Prosodic effects on temporal structure of monosyllabic CVC words in American English

[ac]*Yoonsook Mo*, [ac]*Jennifer Cole, and* [bc]*Mark Hasegawa-Johnson*

[a]Department of Linguistics, [b]Department of Electrical and Computer Engineering,
[c]Beckman Institute, University of Illinois
{ymo; jscole; jhasegaw}@illinois.edu

Abstract

Prosody serves an important function in speech communication: prosodic phrasing groups words into pragmatically and semantically coherent smaller chunks and prosodic prominence encodes the discourse-level status and rhythmic structure of a word within a phrase. Acoustic cues to prosody are available from the speech signal and can be used by listeners to recover the pragmatic and discourse meaning intended by speakers. Effects of prosodic context on the duration of consonants and vowels have been widely reported, and this study extends that line of work by examining how prosodic phrase boundary and prominence influence the temporal structure of the monosyllabic CVC word, based on an analysis of speech excerpts from the Buckeye corpus of spontaneous conversational American English.

Prosody annotation for these speech materials is obtained from 97 untrained, non-expert listeners. The results confirm findings from prior studies, showing that (1) monosyllabic CVC words are lengthened before a prosodic phrase boundary and under prominence, and (2) all subcomponents of a syllable, that is, the onset, nucleus, and coda of the monosyllabic word, are elongated. The findings further show that (3) the magnitude of lengthening associated with prosody varies as a function of syllable position, and (4) the magnitude of lengthening of subcomponents of monosyllabic CVC words varies as a function of prosodic characteristics. Nucleus duration is most strongly affected by both prosodic prominence and boundary and the onset and the coda of the monosyllabic word is also affected but to a lesser degree. The lengthening effect of prosodic phrase boundary on the coda is larger than the lengthening effect on onset duration while lengthening of the onset under prosodic prominence is larger than lengthening of the coda. The findings indicate that prosodic context shapes the internal temporal structure of the monosyllabic CVC word.

1. Introduction

Prosodic structure groups the words of an utterance into hierarchically layered phrasal constituents in which phrasal prominence is assigned to one or more words. Prosodic phrase boundaries and prosodic prominence are signaled through the modulation of many acoustic parameters, including spectral prominences, F0, intensity and duration, and these cues influence listeners' comprehension of the utterance. Among the acoustic correlates of prosody, phone duration is a primary cue that signals both prosodic prominence and boundary. This study asks how the effects of prominence and boundary on the temporal structure of speech differ by examining prosodic effects on phone duration within subcomponents of the syllable.

Prior studies demonstrate that both prosodic phrase boundary and prominence affect the temporal properties of words or syllables through boundary- and prominence-related lengthening [e.g., see 5, 8, 11, 12, 14 for prosodic prominence and 2, 3, 4, 6, 9, 13, 15, 16 for prosodic boundary]. Many studies have looked in particular at the domain of lengthening due to prosodic prominence and boundary. For instance, Turk and Sawusch [12] find that accentual lengthening effects extend through at least one unstressed syllable following the accented syllable, within the word. Cambier-Langeveld and Turk [5] later show that syllables within an accented word are lengthened but the degree of lengthening within a word varies as a function of position within a word. That is, they find that adjacent syllables to the right of the accented syllable are lengthened more than adjacent syllables to the left of an accented syllable. Turk and White [14] find similar effects, where accentual effects on duration spread over syllables which are not adjacent to an accented syllable.

There are also many studies investigating the domain of boundary induced lengthening effects. In an early study, Klatt [9] finds that phrase-final syllables are lengthened, and later Wightman and his colleagues [15] report that segmental lengthening induced by prosodic boundary is restricted to the rhyme of the pre-boundary syllable. In Dutch, Cambier-Langeveld [4] show that regardless of stress position and of the depth of prosodic boundary, final lengthening is not confined only to the final segment or the final rhyme but the amount of lengthening is largest in the final segment and gradually decreases as the distance from the boundary increases. In their articulatory study, Byrd and her colleagues [3] find that segments after a boundary as well as before a boundary are lengthened. The recent study by Turk and Shattuck-Hufnagel [13] confirms earlier findings for boundary lengthening and also shows evidence of boundary-related lengthening in a potentially non-contiguous domain that includes the final syllable and the rhyme of the main-stress syllable.

