

Levels of Prosodic Phrasing:

Acoustic evidence from read and spontaneous
speech corpora

Tae-Jin Yoon, Jennifer Cole & Heejin Kim

University of Illinois at Urbana-Champaign

The 2006 Annual Meeting of the Linguistic Society of America,
Albuquerque, NM., 5-8 January.

Two levels of prosodic phrasing

□ At least two levels of prosodic phrasing have been widely assumed in the description of prosodic structure offered by linguists and speech scientists.

- Pike (1945)
- Householder (1957)
- Trim (1959)
- O'Connor & Arnold (1973)
- Beckman & Pierrehumbert (1986)
- Ladd (1986)

"In a number of descriptions there are, in effect, two types of IP, which we might informally call big ones and little ones." (p. 315)

Beckman & Pierrehumbert (1986)

- There exists a level of phrasing between the prosodic word and the intonational phrase (IP), identified as the immediate phrase (ip) in English, and the accentual phrase (AP) in Japanese.
 - Big IPs are assumed to be set off by audible breaks
 - Little ips are assumed to have a nucleus accents (Ladd 1986).
-

Scant empirical evidence

- There is scant empirical evidence for the intermediate phrase in English, compared with the Accentual Phrase (AP) for Japanese (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988)
-

Previous Phonetic Studies

- Subsequent research provides evidence for differentiating ip from IP based on
 - Articulatory measures (Fougeron & Keating 1997),
 - Acoustic duration (Wightman et al. 1992), &
 - Perceptual judgment of voice quality (Redi & Shattuck-Hufnagel 2001).
-

Pitch & Intensity?

- Prosodic structure is generally defined in terms of categories that are instantiated by pitch, intensity, and duration.
 - However, there remains little evidence for distinction in phrase level in phrase level in terms of F0 or intensity, or their perceptual correlates in pitch and loudness.
-

Purpose of the paper

- Acoustic evidence for the distinction between ip and IP boundaries based on measurements of F0, intensity, duration and glottalization taken from the phrase-final syllable rime.
 - In particular, the distinction between low-toned intermediate phrase and low-toned intonational phrase, or L- vs. L-L% in the ToBI prosodic transcription system (Beckman & Ayers 1997).
-

Corpus

□ Switchboard Corpus

- A WS97 subset of the Switchboard corpus of informal telephone conversations (Godfrey *et al.* 1992).
- 180 files, 79 different speakers, & 1698 words.

□ Boston Radio News Corpus

- Scripted, read speech style of the professional radio announcers (Ostendorf *et al.* 1995).
 - Files from the 'labnews' portion of two speakers (F1A and F2B).
-

Distribution of L- and L-L% tokens: Switchboard

Boundary	Pitch Accent	Plain	Creak
L-	H*	106	3
	L*	7	2
	No PA	92	12
	Total	205	17
L-L%	H*	60	15
	L*	5	4
	No PA	22	11
	Total	87	30

Distribution of L- and L-L% tokens: Radio News

Bnd	PA	Speaker A		Speaker B	
		Plain	Creak	Plain	Creak
L-	H*	54	7	46	38
	L*	2	0	1	1
	No PA	19	5	10	3
	Total	75	12	57	42
L-L%	H*	43	85	55	136
	L*	1	2	2	12
	No PA	0	19	10	37
	Total	44	106	67	185

Measurements (I)

- **Measurement Domain:**
 - preboundary syllable rime
 - **Duration:**
 - Normalized durational difference between L- and L-L%
 - **F0**
 - Beginning F0, End F0, F0 drop & F0 slope
 - **Intensity**
 - Beginning Intensity, End Intensity, & F0 drop
 - **The number of occurrences of the creaky voice quality**
-

Measurements (II)

: F0 & dB

- **Beginning F0:** For preboundary syllables with a H* pitch accent, beginning F0 was measured at the accent peak. For non-pitch-accented syllables, beginning F0 was measured at the rime beginning.
 - **Beginning intensity:** Because maximum intensity in the rime is usually not reached until some point after the rime beginning, beginning intensity was measured at the point of peak intensity in the rime.
 - **End F0 and end intensity:** These measurements were taken at the end of the sonorant portion of the rime.
 - **F0 drop and intensity drop:** F0 drop is equal to end F0 minus beginning F0, and intensity drop is equal to end intensity minus beginning intensity. Bigger negative values indicate greater magnitude of drop.
 - **F0 slope:** This is the F0 drop divided by the duration of the interval from beginning F0 to end F0.
-

