

## Prosodic Modifications of the Internal Phonetic Structure of Monosyllabic CVC Words in Conversational Speech

Mo, Yoonsok<sup>1)</sup>

### ABSTRACT

Previous laboratory studies have shown that prosodic structures are encoded in the modulations of phonetic patterns of speech including suprasegmental as well as segmental features. In particular, effects of prosodic context on duration and intensity of syllables and words have been widely reported. Drawing on prosodically annotated large-scale speech data from the Buckeye corpus of conversational speech of American English, the current study attempted to examine whether and how prosodic prominence and phrase boundary of everyday conversational speech, as determined by a large group of ordinary listeners, are related to the phonetic realization of duration and intensity. The results showed that the patterns of word durations and intensities are influenced by prosodic structure. Closer examinations revealed, however, that the effects of prosodic prominence are not the same as those of prosodic phrase boundary. With regard to intensity measures, the results revealed the systematic changes in the patterns of overall RMS intensity near prosodic phrase boundary but the prominence effects are restricted to the nucleus. In terms of duration measures, both prosodic prominence and phrase boundary are the most closely related to the lengthening of the nucleus. Yet, prosodic prominence is more closely related to the lengthening of the onset while phrase boundary lengthens the coda duration more. The findings from the current study suggest that the phonetic realizations of prosodic prominence are different from those of prosodic phrase boundary, and speakers signal different prosodic structures through deliberate modulations of the internal phonetic structure of words and listeners attend to such phonetic variations.

**Keywords:** prosodic prominence, prosodic boundary, overall intensity

### 1. Introduction

In spoken languages, words are not linearly concatenated but are organized by the prosodic structure comprising prosodic phrases of different sizes (e.g., syllable, foot, word, and phrase) and prominence relations among words or phrases. Prosodic phrasing groups words into pragmatically and semantically coherent small chunks, and prosodic prominence encodes the discourse-level status and rhythmic structure of a word within a phrase. The prosodic structure determines the relative importance of words in utterances. In everyday speech communication, therefore, the speaker shapes speech utterances through the

modifications of the phonetic patterns of speech to signal the prosodic structure, and the listener must attend to such phonetic variation to recover the prosodic structure as intended by the speaker. Especially, elements at the edges of prosodic phrases and those assigned prominence are phonetically distinct from similar elements in different prosodic contexts.

The current paper concerns the prosody-related phonetic modulations of speech in conversational speech. More specifically, the primary goal of this paper is to examine the prosodic modulations of the acoustic parameters of monosyllabic CVC words in conversational speech of American English, where the locations of prosodic prominence and phrase boundary are determined by untrained, ordinary listeners. Among many acoustic parameters, duration and intensity are of interest in the current study.

Much prior research examined the influence of prosodic prominence and phrase boundary on the phonetic patterns of

---

1) Chonnam National University, myoons@jnu.ac.kr

temporal properties of words and syllables. For example, in their study on accentual lengthening, Turk and Sawusch [1] designed a series of sentences with focal accents containing di- or tri-syllabic words varying in terms of the locations of accent (accented vs. unaccented), positions within a phrase (non-final vs. final), and locations of consonants relative to accented syllable (tautosyllabic vs. heterosyllabic). Based on the production data by 20 speakers, they found the accent-related lengthening. In a later study, Cambier-Langeveld and Turk [2] also examined the accent-related lengthening and its distribution with carefully designed target words located in focally accented or unaccented positions. Like Turk and Sawusch [1], they found that the accented words are longer than the unaccented counterparts. Yet, they further indicated that the duration of the unaccented syllable on the left of the accented syllable increases, in addition to the accented syllable and the unaccented syllable on the right of the accented one. In terms of the size of accent-related lengthening, they also found that the size of the lengthening effect of prominence is asymmetrical. In other words, the unaccented syllable right to the accented one is lengthened much more than the unaccented syllable left to the accented one.

With regard to phrase boundary, it was also shown that elements (e.g., syllables, words) are lengthened in the vicinity of a phrase boundary and the size of lengthening is restricted to some elements within a phrase. Klatt [3] examined the variation of vowel durations employing the speech data from a single male speaker's connected discourse consisting of 13 sentences (236 words) under various linguistic contexts. His results showed that lengthening is observed at the end of a phrase which often coincides with a juncture between major syntactic categories. He isolated phrase boundary position as one of the most important factors determining segmental durations but boundary-related lengthening is restricted to the rhyme of the preboundary syllable. In a later study, Wightman and his colleagues [4] also indicated the boundary-related lengthening. They further found that the domain of boundary-associated lengthening is restricted to the rhyme of the preboundary syllable. In their production study of Dutch, Cambier-Langeveld and his colleagues [5] revealed that phrase final lengthening is generally restricted to the final syllable of the preboundary word but when the final syllable contains a schwa, the effect of final lengthening extends to the penultimate syllable. The effect of final lengthening becomes smaller as a function of distance from a phrase boundary. Although they indicated that the

domain of final lengthening is larger than the preboundary syllable, they also claimed that the distribution of final lengthening is not important in perception and has no communicative contribution.

