

Phonetic Encoding of Prosodic Domains in Naturalistic L2 Speech: Evidence from Korean-speaking English Learners' Spoken Data

Miroo Lee
Department of Linguistics, University of Pittsburgh

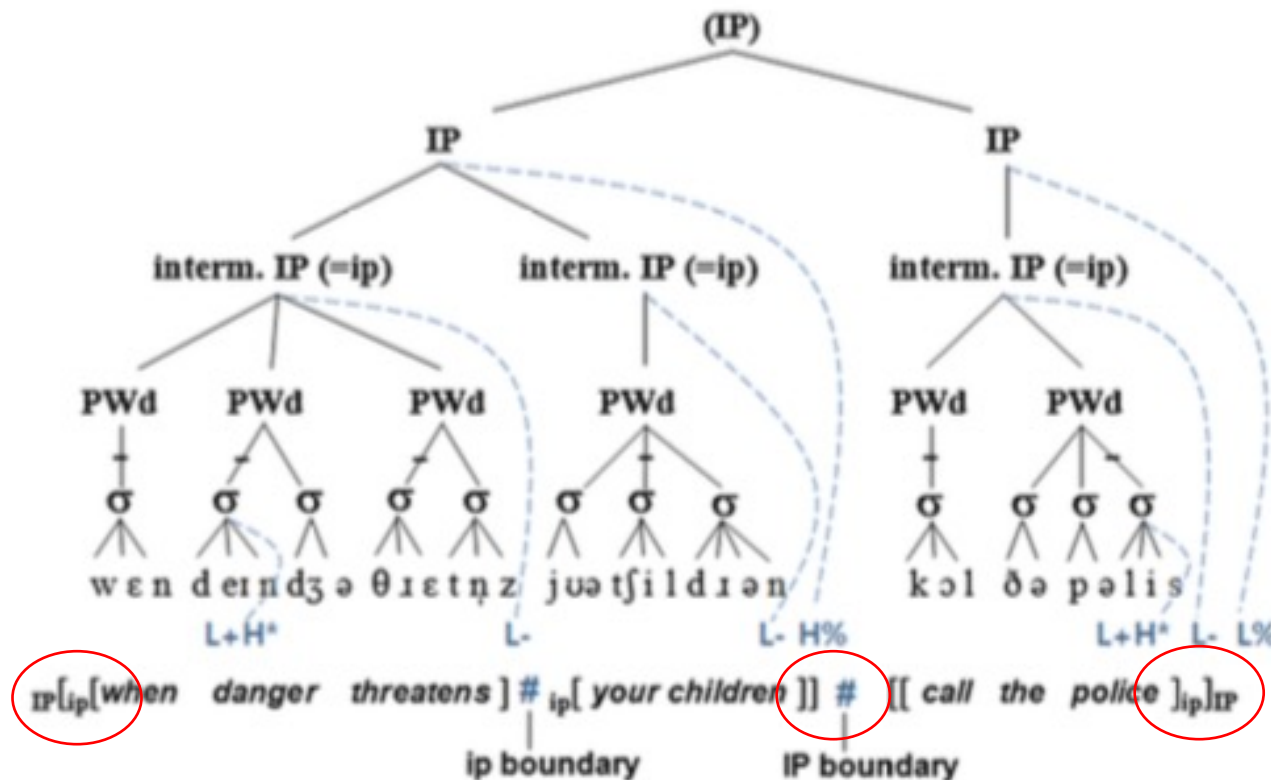
Content

1. Literature Review
2. Research Questions
3. Data Processing (Github)
4. Data Analysis (Github)
5. Notes

What is a prosodic domain?

The Intonational Framework

(Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980)



The syntactic approach

- focuses on how to predict a prosodic boundary from a *sentence*

The intonational approach

- focuses on how to detect a prosodic boundary from an *utterance*.

Figure 1

A prosodic structure of "When danger threatens your children, call the police."

Reprinted from Cho, T. (2016).

- In the intonational framework, prosodic domains are determined using pitch.

