ON THE EDGE: ACOUSTIC CUES TO LAYERED PROSODIC DOMAINS

Tae-Jin Yoon, Jennifer Cole, & Mark Hasegawa-Johnson

University of Illinois at Urbana-Champaign, USA
{tyoon, jscole, jhasegaw}@uiuc.edu

ABSTRACT

Prosodic structure encodes the grouping of words into hierarchically layered prosodic constituents, including the prosodic word, intermediate phrase (ip) and intonational phrase (IP). This paper investigates the phonetic encoding of prosodic structure from a corpus of scripted broadcast news speech in American English through analysis of the acoustic correlates of prosodic boundary and their interaction with phrasal stress (pitch-accent) at three levels of prosodic structure: Word, ip, and IP. Evidence for acoustic effects of prosodic boundary is shown in measures of duration local to the domain-final rime. These findings provide strong evidence for prosodic theory, showing acoustic correlates of a 3-way distinction in boundary level.

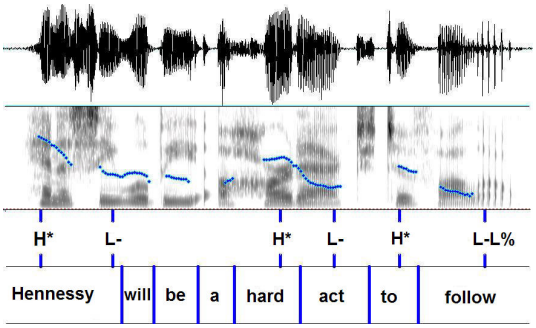
Keywords: prosodic hierarchy, levels of prosodic boundaries, normalized duration

1. INTRODUCTION

Prosodic structure encodes the grouping of words into hierarchically layered prosodic constituents, including for English the prosodic word, intermediate phrase (ip) and intonational phrase (IP), as illustrated in Figure 1 (cf. [1][4][5]). Given the hierarchical organization, we expect to find audible acoustic correlates of prosodic boundaries at each of these levels, but especially at phrasal junctures of ip and IP, to guide the listener in chunking the speech signal. Acoustical cues to prosodic boundaries are observed in the lengthening of segments in the preboundary syllable rime, with greater effects of lengthening at successively higher levels of prosodic domains [9]. A second dimension of prosodic structure is the encoding of prominence, which also gives rise to lengthening effects in the prominent syllable (stressed or accented) and syllables to the right of the prominent syllable [7].

Given two distinctive sources of lengthening, the question arises whether lengthening on its own can serve as a cue to either prosodic context, a question which is not addressed in any prior work. An additional question is whether the acoustic correlates of prominence and juncture are differentiated for syllables that are doubly marked (i.e., both accented and phrase-final). This paper investigates the phonetic encoding of prosodic structure through a study of duration as an acoustic correlate of prosodic boundary and the interaction between boundary and accent effects at three levels of prosodic structure: Word, ip, and IP. Guided by earlier evidence that boundary cues are local [9], evidence for acoustic effects of prosodic boundary is considered in measures of duration local to the domain-final rime.

Figure 1: An example that illustrates low-toned ip (L-) and low-toned IP (L-L%), taken from the Boston University Radio Speech Corpus [6].



2. ACOUSTIC CUES FOR PROSODIC BOUNDARY

Silent pause and pre-boundary lengthening are known to be acoustic correlates of prosodic boundary in English [9]. While silent pause is neither a necessary nor sufficient boundary cue, the potential value of lengthening as a boundary cue is questionable given that there are two distinct sources of lengthening: boundary and accent. The possibly confounding interaction between boundary and accent lengthening motivates the current study.

2.1. Silent pause

There is a strong correlation between the presence of a pause and the perception of a prosodic boundary; however, the perception of prosodic boundary does not depend on the occurrence of silent pause. Table 1 presents the correlation between the presence or absence of silent pause and the presence or absence of phrasal boundary (ip or IP) in a prosodic transcription of speech from the Boston University

Radio Speech corpus [6]. Phrasal boundaries are signaled by the presence of silent pause about 40% of the time, whereas the remaining 60% of boundary labels occur with no silent pause.