In addition to temporal adjustment by prosody, intensity profile varies depending on the prosodic structure. Phonetic variation in the patterns of overall intensity influences prosodic judgments. For instance, Fry [6, 7] and Lieberman [8] tested overall intensity as a relevant correlate of linguistic stress and its relative importance in stress judgments. Both found that overall intensity is a relevant acoustic correlate of stress. Yet, in terms of the relative importance, Lieberman [8] found that overall intensity is more reliable cue to linguistic stress, while Fry found syllable duration as a more important acoustic correlate of linguistic stress. Later, Heldner [9, 10] investigated the effects of focal accents on the phonetic modulations of overall intensity in Swedish. His findings indicated that the overall intensity is a reliable correlate of focal accents. A later study by Turk and Sawusch [1] also found the importance of loudness on the prediction of prominence. When looking at the relationship between manipulated duration and intensity of reiterant two-syllable words and a lexical stress, however, they found that durational information predicts prominence more consistently than loudness, and loudness alone does not contribute much to prominence prediction in English. More recently, Fant and his colleagues [11] also tested the relative importance of various acoustic measures including overall intensity in predicting prominence, revealing that syllable duration is a better predictor of prominence than overall intensity. Kochanski and his colleagues [12, 13] evaluated the relative importance of the acoustic cues to prominence classification. They found that both loudness and duration are better cues to prominence classification but loudness is a better predictor. Although there are many studies looking at the relationship between the phonetic patterns of overall intensity and prominence, there is no study which found any significant influence of the presence/ absence of a prosodic phrase on intensity measures, to my knowledge.

As discussed, in many prior studies, it was shown that prosody modifies the phonetic patterns of temporal and intensity profiles of speech. Along with prior studies, the current study investigates the phonetic modulations of prosody. However, the current study differs in several ways from prior laboratory studies. First of all, it should be noted that in the current study the definition of prominence is broader than that in many early

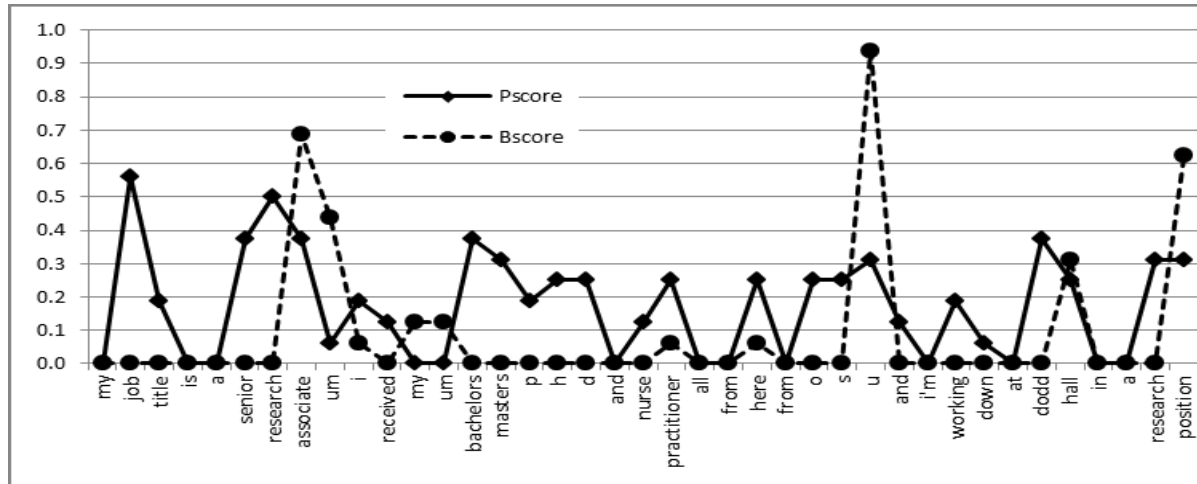


Figure 1. The distribution of probabilistic prominence scores (P-scores, solid line) and boundary scores (B-scores, dotted line) for each word in a sample utterance from the excerpts of Speaker 01

acoustic studies. In prior studies, the level of prominence of interest was lexical stress or focal accents. But the current study examines the phonetic modulations of any phrase- and sentence-level prominence. Second of all, most speech materials used in the prior studies are different from speech materials in conversational speech. That is, most prior studies employed controlled laboratory speech including isolated words or phrases, sentences, or read speech. Although findings from research with laboratory speech provide great insights about human speech communication, looking at the relationship of the prosodic structure with the patterns of duration and overall RMS intensity in laboratory speech may not directly reflect that in natural conversational speech. Third of all but related to the second point, in most prior production studies, the locations of prosodic prominence and phrase boundary are artificially manipulated by using punctuation marks, underlining, italics or boldface, which limits the prosodic structure among all other possible prosodic structures, or in perception studies, highly trained expert listeners mark prosodic prominence and phrase boundary based on their expert linguistic knowledge with visual as well as auditory aids. In other words, prosodic structures examined in the prior studies are not the same as those in naturally produced or perceived speech. Fourth of all, most prior studies examined the relationship between the prosodic structure and duration and intensity of syllables or words. Yet, taking into account that prosodic effects on duration are not equal within a word, it is important to understand the effects of prosodic prominence and phrase boundary on the subsyllabic structure of words.