The studies cited above have examined durational effects of prosody in read speech that is elicited in the laboratory, where punctuation or other text marking devices are used to evoke the desired prosodic structures. Among the few studies to examine the temporal encoding of prosody in spontaneous speech, Aylett & Turk [1] report prominence-induced lengthening of syllables in spontaneous speech from a Map Task corpus, in analysis that considers the number of phones in the syllable, but which does not report prosodic lengthening effects on sub-constituents of the syllable, or on individual phones. Greenberg et al. [8], on the other hand, does examine lengthening effects at the phone level due to prominence ('stress accent' in the terminology of that study), with spontaneous speech data from the Switchboard corpus of telephone conversations in American English. Findings from this study indicate that the magnitude of the accentual lengthening effect is largest in the nucleus of an accented syllable, with a smaller effect on onset consonants, and no significant effect on coda consonants.

The current study asks whether prosodic prominence and boundary exert similar effects on the temporal structure of subcomponents of monosyllabic CVC words in spontaneous speech, extending the approach of Greenberg et al. [8] in two ways. First, we consider accent and boundary effects and their interaction on temporal patterns. Second, whereas Greenberg et al. identify 'stress accent' based on the transcriptions of two trained, expert transcribers, we examine prosody from the perspective of a population of ordinary listeners.

2. Methodology

2.1. Rapid Prosody Transcription (RPT)

97 ordinary listeners from undergraduate linguistics courses at the University of Illinois at Urbana-Champaign participated in a prosody transcription task which included a total of 54 speech excerpts from 38 speakers that were extracted from the Buckeye corpus of American English spontaneous speech [10]. Transcribers were given minimal definitions of prosodic prominence and boundary as follows.

"In normal speech, speakers pronounce some word or words in a sentence with more prominence than others. The prominent words are in a sense highlighted for the listener, and stand out from other non-prominent words. In some of the excerpts you will hear, you will be asked to mark all prominent words by underlining them….

Another feature of normal speech that we are interested in is the way speakers break up an utterance into chunks. These chunks group words in a way that helps the listener interpret the utterance, and are especially important when the speaker produces long stretches of continuous speech.… It is important for you to know that the boundary between two chunks does not necessarily correspond to the location where you would place a comma, period, or other punctuation mark, so you must really listen and mark the boundary where you hear a juncture between chunks. A chunk may be as small as a single word, or it may contain many words, and speakers can vary quite a bit in the size of the chunks they produce in a given utterance."

Listeners were seated at a computer equipped with personal headphones and then asked to mark the locations of prosodic prominence and boundary on words on a printed transcript where all punctuation and capitalization had been removed. Transcription was done in real time and was not aided by any visual display of speech. In each session, listeners marked only one prosodic feature (either prosodic prominence or boundary) on the transcripts.

Each excerpt was transcribed by a group of between 10-22 transcribers. Transcriptions from 6 transcribers were excluded because of failure to follow transcription instructions or because they identified themselves as non-native speakers of American English on the language background questionnaire. Transcriptions are pooled across listeners and each word is assigned a probabilistic prominence (P-score) and boundary (B-score), depending on the number of listeners who marked the word as prominent or followed by a boundary. The P-and B-scores range from 0 to 1, as shown in Fig.1. For example, in Fig.1, around 35% of listeners heard the first word "I" as prominent but no listener heard it as followed by a boundary.

2.2. Reliability of RPT

We calculated Fleiss' kappa statistic for multi-rater agreement to access the reliability of transcriptions [7]. Kappa scores (.392 for P-scores and .580 for B-scores across 97 listeners) were converted to Z-statistics to evaluate significance. The results showed that all Fleiss' kappa scores are much above chance ($p < 0.001$) with a 99% confidence interval, confirming that transcriptions of prosodic prominence and boundary by untrained non-expert listeners using the RPT method are reliable and consistent.

2.3. Durational measurements

Using the word and phone transcriptions available from the Buckeye corpus, 771 monosyllabic CVC words were extracted from the speech excerpts. The durations of the onset, nucleus, and coda of each monosyllabic CVC word were measured and the relative proportions of the onset, nucleus, and coda within the monosyllabic word were calculated.