Table 1: Correlation of the presence/absence of silent pause and the presence/absence of phrasal boundary (ip or IP) in the Boston University Radio Speech Corpus

	Phrasal Boundary	No Boundary
Silence	984 (40.6%)	67 (0.8%)
No silence	1439 (59.4%)	8056 (99.2%)

2.2. Pre-boundary & Accentual Lengthening

Given the somewhat weak role of silent pause as a cue to prosodic boundary, it is likely that phrasal boundaries are also signaled by other cues, such as the pre-boundary lengthening that lengthens the final rime in a prosodic domain, reflecting a reduction of the articulation rate at the end of the phrase. Yet, as noted above, lengthening is also an effect of accent that encodes phrasal stress. For English, durational effects of accent extend beyond the stressed syllable nucleus or rime [2][7][8] (cf. [3]). The domain of accentual lengthening begins with the onset of the primary stressed syllable, and extends rightward until the end of the word.

3. EXPERIMENT

3.1. Corpus

We investigate prosodic lengthening effects in a acoustic analysis of speech from the Boston University Radio Speech Corpus ([6]). The speech materials are taken from read paragraphs and consist of about 10,000 words produced by 5 speakers. The prosodic labels were transcribed (by other researchers) based on auditory impression and visual inspection of the speech signal, using the ToBI standard [1]. According to the study of Inter-transcriber agreement reported in [6], when labelers agree that a phrasal tone was present they agreed on the type of the phrasal tone (i.e., ip or IP) 91% of the time. Our analysis is based on the agreed-upon labels.

3.2. Normalized duration

Duration measures are taken for each segment following segmentation and phone labeling of the speech signal. Segmentation and labeling is automated by doing a forced alignment of the speech signal to a phone string. The phone string is taken from the dictionary encoding of each word, and forced alignment is done using the HTK Hidden Markov Model Toolkit. Normalized duration measure is calculated based on observed segment durations, using the normalization method employed in [9]. The normalized duration of a segment is measured as the number of standard deviation units (σ) from the mean duration (μ) of that segment, as observed over the entire corpus. The phone-based normalization formula is given in (1):

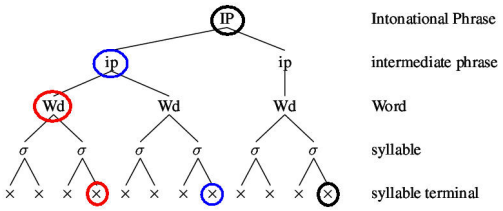
$$\bar{d}_i^k = \frac{x_i^k - \mu_i^k}{\sigma^k} \quad (1)$$

where x_i^k is the observed duration of token x_i, belonging to vowel phone class k.

3.3. Measurement domain

Duration measures are taken from the nucleus segment(s) of syllables in word-final position in three prosodic contexts, as illustrated in Figure 2: (1) phrase-medial position, (2) intermediate-phrase final position, and (3) intonational-phrase final position.

Figure 2: Measurement domain.



4. RESULTS

Lexical stress (i.e., word-level prominence) may also influence segment duration, so separate analyses are performed based on the location of lexical stress relative to the target word-final syllable, as shown in the schematic diagram in Figure 3. As the present study is concerned with the interaction of final lengthening with accentual lengthening, the data are restricted to those syllables that are final in a prosodic domain, and in the domain of accentual lengthening (i.e., in the accented stress foot). Final syllables in words where lexical stress falls on a syllable that precedes the penultimate syllable are not eligible for accentual lengthening, based on the findings of [7].

Figure 3: Schematic diagram of the two locations of word-level stress for words in the present study. Duration measurements are taken from the word-final syllable at the prosodic boundary.

(a)	(b)
$[\ldots \sigma\ \sigma\ \acute{\sigma}\ ]_{\mathrm{BND}}$	$[\ldots \sigma\ \acute{\sigma}\ \sigma\ ]_{\mathrm{BND}}$

Thus the present study is limited to analysis of duration from word-final syllables in words with lexical stress on the final or penultimate syllables, as illustrated schematically in Figure 3.

Prosodic effects on normalized duration measures are tested using ANOVA with the independent factors of Boundary (Word, ip, IP) and Accent (Accented, Unaccented), and with separate ANOVAs for the two conditions of lexical stress location shown in Figure 3.

4.1. Lexical stress on the pre-boundary syllable

The results of duration measures for the condition with lexical stress on the final syllable (Figure 3a) are presented first. Table 2 shows the number of tokens available for analysis under the two conditions of pitch accent and three boundary levels, from words where stress falls on the final syllable (the 3rd column), and from words where stress falls on the penult (the last column).