The current study examines the relationship of perceived prosodic structures with the durations and the overall RMS

intensities of monosyllabic CVC words. More specifically, this study investigates how the patterns of durations and overall RMS intensities of subsyllabic components of monosyllabic CVC words are related to perceived prosodic prominence and phrase boundary, employing the speech excerpts from the Buckeye corpus of spontaneous, conversational speech of American English [14]. Building on findings from prior studies, it is hypothesized that (1) the patterns of the durations and overall RMS intensities of monosyllabic CVC words would be closely related to the presence/ absence of perceived prosodic prominence and phrase boundary if speakers signal the prosodic structures through phonetic modulations and listeners are sensitive to them and (2) the acoustic parameters would phonetically be enhanced if prosodic prominence and phrase boundary play important roles in speech communication, but (3) the phonetic patterns associated to prosodic prominence would be different from phrase boundary associated phonetic patterns if prosodic prominence highlights a word or a phrase within a discourse while prosodic phrase boundary demarcates a juncture between words within a long stretch of words.

## 2. Materials

The current study employed a total of 54 short speech excerpts (11-55s long each), two of which were extracted from each of 27 speaker's 30- to 60-minute interview in the Buckeye corpus of spontaneous conversational speech of American English [14]. The speech excerpts were carefully selected to contain the equal number of speech excerpts from the beginning, middle, and end part of interviews and to minimize

the number of disfluencies including silent pauses, cut-offs, and repetitions. Yet, it is impossible to completely remove all the disfluencies and in a few speech excerpts, some cut-offs and repetitions were included, though cut-offs were discarded in the analysis. After extracting the speech excerpts, I normalized the loudness by dividing the mean RMS intensity of each sound file by the maximum mean RMS intensity and then scaling the maximum peak value to 1, in order to make the loudness of the sound files approximately equal. This step is essential because some speech excerpts are much louder than others due to recording environments or differences in interviewees' intrinsic voice.

Along with a set of 54 short speech excerpts, the corresponding orthographic transcripts were also prepared in a separate paper. In the orthographic transcripts, words were separated by a space, all punctuation marks were removed and no words were capitalized. Then, a set of the extracted speech excerpts were divided into two blocks: a half for prosodic prominence and the other half for prosodic phrase boundary. Within each block, the sound files were randomized and the orthographic transcripts were also prepared in the same order as the speech excerpts in order for the listeners to mark prosodic prominence and boundary. The transcripts were also prepared in the same order of the speech files.

### 3. Participants

97 college students were recruited from undergraduate Linguistic courses at the University of Illinois at Urbana-Champaign. They are so-called "ordinary" (untrained, non-expert) listeners because they were never trained in terms of phonology and phonetics of prosody annotation. They were all native speakers of American English and confirmed no known hearing disorders. The participants who reported that they lived abroad more than 3 years before puberty were excluded in data analysis. The bilingual participants were also excluded in the data analysis. All the participants reported never to participate in any similar prosody annotation experiments.

### 4. Rapid Prosody Transcription (RPT)

The current study employed RPT to collect prosody annotation data from a large number of ordinary listeners, which is a newly developed in the Prosody-ASR group at University of Illinois at Urbana-Champaign. The detailed information about

RPT was reported in Mo [16, 17]. The main characteristics of RPT are as follows. First, RPT is time efficient. In RPT, prosody annotation is instantaneously done while the listeners are listening to speech excerpts in real time. Second, it allows to obtain prosody annotations from a group of untrained, so-called "ordinary" listeners instead of a few highly trained expert listeners. Third, in RPT, prosody annotation is solely based on the auditory impression of speech, not aided by visual display of speech waveforms. Therefore, RPT better reflects prosody perception in everyday speech communication.

The RPT data collection was done at a computer lab in the basement of the Foreign Language Building at University of Illinois at Urbana-Champaign. During RPT, the listeners were seated in front of a desktop computer equipped with a headphone. The sound files were uploaded at each computer and the corresponding transcripts were prepared beforehand.

Before the first transcription session, they were provided a 5-min introduction with simple 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.

After completing the language background survey, they were also provided a pair of speech excerpts for practice before each session. During the transcription tasks, the subjects marked the locations of prominence and boundary on the transcripts, while listening to the speech excerpts in the chronological order. In one transcription session, however, the subjects marked only one prosodic feature for one speech excerpt. One week later, the other prosodic feature was marked on the same speech excerpt.