Overall Results

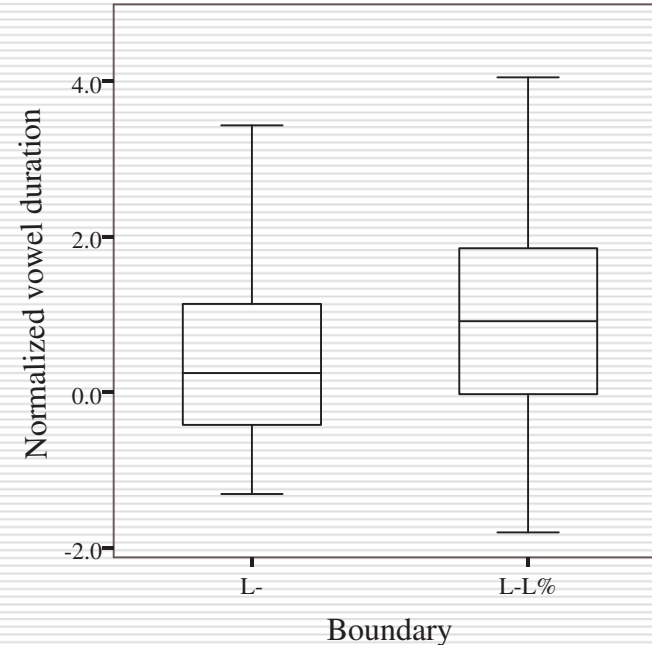
- ❑ Significant differences in F0 at the rime end, in peak intensity, and in the frequency of creak occurrence in the preboundary rime for the two boundary levels, L- and L-L% in both corpora.
 - ❑ An additional effect of boundary level on F0 drop and F0 slope in Radio News Corpus. The effect of boundary on intensity.
 - ❑ Variation of the effect of boundary on intensity depending on speaker and voice quality in Radio News Corpus.
-

Normalized nucleus duration

- Vowel Duration Normalization
 - $E(d_i^k) = (d_i^k - \mu^k) / \sigma^k$
 - Normalization based on each vowel type
 - The sig. greater duration at L-L% compared L- indicates that degree of preboundary lengthening differentiates levels of phrasing across speakers in both corpora.
-

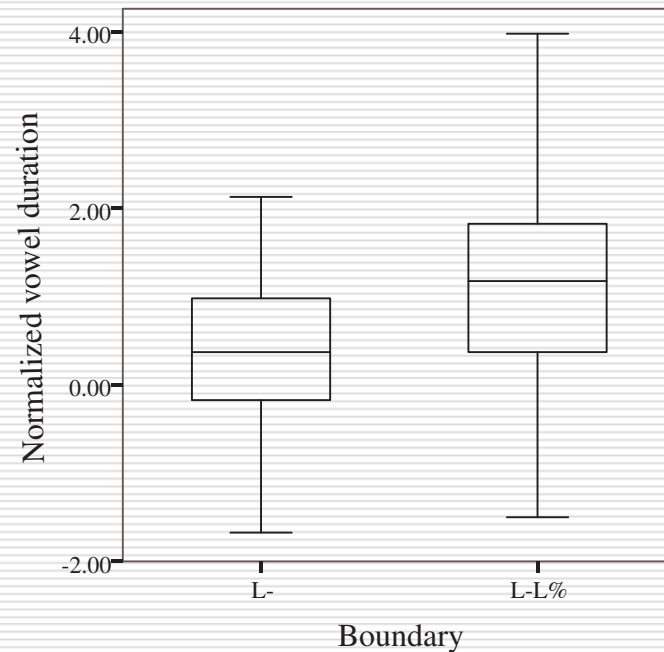
Switchboard vowel

- The difference in normalized nucleus duration between L- and L-L%:
 - $F(1, 313) = 15.748,$
 - $p < 0.001$



Boston Radio News vowel

- Speaker A:
 - $F(1, 245) = 20.969,$
 - $P < 0.001$
- Speaker B:
 - $F(1, 362) = 7.967,$
 - $p < 0.01$