Table 2: Frequency table (penult stress and final stress)

	Boundary	No. of tokens (Final stress)	No. of tokens (Penult stress)
No Pitch Accent	ip	136	53
	IP	272	171
	Word	3864	288
Pitch Accent	ip	283	256
	IP	347	364
	Word	1424	1512
	Total	6424	2644

Figure 4a shows the effects of prosodic boundary on final nucleus duration. In both accented and unaccented contexts, the normalized duration increases with the level of the prosodic boundary in the hierarchy (Word $<$ ip $<$ IP) (Accented: $F(2, 2051)=341.8$, $p<.001$; Unaccented: $F(2, 4269)=345.1$, $p<.001$). This demonstrates a three-way distinction of prosodic boundaries.

Figure 4b shows the effect of pitch accent on the nucleus duration. The normalized duration is significantly longer when the nucleus is accented than when the nucleus is unaccented, for all levels of prosodic boundary (Word: $F(1, 5286)=449.5$, $p<.001$; ip: $F(1, 417)=13.2$, $p<.001$; IP: $F(1, 617)=58.5$, $p<.001$). This finding demonstrates an accumulative effect of lengthening due to accent and prosodic boundary.

4.2. Lexical Stress on the penultimate syllable

Next we present results of duration measures for the condition in Figure 3b, where lexical stress occurs on the penultimate syllable. Bear in mind that the duration measurements are taken from the final nucleus, and not from the penultimate stressed syllable. See Table 2 (the last column) for the number of words where stress falls on the penultimate syllable.

Figure 4: (1) Effect of prosodic boundary on final nucleus duration (final stress), and (2) effect of pitch accent on final nucleus duration (final stress). The asterisk * indicates that the difference between the adjacent mean values on the plot is statistically significant.

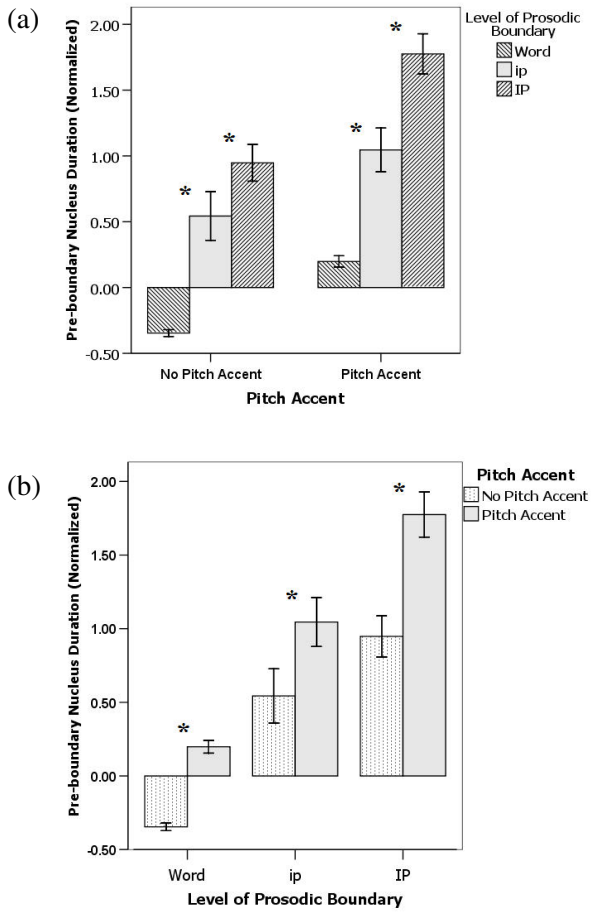
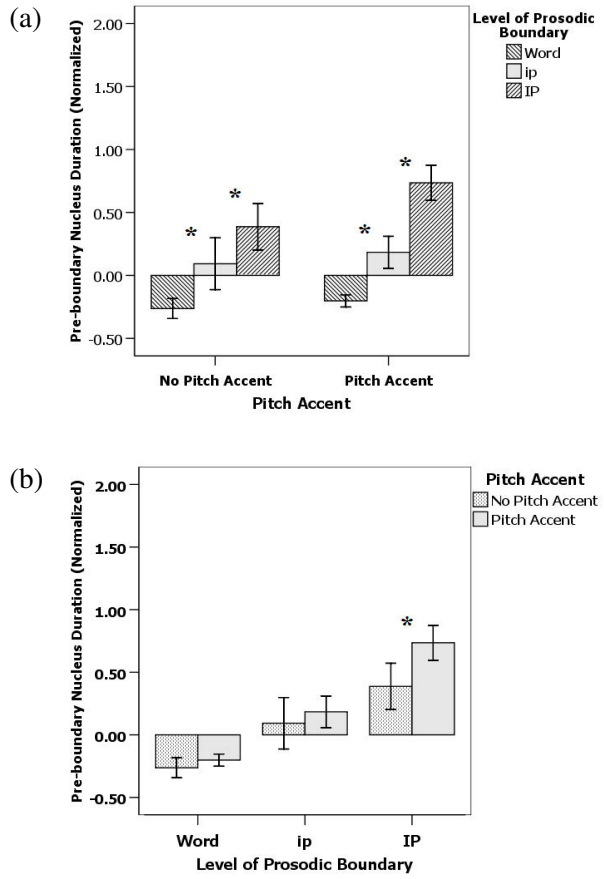