The order of prosody annotation was balanced as follows. First, the sound files were randomly blocked for prominence and boundary annotation. Second, within each block, the sound files were randomized. Third, the order of prosody annotation was also balanced as described in the following: half of the listeners marked prosodic prominence first and then prosodic phrase boundary and the other half did in the reverse order. Fourth, a listener who marked prosodic prominence first and

then boundary in the first transcription session had to mark prosodic boundary first, and then prominence in the next session.

## 5. Reliability Test

After collecting prosodic transcriptions, I pooled prosody annotations from all the listeners and each word in the speech excerpts were assigned a probabilistic prominence (P-score) and boundary (B-score) scores depending on the number of listeners who hear the word as prominent or as followed by a boundary as in Figure 1. For example, if all the subjects marked a word as prominent, then its P-score is 1. If a half of the listeners marked a word as followed by a boundary, its B-score is 0.5. In Figure 1, no listener heard the first word, 'my', as prominent or as followed by a prosodic phrase boundary. As described in [16, 17], the reliability of prosody annotation was evaluated by using Fleiss'  $\kappa$  multi-raters agreement scores. It was confirmed that ordinary listeners' prosody annotation is statistically consistent and reliable across listeners.

## 6. Acoustic measurements

### 6.1. Duration measurements

Using the word and phone transcriptions provided in the Buckeye corpus, 862 monosyllabic CVC words were extracted from the speech excerpts. The words were not selected if phone boundaries are not clear (e.g., words containing a stop before and after a pause). Then the durations of the onset, nucleus and coda of each monosyllabic CVC word were measured and the relative durations of each subsyllabic component within the word were also calculated. For example, if the duration of a word is 300ms and the duration of the nucleus is 150ms, the relative duration of the nucleus is 0.5. In the current study, variations in duration associated with phone types were not considered due to the sparseness of data.

### 6.2. Intensity measurements

After extracting the CVC monosyllabic words, the average RMS intensity from the onset, nucleus, and coda of each word were automatically measured by using a Praat script. The extracted RMS intensities were then normalized within a speaker across phones within each subsyllabic position. In other words, the RMS intensities of the onset, the nucleus, and the coda phones were normalized.

## 7. Results

### 7.1 Spearman's non-parametric correlation analyses

In order to examine how prosody influences the internal phonetic structure of the monosyllabic CVC words, the relationship between the prosodic structures and the acoustic measures of each subsyllabic component was evaluated using Spearman's non-parametric correlation analyses. More specifically, the correlations between probabilistic prominence scores (P-scores) and the acoustic measures of subsyllabic components including the raw phone duration, the relative phone duration, and the RMS intensity and between probabilistic boundary scores (B-scores) and the same acoustic measures from the same subsyllabic components were statistically evaluated.

Among many correlation analyses, Spearman's non-parametric correlation analysis was chosen because P- and B-scores are highly negatively skewed (many non-prominent words and few prominent words and many non-boundary words and few boundary words) and Spearman's correlation analysis does not require the variables (P-scores, B-scores, and the acoustic measures) to be normally distributed.

Table 1. The summary of Spearman's non-parametric correlation analyses between P- and B-scores and raw word duration

	P-score	B-score
N. of tokens	862	
Word duration	.387*	.491*

\* indicates that the correlation is significant with the 99% confidence interval.

Table 1 summarizes the results from Spearman's non-parametric correlation analyses of P- and B-scores with the duration of the monosyllabic CVC words. As shown in Table 1, word duration is significantly correlated with both P- and B-scores in the positive direction. This tells us that the duration of the monosyllabic CVC word is longer as perceived as prominent and as perceived as followed by a phrase boundary. However, the correlation coefficient ( $\rho$ ) of P-scores with word duration (0.387) is lower than that between B-scores and word duration (0.491), suggesting that word duration is more closely correlated with B-scores than P-scores.

Table 2. The summary of Spearman’s non-parametric correlation analyses of P-scores with the durations and the overall RMS intensities of the subsyllabic components (Onset, Nucleus, and Coda)

	Onset	Nucleus	Coda
Duration	.252*	.378*	.125*
Overall RMS Intensity	.015	.302*	.079*

\* indicates that the correlation is significant with the 99% confidence interval.

The Spearman’s non-parametric correlation coefficients ( $\rho$ ) of P-scores and B-scores with the duration and the overall RMS intensity of each subsyllabic component (onset, nucleus, and coda) are summarized in separate tables (Table 2 and 3). As seen in Table 2, the duration of each subsyllabic component of the monosyllabic CVC words is positively correlated with P-scores. The correlations of the duration of the onset, nucleus, and coda with P-scores are all significant with the 99% confidence interval. Looking closely, the correlation coefficient of the nucleus duration ( $\rho = 0.378$ ) with P-scores is the largest, followed by that of the onset duration ( $\rho = 0.252$ ) and of the coda duration ( $\rho = 0.125$ ) in order. This suggests that the duration of the nucleus is most strongly correlated with prominence perception and the coda duration is the least strongly correlated with prominence perception.