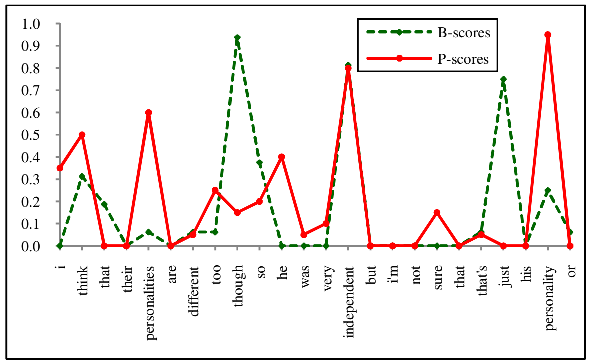

Figure 1. *Graph of probabilistic P(rominence)-scores (solid line) and B(oundary)-scores (dotted line) for each word in a sample utterance (speaker 26) from the Buckeye corpus. Prosody scores are based on pooled transcriptions of 22 transcribers*

3. Results

The following graphs (Figs. 2, 3) demonstrate the relationship between prosody scores (P- and B-scores) and the raw durations of the onset, nucleus, and coda of the monosyllabic CVC words, where non-parametric Spearman's correlation analyses show that prosody scores are positively correlated with each raw duration ($p<0.001$). Duration measures of the onset, nucleus, and coda of the monosyllabic CVC words are all longer in words that are more likely to be perceived as prominent or as followed by a prosodic boundary, and the same durations are all similar to one another in words that are more likely to be perceived as not prominent, or as followed by a prosodic boundary. As prosody scores (P- and B-scores) increase, the durations of the onset, nucleus, and coda of the monosyllabic CVC words increase monotonically. Looking closely, however, the results show that although the durations of all the subcomponents of the monosyllabic words are longer in words with that are more often perceived as prominent or as followed by a prosodic boundary, the magnitude of lengthening varies as a function of the position within a syllable. The lengthening effect of the nucleus is largest, both for prominent words (with high P-scores) and for phrase-final words (with high B-scores). On the other hand, the duration of onsets shows a greater lengthening effect of perceived prominence than do codas, while the duration of codas are more affected by prosodic boundary than are onsets.

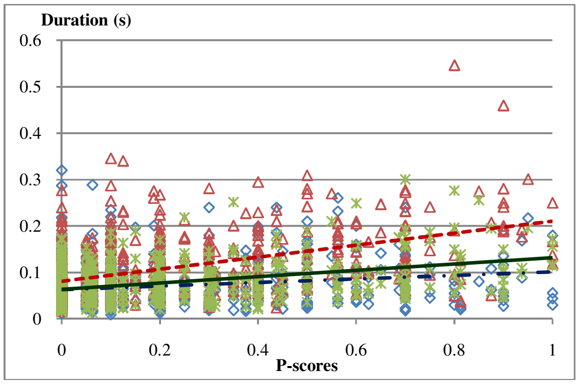

Figure 2. *Scatterplot with regression lines between P-scores and the raw durations of the onset (∗, —), nucleus (△, ---), and coda (◇, –·–) of the monosyllabic CVC words*

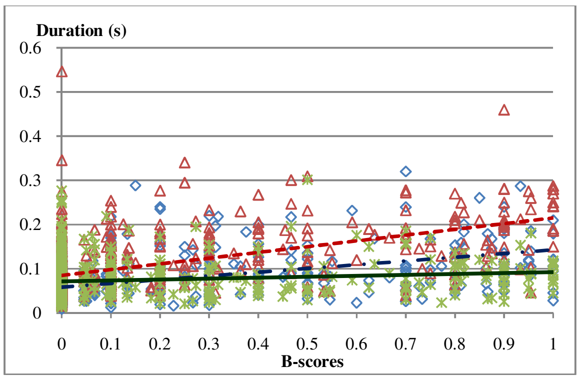

Figure 3. *Scatterplot with regression lines between B-scores and the raw durations of the onset (∗, —), nucleus (△, ---), and coda (◇, –·–) of the monosyllabic CVC words*

Figs. 4 and 5 demonstrate the relative effects of prosody on the temporal structures of the subcomponents of the monosyllabic CVC words. As perceived prominence of a word increases, the ratio of nucleus duration to overall duration in monosyllabic CVC words increases but the ratio of the onset levels off and the coda ratio decreases. As for the effect of prosodic boundary, when a word is perceived as followed by a boundary, the ratio of nucleus duration to overall duration increases for the same data, while the ratio of the coda remains almost the same and the onset ratio decreases.