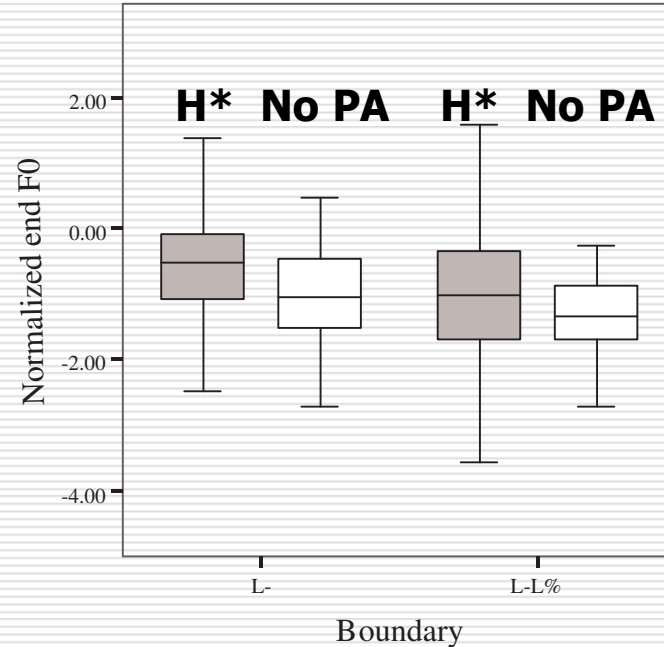


Pitch

- Switchboard: Short-Term Pitch Normalization
 - $E(F0_i) = (F0 - \mu^k) / \sigma^k$
 - Normalization based on intonational boundary
 - No normalization for Boston Radio News Corpus
-

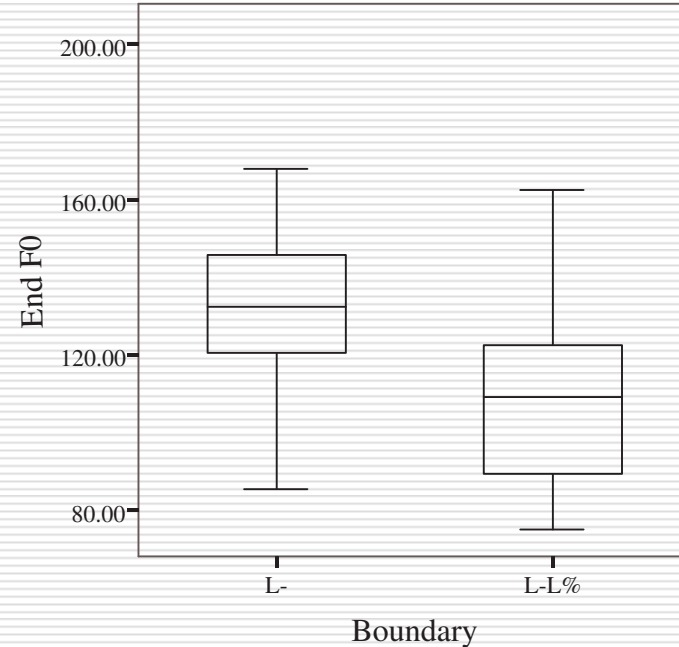
Switchboard: End F0

- L- tends to have a higher rime beginning F0 than L-L%, but the difference is not significant.
- Rime end F0 is significantly lower for L-L% than for L-.
 - $F(1, 276) = 7.597,$
 - $p < 0.01$



Radio News: End F0

- As in Switchboard, both speakers show F0 at the rime end is lower for L-L% than L-.
- Speaker A:
 - $F(1, 90) = 20.371,$
 - $p < 0.001$
- Speaker B:
 - $F(1, 94) = 19.316,$
 - $p < 0.001$



Radio News: F0 Drop

□ Besides End F0, **F0 drop** and **F0 slope** show sig. difference between L- and L-L%.

■ Speaker A:

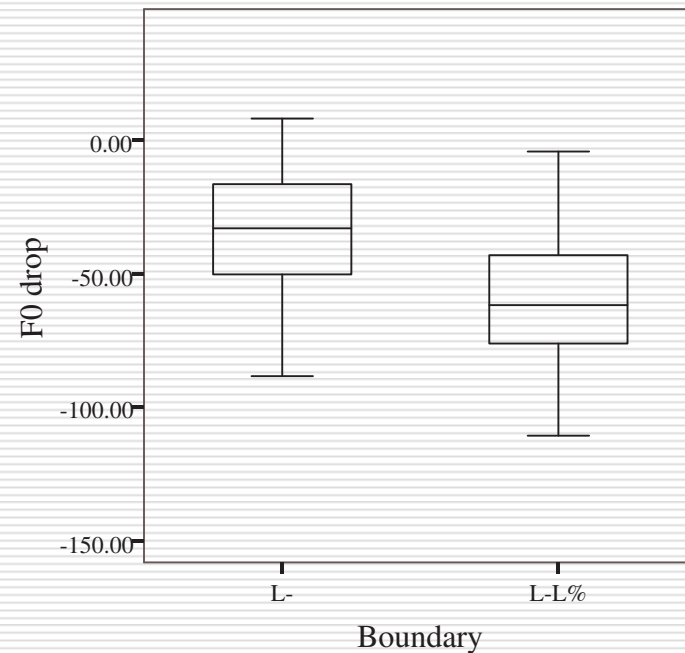
□ $F(1, 90) = 10.824,$

□ $p < 0.05$

■ Speaker B:

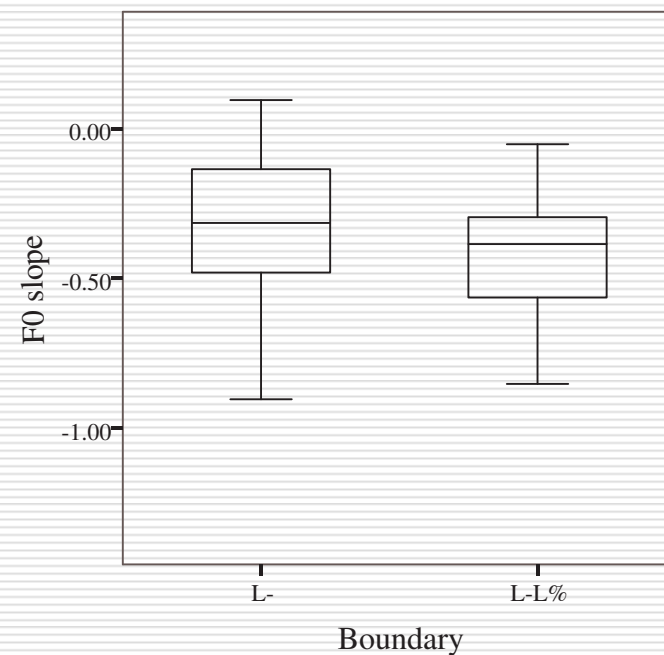
□ $F(1, 94) = 8.124,$

□ $p < 0.01$



Radio News: F0 Slope

- Speaker A:
 - $F(1, 90) = 4.929,$
 - $p < 0.01$
- Speaker B:
 - $F(1, 94) = 7.789,$
 - $p < 0.01$



Intensity

□ **Switchboard Corpus:**

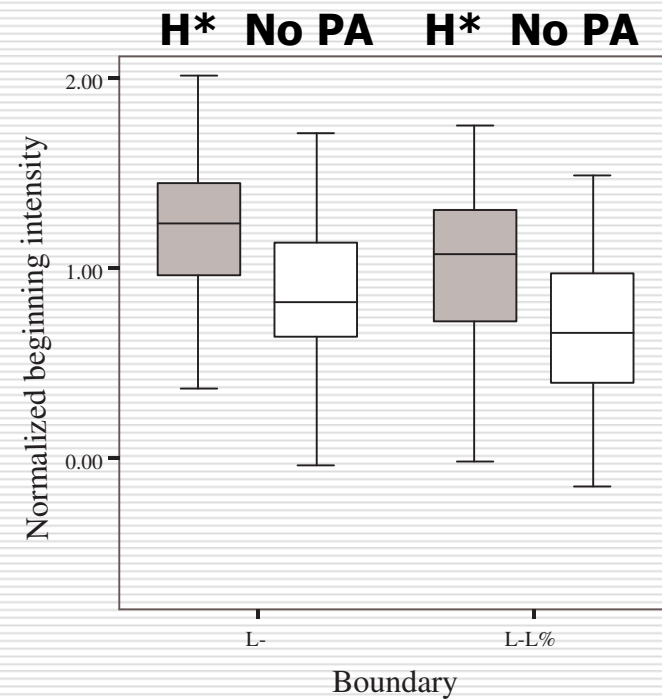
- For plain tokens, intensity at the rime end is not significantly different between L- and L-L⁰.
- Intensity at the rime beginning is sig. lower for L-L⁰ than for L-.