On the other hand, the overall RMS intensities show significant positive correlations with P-scores of the nucleus and the coda but the onset does not show a significant correlation with P-scores and its overall RMS intensities. Compared to the correlation coefficients between P-scores and the overall RMS intensities of the subsyllabic components, the correlation coefficient of the overall RMS intensity of the nucleus is the largest ( $\rho = 0.302$ ). The other coefficients are smaller than 0.100. This shows that the overall RMS intensity of the nucleus is most strongly correlated with prominence perception and the onset intensity does not show any systematic correlation with prominence perception at all. When comparing the correlation coefficients between P-scores and the subsyllabic durations with those with the subsyllabic overall RMS intensities, it is shown that the duration and the overall RMS intensity of the nucleus show much higher correlation coefficients with P-scores. This suggests that the duration and the overall RMS intensity of the nucleus of the monosyllabic CVC words both are most strongly correlated with prominence perception.

The following table (Table 3) summarizes the results of the Spearman’s non-parametric correlation analyses between B-scores and the durations and the overall RMS intensities of the subsyllabic components of the monosyllabic CVC words. Similar to P-scores, B-scores are significantly correlated with the durations of the subsyllabic components of the monosyllabic CVC words. Looking closely, however, the results show that the correlation coefficient between B-scores and the nucleus durations is the largest, followed by that of the coda duration and of the onset duration in order, suggesting that boundary perception is more closely correlated with the nucleus duration and the coda duration than the onset duration.

As for the correlation between B-scores and the overall RMS intensities of the subsyllabic components of the monosyllabic CVC words, it is shown that the overall RMS intensity of the nucleus ( $\rho = .122$ ) and of the coda ( $\rho = -.157$ ) is negatively correlated with B-scores and there is no systematic correlation between B-scores and the onset intensity. The results from Spearman’s non-parametric correlation analyses between

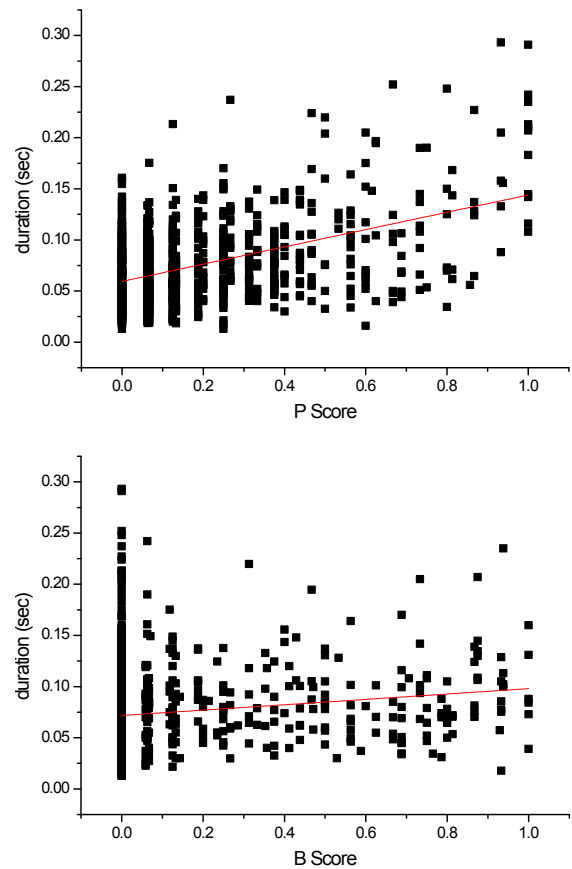


Figure 2. The scatter plots and regression lines of P- and B-scores with onset durations (in seconds)  
As for the correlation between B-scores and the overall RMS

B-scores and the overall intensities of the subsyllabic components demonstrate that the degree of correlation decreases as the distance from a prosodic phrase boundary increases.

Table 3. The summary of Spearman's non-parametric correlation analyses of B-scores with the durations and the overall RMS intensities of the subsyllabic components (Onset, Nucleus, and Coda)

	Onset	Nucleus	Coda
Duration	.156*	.477*	.410*
Overall RMS Intensity	-.050	-.122*	-.157*

\* indicates that the correlation is significant with the 99% confidence interval.

## 7.2 Linear regression analyses

Figure 2 shows the scatterplots with a regression line between P- and B-scores and onset durations, in which the slope of each regression line demonstrates the relationship between changes in onset duration and the listeners' perception of prosody. The slope of P-scores and onset durations ( $a=85\text{ms}$ ) is steeper than that of B-scores and onset durations ( $a=26\text{ms}$ ), showing that onset durations increase by 8.5ms when P-scores increase by 10%, and by 2.6ms when B-scores increase by 10%. Figure 3 and 4 show the scatterplots with regression lines of nucleus durations (Fig. 3) and coda durations (Fig. 4) with perceived prosody scores. As seen in Figure 2 and 3, prosody scores increase as the durations of nucleus and coda increase. Specifically, the slopes of P-scores and of B-scores with nucleus durations are 140ms and 111ms and the slopes of P-scores and B-scores with coda durations are 47ms and 100ms.