4. Discussion

The current study examines whether prosodic context affects the temporal structures of the subcomponents of the monosyllabic CVC words and the effects are uniform on the onset, nucleus, and coda. The current study investigates whether the temporal effects of prosodic prominence and boundary, the two major sources of lengthening, are similar on the syllabic structures of monosyllabic CVC words.

The findings from this study show that the durations of monosyllabic CVC words tend to be longer for words that listeners hear as prominent or followed by a boundary. Yet the magnitude of temporal effects of prosodic context is not uniform over all syllable positions within the monosyllabic CVC word. Regarding the effects of perceived prosodic prominence, nucleus duration shows the greatest increase due to prominence, followed by the durations of the onset and of the coda, in order. These findings are consistent with those of Greenberg et al. [8], confirming that the durations of subcomponents of the syllable are affected by prosodic prominence in spontaneous speech, where prosody is identified as it is perceived by untrained, non-expert listeners.

Similarly, regarding the effects of perceived prosodic boundary, nucleus duration again shows the largest lengthening effect. But contrary to the effects of prosodic prominence, codas showed a bigger lengthening effect than onsets when perceived with a following prosodic boundary. The findings are somewhat contradictory to those from prior studies which found that the segment nearest to the boundary is lengthened to the largest degree and the magnitude of lengthening by boundary decreases as the distance from a boundary increases. This difference might be due to the nature of the speech materials used in the current study, which are excerpted from conversational speech. The number and the location of prosodic prominence and boundary may differ in spontaneous vs. scripted materials, and the frequency of words that are both prominent and phrase-final, which is common in our materials, may also be different from the scripted laboratory materials. To further explore this matter, we will have to analyze a larger corpus from which all words with high P-scores *and* high B-scores are excluded.

How does prominence- and boundary-related lengthening modulate the temporal structures of the syllable of the monosyllabic CVC words? In Figs. 6 and 7, the proportion of each subcomponent within the monosyllabic CVC words is schematically displayed. As shown in Fig. 6, when the monosyllabic CVC words are not typically perceived as prominent each subcomponent of the syllable consists of almost equal duration. For words perceived as prominent, however, the relative duration of the nucleus increases, apparently consuming some of the coda duration, while the relative duration of the onset remains the same. Fig. 7 demonstrates, on the other hand, that the durations of onset, nucleus, and coda duration relative to total syllable duration are almost equal for words that are perceived as not followed by a boundary. The relative duration of the nucleus of the syllable greatly increases, taking over that of the onset, in words perceived as followed by a boundary. The proportional duration of the coda remains unchanged in words that are followed by a boundary.

5. Conclusions

This study shows that prosodic features modulate the temporal structures of the subcomponents of the monosyllabic CVC word in everyday conversational speech, and that untrained non-expert listeners are sensitive to the modulation of temporal structures in their perception of prosody. When words are perceived as prominent the proportional duration of the nucleus increases while that of the coda decreases; when words are perceived as followed by a boundary the proportional duration of the nucleus increases while that of the onset decreases.

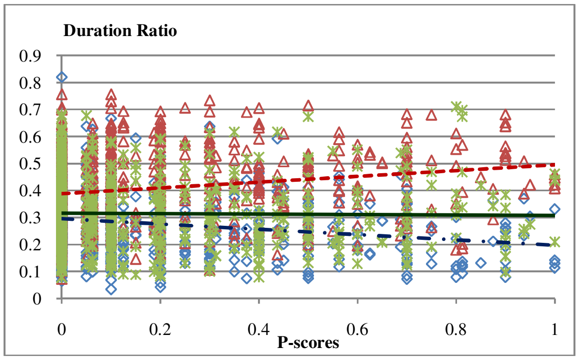

Figure 4. *Scatterplot with regression lines between P-scores and the ratio of the onset (✳, —), nucleus (△, ---), and coda (◇, –·–) duration to syllable duration in monosyllabic CVC words*

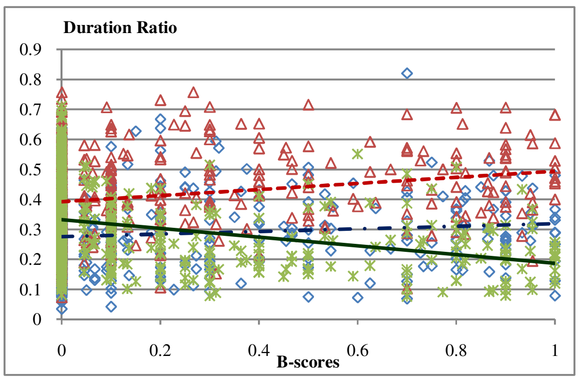

Figure 5. *Scatterplot with regression lines between B-scores and the ratio of the onset (✳, —), nucleus (△, ---), and coda (◇, –·–) duration to syllable duration in monosyllabic CVC words*

6. Acknowledgement

This research is supported by NSF grants IIS 07-03624 to Jennifer Cole. I would like to thank the members of Prosody-ASR group for their comments.