□ **Radio News Corpus** (speaker variation):

- Speaker A shows no sig. difference between the two boundary types for any measurement of intensity.
 - Speaker B shows significant difference in beginning intensity and end intensity between L- and L-L⁰ with lower intensity values for L-L⁰.
-

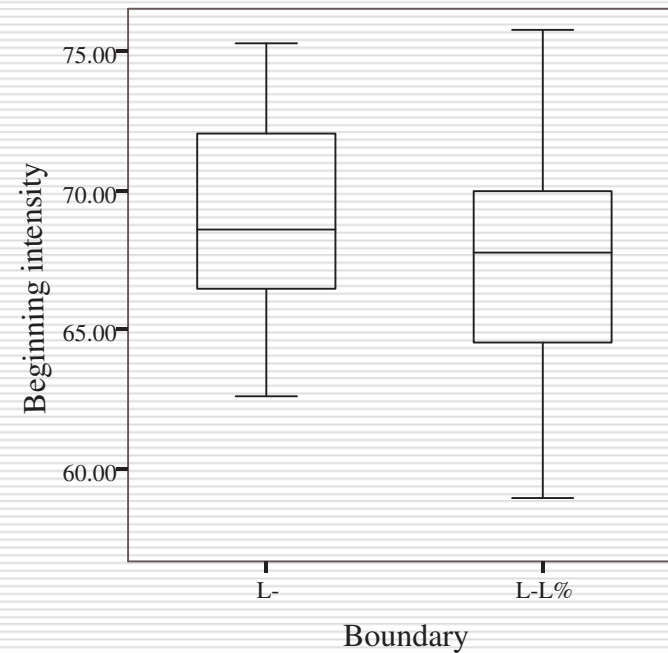
Switchboard: Normalized Beginning Intensity

- Intensity at the rime beginning is sig. lower for L-L% than for L-.
- $F(1, 276) = 12.769,$
- $p < 0.01$



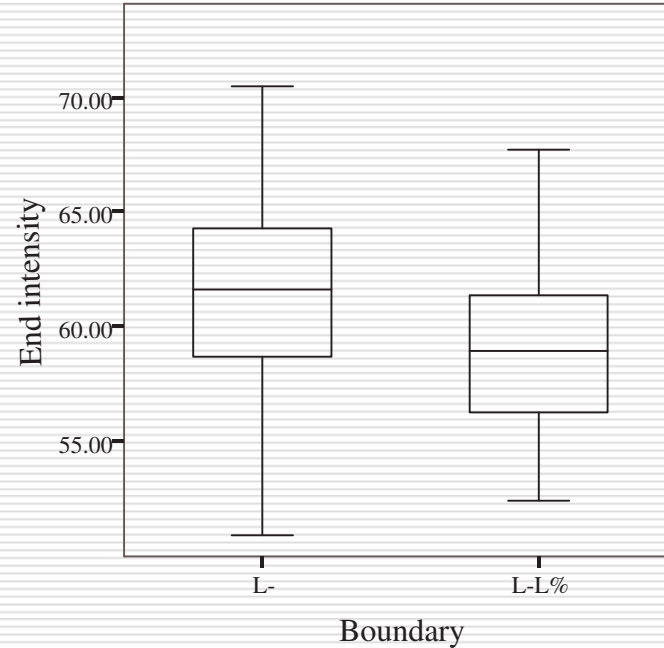
Radio News: Beginning intensity

- Beginning intensity
 - $F(1, 94) = 13.899$,
 $p < 0.001$



Radio News: End intensity

- End intensity
 - $F(1, 94) = 10.344,$
 $p < 0.01$



Voice Quality

- The creak distribution includes cases of complete pitch failure but did not include other instances of pitch error such as doubling and halving.
 - Frequency of creaky occurrence is greater for L-L⁰ than for L- (Redi & Shattuck-Hufnagel 2001).
-

Voice Quality: Switchboard

Boundary	Percentage of creak
L-	7.66% (17/222)
L-L ⁰ %	25.6% (30/117)

Voice Quality: Radio News

Boundary	Percentage of creak	
	Speaker A	Speaker B
L-	13.79% (12/87)	42.42% (42/99)
L-L%	70.67% (106/150)	73.41% (185/252)

Discussions & Conclusion

- ❑ Preboundary lengthening, preboundary F0, and the frequency of glottalization differentiate boundary levels across speakers, although no single feature is likely to serve as an effective classifier.
- ❑ Our findings provide important empirical support from non-laboratory speech for the Pierrehumbert-Beckman model in its distinction of two levels of phrase juncture.
- ❑ Our findings of acoustic correlates of phrase level in the phrase-final rime, and most often at the rime-end, offers critical support for the claim that prosodic features are locally rather than globally associated in phonological structure (Beckman & Ayers 1997, Ladd 2000).

Discussions & Conclusion

- Our speaker-dependent analysis of Radio News speech shows that speakers may vary in the prosodic features they use to mark boundary level distinctions.
 - We expect that a similar in-depth look at individual speaker data in Switchboard would result in additional findings of boundary level differentiation for conversational speech.
-