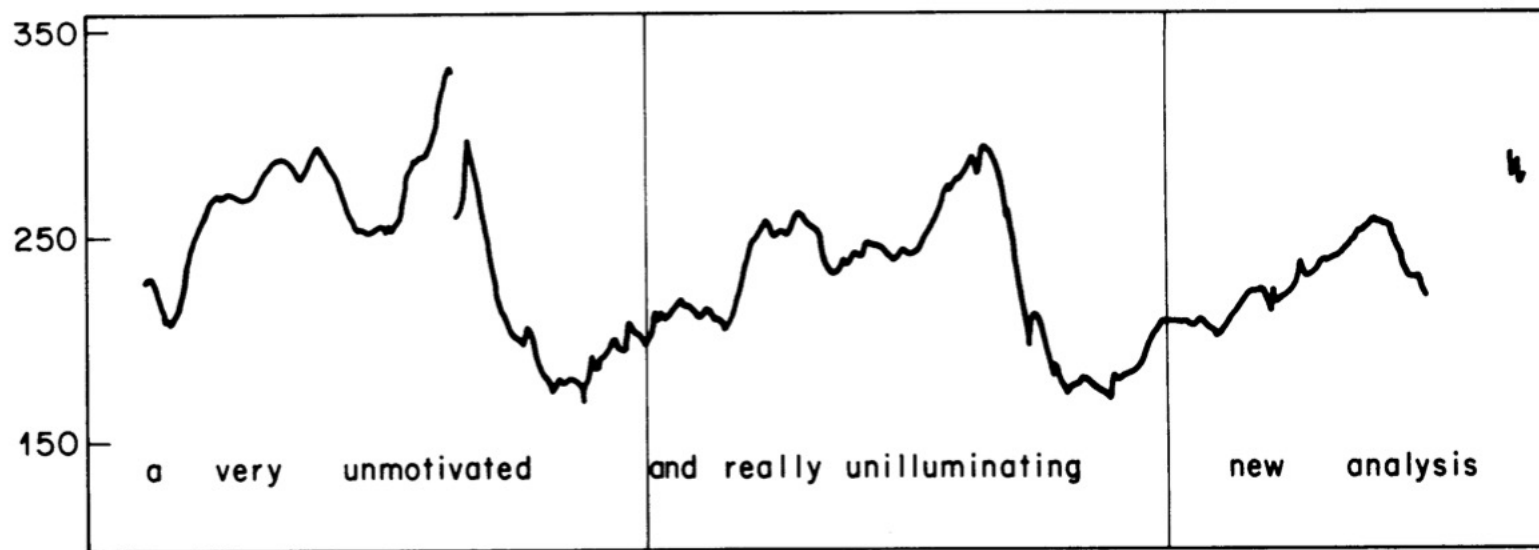


Figure 35

F_0 contour of three intonation phrases, in which phrasal manipulations of overall pitch range mimic catathesis.

Phonetic encoding of prosodic domains

So what is happening at the prosodic domain boundary?

- The durations of phonetic segments in the vicinity of prosodic domain boundaries tend to be *longer* than segments in the middle of prosodic domains (Byrd & Saltzman, 2003; Cho, 2016; S. Kim & Cho, 2013; A. E. Turk & Shattuck-Hufnagel, 2000).
- Prosodic domain boundary lengthening effects have been found cross-linguistically, but the detailed scope of the effects may vary (Cho, 2016; Jun, 1998b; Li & Post, 2014).

[Say 'ice']_{ip}, # ['c can again' with me]_{ip}.

[Say 'ice can again' with me]_{ip}. (Modified from Cho, Lee, & Kim, 2014)

Prosodic gestures and temporal modification

The π -gesture (Byrd & Saltzman, 2003).

- An abstract, non-tract gesture that occurs *at the near end of the prosodic phrase* and modulates the rate of the “clock” of articulatory gestures around the boundaries.
- The π -gesture is used to explain boundary-induced lengthening effects in intonational framework.