Figure 5a illustrates the effect of prosodic boundary on final nucleus duration. The normalized duration of the final post-stress syllable nuclei in Figure 5a are shorter than the normalized durations from the final stressed syllables in Figure 4a. Nevertheless, Figure 5a shows the same three-way contrast in duration according to the level of prosodic boundary as in Figure 4a (Accented: $F(2, 2129)=121.6$, $p<.001$; Unaccented: $F(2, 509)=27.8$, $p<.001$).

Figure 5b illustrates the effect of accent-induced lengthening on the final nucleus duration. Based on the work by Turk and her colleagues, we hy-

pothesized that accent would cause lengthening of the post-accented syllable in each of the prosodic boundary conditions. The results show that this hypothesis is confirmed only for final post-accented syllables in IP-final position, but not at the two lower levels of prosodic boundary (Word: $F(1, 1798)$=1.5, p>.2; ip: $F(1, 307)$=.37, p>.5; IP: $F(1, 533)$=8.1, p<.005).

Figure 5: (a) Effect of prosodic boundary on final nucleus duration, and (b) effect of accent-induced lengthening on final nucleus duration (penult stress).



In summary, there are significant and increasing effects of final lengthening for the nucleus in the final syllable of the Word, ip and IP, supporting a 3-way distinction for word-final syllables according to the prosodic phrase context. As expected, pitch accent also induces lengthening of the accented syllable nucleus, but accentual lengthening effects on the post-accented, word-final syllable are observed only for syllables that are final in the IP, and not in final position of the ip or Word. The discrepancy between our findings and those of Turk and her colleagues may be due to differences in the focus conditions for the accents. In the radio news materials, most pitch accents mark broad focus (new information), with relatively few emphatic or contrastive focal accents. In the materials used by Turk & her colleagues [2][7][8], accents mark contrastive focus. Contrastive focal accents are observed to have larger pitch movements, and may also exhibit stronger effects of accentual lengthening.

5. CONCLUSION

In conclusion, we find strong evidence for lengthening effects conditioned by prosodic boundaries and by phrasal prominence. The boundary lengthening effects distinguish three levels of prosodic domains, and thus support a theory of prosodic structure that discriminates between levels of prosody phrasing, such as ip and IP, in addition to the prosodic word. We also demonstrate that lengthening effects due to accent and boundary are fully accumulative for final accented syllables, and partially accumulative for final post-accented syllables. Finally, our study provides evidence for local effects of prosodic domains in the syllable at the right edge.

ACKNOWLEDGMENTS

This work is supported by NSF (IIS-0414117). Statements in this paper reflect the opinions and conclusions of the authors, and are not necessarily endorsed by the NSF.

6. REFERENCES

[1] Beckman, M. E., Ayers, G. 1997. *Guidelines for ToBI labeling*. The Ohio State University.

[2] Cambier-Langeveld, T., Turk, A. E. 1999. A cross-linguistic study of accentual lengthening: Dutch vs. english. *Journal of Phonetics* 27, 255–280.

[3] Chen, Y. 2006. Durational adjustment under corrective focus in standard chinese. *Journal of Phonetics* 34, 176–201.

[4] Ladd, D. R. 1986. Intonational phrasing: The case of recursive prosodic structure. *Phonology* 3, 311–340.

[5] Ladd, D. R. 1996. *Intonational Phonology*. Cambridge, UK: Cambridge University Press.

[6] Ostendorf, M., Price, P., Shattuck-Hufnagel, S. 1995. *The Boston University Radio News Corpus*. ECS Department, Boston University.

[7] Turk, A. E., Sawusch, J. R. 1997. The domain of accentual lengthening in american english. *Journal of Phonetics* 25, 25–41.

[8] Turk, A. E., White, L. 1999. Structural influences on accentual lengthening in english. *Journal of Phonetics* 27, 171–206.

[9] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P. J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *J. Acoust. Soc. Am.* 91, 1707–1717.