way ANOVA confirm that the mean duration of the rhyme of stressed syllables is significantly different depending on the number of syllables in a foot only under the within-ip condition (rhymes with one segment: $F(3, 329) = 17.988, p < .001$, rhymes with two segments: $F(3, 238) = 3.801, p < .05$). In stark contrast to this pattern for stressed syllables, we observe that rhyme duration of unstressed syllables does not decrease in any of the prosodic conditions, as shown in Figure 4.

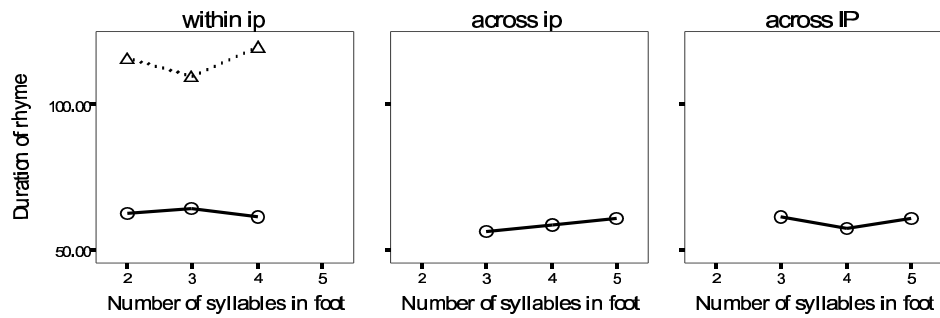


Figure 4: Mean duration of the rhyme of unstressed syllables against the number of syllables in a foot. ○-when rhyme consists of one segment, △-when rhyme consists of two segments. (Since the across-ip and across-IP conditions have only few tokens for rhymes with two segments, those cases are not shown)

5.3. Discussion

Our first main finding in this section is that only stressed syllables are shortened as a function of the increased number of syllables in a foot. Thus, it is clear that the decrease of mean duration of syllables found in section 4 reflects the reduced duration of stressed syllables. As to the effect of prosodic context, we found that stressed syllables are shortened only when the foot occurs within the ip. Thus, the longer duration of the foot that crosses ip and IP boundaries (Figure 1) is not only due to lengthening or a pause at the boundary position, but also due to the absence of stressed-syllable shortening as a function of added syllables after the boundary.

6. Conclusion

To summarize, as the number of syllables in a foot increases, the duration of the foot increases, but the duration of the rhyme of stressed syllable in the foot is shortened. This finding suggests that rhythmic regularization in American English is manifested not in terms of isochronous foot duration, but through an adjustment of the duration of stressed syllables. Another interesting finding is that the shortening effect occurs only within the intermediate phrase, suggesting that prosodic structure plays an important role in rhythmic organization in American English. Taking these findings together, we claim that the foot within the ip is a timing unit where a certain level of rhythmic stability exists.

In ongoing work, we are examining evidence for foot-level shortening in other speakers from the Radio News corpus. We are also investigating the temporal patterning of larger prosodic domains such as the interval between pitch accented syllables.

7. References

- [1] Pike, K.L., *The intonation of American English*, University of Michigan Press, Ann Arbor, MI, 1945.
- [2] Abercrombie, D., *Elements of general phonetics*, Edinburgh University Press, Edinburgh, 1967.
- [3] Fowler, C., *Timing Control in Speech Production*, Indiana University Linguistics Club, Bloomington, IN, 1977.
- [4] Huggins, A. W. F., "On the perception of temporal phenomena in speech", *Journal of the acoustical society of America*, Vol. 51. pp. 1279-1290, 1972.
- [5] Lehiste, I., "Isochrony reconsidered", *Journal of Phonetics*, Vol. 5, pp. 253-263, 1977.
- [6] Roach, P., "On the distinction between 'stress-timed' and 'syllable-timed' language", *Linguistic controversies* (D. Crystal, ed.), pp. 73-79, Edward Arnold, London, 1982.
- [7] Scott, D. R., Isard, S. D., and Boysson-Bardies, B. D., "Perceptual isochrony in English and in French", *Journal of Phonetics*, Vol. 13, pp. 155-162, 1985.
- [8] Fant, G., Kruckenberg, A., and Nord, L., "Durational correlates of stress in Swedish, French and English", *Journal of Phonetics*, 19(3,4): 351-365, 1991.
- [9] Dasher, R. and Bolinger, D., "On pre-accentual lengthening", *Journal of the International Phonetic Association*, Vol. 12, pp. 58-69, 1982.
- [10] Ramus, F., Nespors, M., and Mehler, J., "Correlates of linguistic rhythm in speech", *Cognition*, Vol. 73, pp. 265-292, 1999.
- [11] Grabe, E. and Low, E., "Durational Variability in Speech and the Rhythm Class Hypothesis", *Studies in Laboratory Phonology VII*, CUP, 2002.
- [12] Ostendorf, M., Price, P. J., and Shattuck-Hufagel, S., *The Boston University Radio News Corpus*, Linguistics Data Consortium, <http://www ldc.upenn.edu>, 1995.
- [13] Bond, Z. S. and Fokes, J., "Non-native patterns of English syllable timing", *Journal of Phonetics*, Vol. 13, pp. 407-420, 1985.
- [14] Ericksson, A., *Aspects of Swedish Speech Rhythm*, University of Göteborg, Göteborg, 1991.
- [15] O'Dell, M. L. and Nieminen, T., "Coupled oscillator model of speech rhythm", *Proceedings of the 14th International Congress of Phonetic Sciences*, Vol. 2, pp. 1075-1078, 1999.
- [16] Rakerd, B., Sennett, W., and Fowler, C. A., "Domain-final lengthening and foot-level shortening in spoken English", *Phonetica*, 44(3):147-155, 1984.
- [17] Fach, M. L., "A comparison between syntactic and prosodic phrasing", *Proceedings of Eurospeech*, Vol.1:527-530, 1999.
- [18] Beckman, M. E. and Ayers, G., *Guidelines for ToBI Labelling*, m.s., The Ohio State University, 1997.
- [19] Selkirk, E., *Phonology and Syntax: The Relation between Sound and Structure*, MIT Press, Cambridge, MA, 1984.