The scatter plots and regression lines of P- and B-scores, and the overall intensities of the nuclei and the codas are plotted in Figure 5 and 6 because the overall intensities of the onsets show significant correlations with neither P- nor B-scores. The slopes of the overall RMS intensities of the nuclei with P-scores and with B-scores are 2.65dB and -3.47dB, respectively; the slopes of the overall RMS intensities of the codas with P- and B-scores are 0.06dB and -5.72dB, respectively.

## 8. Discussion

The present study examined whether and how prosodic prominence and boundary as determined by ordinary listeners is related to the phonetic realizations of the durations and overall

monosyllabic CVC words in everyday conversational speech of American English. More specifically, the present study examined (1) whether the effects of prosodic prominence and phrase boundary are uniform on each subsyllabic component including the onset, nucleus, and coda, and (2) whether the effects are similar on the phonetic realizations of two different acoustic features, namely, duration and intensity.

The findings from Spearman's non-parametric correlation analyses showed that the patterns of the phonetic characteristics of the subsyllabic components of the monosyllabic CVC words are influenced by the presence or absence of prosodic prominence and phrase boundary. This study further showed that the temporal effects of prosody extend to all the subsyllabic components of the words. Yet, this study also showed that the temporal effects of prosodic prominence are not the same as those of prosodic phrase boundary and the correlation strength of each subcomponent varies. As for prosodic prominence, the nucleus duration is most strongly correlated with prosodic prominence, followed by the duration of the onset and of the coda, in order. Similar to prosodic prominence, the nucleus duration shows the strongest correlation ( $\rho=.477$ ) with phrase boundary. Different from prosodic prominence, prosodic phrase boundary shows the weakest correlation ( $\rho=.156$ ) with the onset duration and the strength of correlation of phrase boundary ( $\rho=.410$ ) with the coda duration is quite comparable to the correlation with the coda duration.

Changes in the patterns of the overall RMS intensity were shown to be correlated with the prosodic structure. Yet, the effects of prosodic prominence on the overall intensity of the word are not the same as those of prosodic phrase boundary and the prosodic effects on the overall intensities of the onset, nucleus, and coda are not uniform. Firstly, the present study showed that only the overall RMS intensities of the nucleus are positively correlated with prosodic prominence and the intensities of the onset and the coda do not show any significant correlation with prosodic prominence, suggesting that under prosodic prominence, the core element of the word becomes louder. On the other hand, prosodic phrase boundaries were shown to be negatively correlated with the overall intensities of the nucleus and the coda but not of the onset, suggesting that before a prosodic phrase boundary, the rhyme of the word becomes softer. The finding that the overall intensities of the coda are more strongly correlated with those of the nucleus suggested that when the distance from the phrase boundary increases, the overall intensities are less likely to be

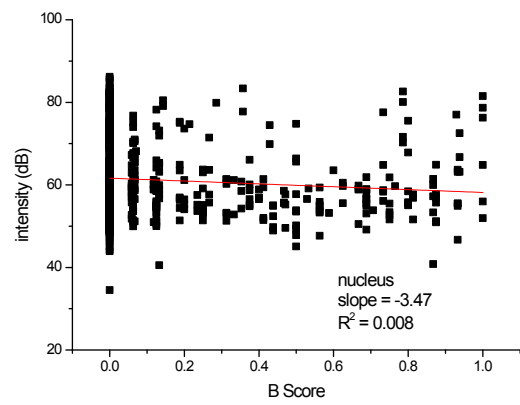
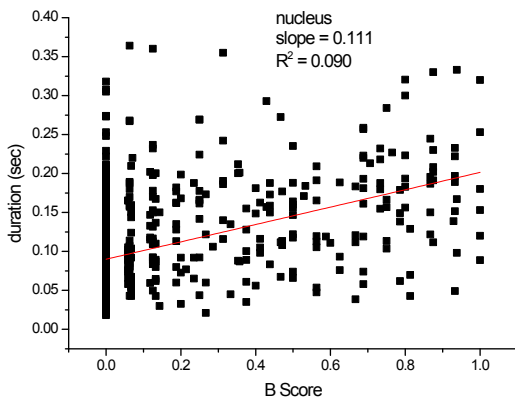
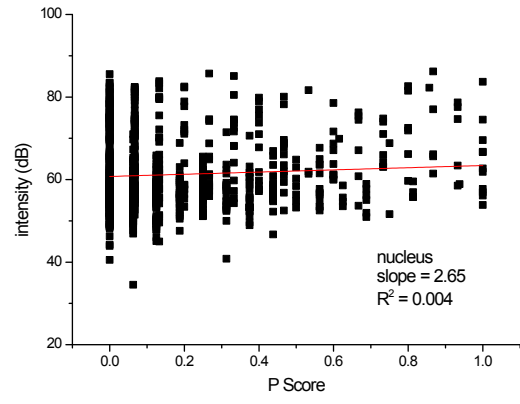
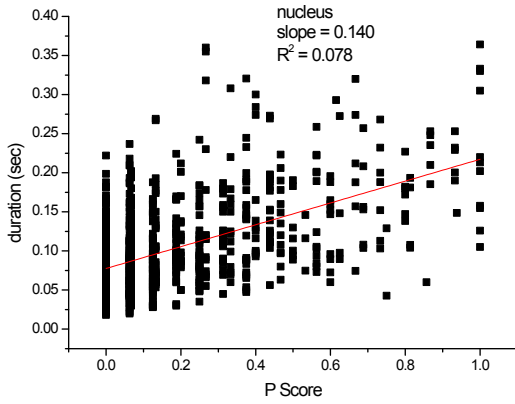