7. References

[1] Aylett, M. and Turk, A., The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech,* 47 (1), 31-56, 2004.

[2] Beckman, M. E. and Edwards, J., Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston and M. E. Beckman (editors), *Papers in laboratory phonology I: between the grammar and physics of speech*, Cambridge University Press: Cambridge. 1990.

[3] Byrd, D., Krivokapic, J., & Lee, S., How far, how long: On the temporal scope of prosodic boundary effects. *Journal of the Acoustical Society of America*, 120(3), 1589–1599. 2006.

[4] Cambier-Langeveld, T., The domain of final lengthening in the production of Dutch. In J. Coerts and H. d. Hoop (editors), *Linguistics in the Netherlands*, John Benjamins: Amsterdam, 13–24, 1997.

[5] Cambier-Langeveld, T. and Turk, A. E., A comparison of accentual lengthening in Dutch and English. *Journal of Phonetics,* 27, 255–280, 1999.

[6] Crystal, T. H. and House, A. S., Segmental durations in connected-speech signals: Current results. Journal of the Acoustical Society of America, 83, 1553-1573, 1988.

[7] Fleiss, J. L., Measuring nomial scale agreement among many raters. Psychological Bulletin, 76 (5), 378-382, 1971.

[8] Greenberg, S., Carvey, H., Hitchcok, L., and Chang, S., Temporal properties of spontaneous speech- a syllable-centric perspective. *Journal of Phonetics*, 31, 465-485, 2003.

[9] Klatt, D. H., Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3, 129–140, 1975.

[10] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E., *Buckeye Corpus of Conversational Speech* (2nd release), Department of Psychology, Ohio State University (Distributor), Columbus, OH, Online: www.buckeyecorpus.osu.edu, 2007.

[11] Turk, A. E. and Sawusch, J. R., The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, 99 (6), 3782-3790, 1996.

[12] Turk, A. E. and Sawusch, J. R., The domain of accentual lengthening in American English. *Journal of Phonetics*, 25, 25-41. 1997.

[13] Turk, A. E. and Shattuck-Hufnagel, S., Multiple targets of phrase-fianl lengthening in American English words. Journal of Phonetics, 35, 445-472. 2007.

[14] Turk, A. E. and White, L., Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27, 171-206, 1999.

[15] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J., Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91(3), 1707–1717, 1992.

[16] Yoon, T-J., Cole, J., & Hasegawa-Johnson, M., On the edge: Acoustic cues to layered prosodic domains, In *proceedings of 14th International Congress of Phonetic Sciences*. Saarbrucken, Germany (Aug. 6-10), 2007.

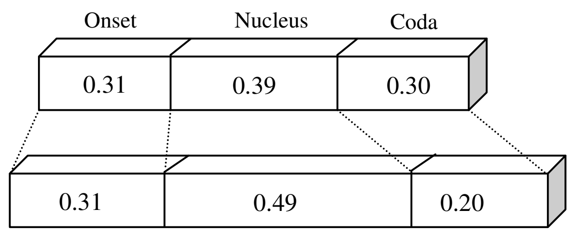

Figure 6. *Schematic representation of temporal structure of the monosyllabic CVC word: non-prominent word (P-score =0, top) vs. prominent word (P-score =1, bottom)*

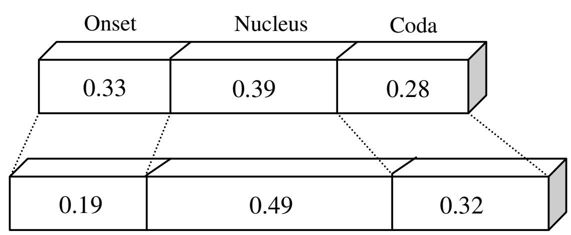

Figure 7. *Schematic representation of temporal structure of the monosyllabic CVC word: phrase-medial word (B-score=0, top) vs. phrase-final word (B-score=1, bottom)*