Figure 3. The scatter plots and regression lines of P- and B-scores with nucleus durations

Figure 5. The scatter plots and regression lines of P- and B-scores with nucleus intensities

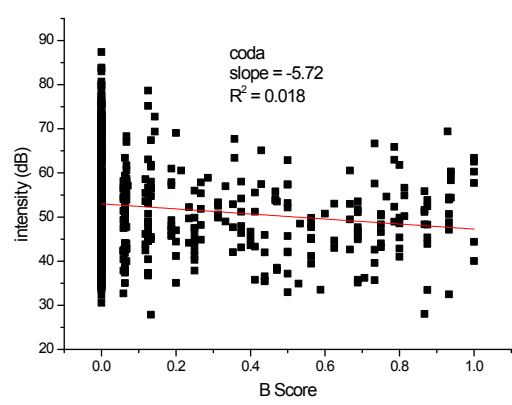
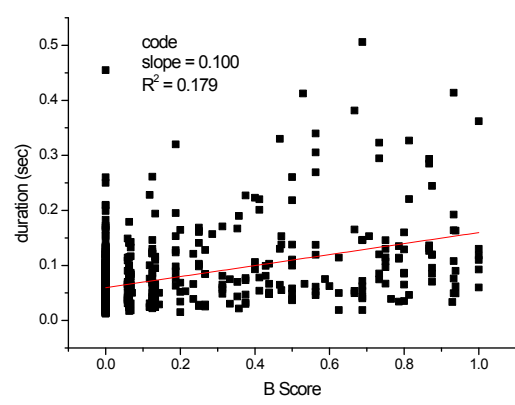
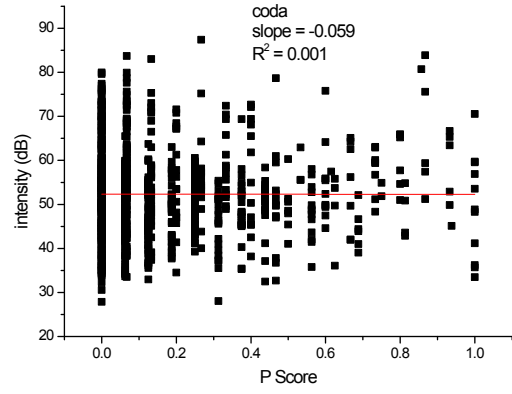
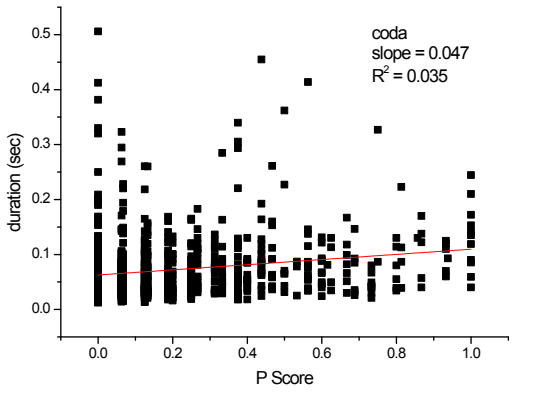


Figure 4. The scatter plots and regression lines of P- and B-scores with coda durations

Figure 6. The scatter plots and regression lines of P- and B-scores with overall RMS intensities of codas



correlated with listeners' perception of the boundary.

In addition to correlation analyses, the current study examined the slope of the linear regression lines between prosody scores and the acoustic measures (duration and overall RMS intensity) of the subsyllabic components. The findings from the linear regression analyses demonstrated that the amount of lengthening effects varies as a function of prosodic prominence and phrase boundary. The durational difference of the nucleus between the word perceived as prominent and the non-prominent word is the largest (140ms), followed by that of the onset (85ms) and of the coda (47ms), in order. The durational difference of the nucleus between the phrase-final and the phrase-medial word is the largest (111ms), that of the coda (100ms) the second, and that of the onset (26ms) the third. This study also showed that the amount of prosodic effects on the overall RMS intensity varies depending on the locations of prosodic prominence and phrase boundary. The onset of the word perceived as prominent is louder than the non-prominent word by 2.65dB and the nucleus and the coda of the phrase-medial word are louder than those of the phrase-final word by 3.47dB and by 5.72 dB, respectively.

Taking the results from Spearman's non-parametric correlation and linear regression analyses together, the findings from the current study can be summarized as in the following. Firstly, not only the nucleus duration is the most likely to be consistently increased but the amount of durational increase tends to be the largest as well, and the lengthening effects of the onset and the coda duration are less consistent and the amount of lengthening of the onset and the coda is relatively small. Secondly, similar to the lengthening effects of prosodic prominence, the consistency and the amount of lengthening of the nucleus duration is the greatest, followed by the coda and the onset. Thirdly, both prosodic prominence and phrase boundary increase the duration of each subsyllabic component as well as the whole word. Yet, between prosodic prominence and phrase boundary, boundary-related lengthening effects are more consistent and greater.

Fourthly, similar to the lengthening effects, prosodic prominence increases the overall intensity of the nucleus in the most consistent pattern and the amount of the increase of the overall intensity of the nucleus is the largest. The overall intensity of the onset and of the coda does not show any consistent correlation pattern at all. Fifthly, prosodic phrase boundary influences the patterns of the overall intensity of the subsyllabic components. The overall intensity shows the

consistent decrease as the distance from a phrase boundary increases. The consistency and the amount of the decrease of the overall intensity of the coda is the greatest, followed by the nucleus. The onset intensity does not show a consistent correlation with the presence/ absence of the phrase boundary. Sixthly, both prosodic prominence and phrase boundary influence the phonetic patterns of the overall RMS intensity of the monosyllabic CVC word but the direction of prominence effects and boundary effects are opposite. That is to say, the prosodic phrase boundary decreases the overall RMS intensity of the word while the prosodic prominence increases it. Lastly, prosody-related durational effects are more consistent and their magnitudes are greater than those of prosodic effects on the overall intensity.

## 9. Conclusion

The present study demonstrates that prosody modulates the phonetic patterns of the duration and the overall RMS intensity of the subsyllabic components of the monosyllabic CVC words in everyday conversational speech of American English. Under prosodic prominence, both duration and intensity consistently increase; before a phrase boundary, intensity decreases while duration increases. Between prosodic prominence and phrase boundary, the variation induced from prosodic phrase boundary is generally larger than the prominence-induced variation. Between duration and intensity, the durational variation is larger than the intensity variation. The phonetic characteristics of the nucleus, the most salient element within the syllable, are most strongly influenced by both prominence and boundary.

## References

- [1] Turk, A. E. & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, Vol. 99, No. 6, 3782-3790.
- [2] Cambier-Langeveld, T. & Turk, A. E. (1999). A cross-linguistic study of accentual lengthening: Dutch vs. English. *Journal of Phonetics*, Vol. 27, 255-280.
- [3] Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, Vol. 3, 129-140.
- [4] 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*, Vol. 91, No. 3, 1707-1717.
- [5] Cambier-Langeveld, T., Nespors, M., & van Heuven, V. J. (1997). The domain of final lengthening in production and perception in Dutch. *The Proceedings of the 5<sup>th</sup> European Conference on Speech Communication and Technology*, (September 22-25, Rhodes, Greece).
- [6] Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, Vol. 27, 1765-8.
- [7] Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, Vol. 1, No. 3, 126-150.
- [8] Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, Vol. 32, No. 4, 451-454.
- [9] Heldner, M. (2001). Spectral emphasis as a perceptual cue to prominence. *TMH-QPSR*, Vol. 42, No. 1, 51-57.
- [10] Heldner, M. (2003). On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, Vol. 31, No. 1, 39-62.
- [11] Fant, G., Kruckenberg, A., Liljencrants, J., & Hertegard, S. (2000). Acoustic-phonetic studies of prominence in Swedish. *TMH-QPSR*, Vol. 41, No. 2-3, 1-51.
- [12] Kochanski, G., Grabe, J., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, Vol. 118, No. 2, 1038-1054.
- [13] Kochanski, G. (2006). Prosody beyond fundamental frequency. *Methods in Empirical Prosody Research* (Sudhoff, S., Lenertov'a, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., & Schlieber, J. Eds.), 1-43. New York: De Gruyter.
- [14] Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication*, Vol. 45, 89-95.
- [15] Mo, Y. (2010). *Prosody production and perception with conversational speech*. Ph. D. Dissertation, Department of Linguistics, University of Illinois, Urbana-Champaign.
- [16] Mo, Y. (2011). Acoustic correlates of prosodic prominence in conversational speech of American English, as perceived by ordinary listeners. *Journal of the Korean Society of Speech Sciences*, Vol. 3, No. 3, 19-26.

• **Mo, Yoonsook**

Department of English Education  
Chonnam National University  
Email: myoons@jnu.ac.kr