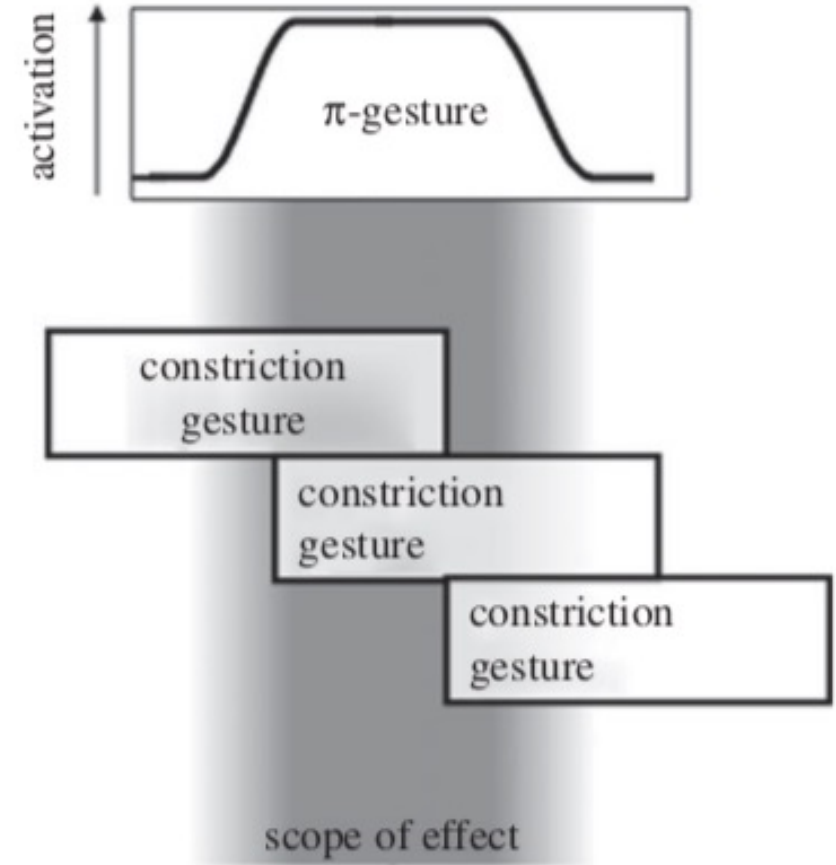


Figure 2

A schematic representation of the π -gesture.
Reprinted from Krivokapić (2014)

Two types of prosodic boundary effects

[Say 'ice']_{ip}, # ['can again' with me]_{ip}.



Domain Final Lengthening (DFL)

Significantly increase the duration of vowel and/or rhyme (Cho, 2016; Jun, 1998b).

Domain Initial Strengthening (DIS)

'Strengthened' articulatory gestures such as longer Voice Onset Time (VOT) and longer closure duration for aspirated stop consonants (Cho, 2016; Cho & Keating, 2009; Fougeron & Keating, 1997; Lisker & Abramson, 1964).

Prosodic gestures and temporal modification

The μ -gesture (Byrd & Saltzman, 2003).

- Anchored at the stressed syllables and modulate spatial and temporal aspects of phonetic segments.
- Explains longer syllable durations of stressed syllables compared to unstressed syllables.

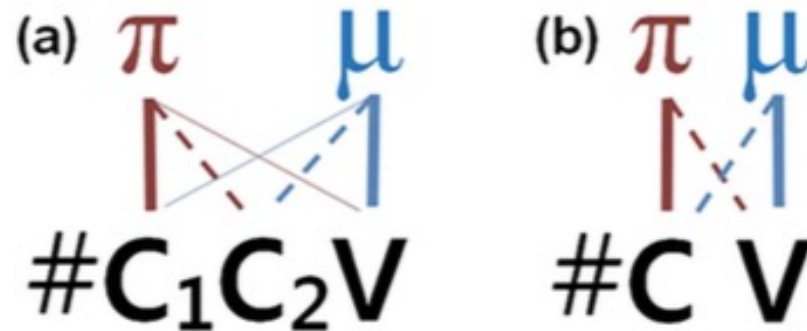


Figure 3

Schema for segmental associations of the π -gesture vs. the μ -gesture. The segment linked by a thick solid line refers to a “docking site” of a modulation gesture.

Reprinted from Cho (2016)

Previous Studies on L2 prosodic properties

Li & Post, 2014

- Examined rhythmic patterns in L2 English produced by German and Mandarin learners at intermediate vs. advanced learners.
- Found L2 speakers with low proficiency level showed significantly less accentual and final lengthening on syllable durations compared to high level learners who showed similar patterns to native English speakers.
- Most previous studies used read speech
- Syllable durations were used as the measure for the final lengthening effect

Challenges in studying L2 prosody

Languages use different tonal sequences to demarcate prosodic domains.

- English : the prosodic unit higher than a word is intermediate phrase (ip) and marked by a phrase accent, H- or L- (Beckman & Pierrehumbert, 1986).
- Korean: the prosodic unit higher than a word is accentual phrase (AP) and delimited by a phrase-final High tone (Jun, 1998b).

The ToBI system of prosodic labels

- **Orthographic tier:** time-aligned words
- **Break index tier:** degrees of junction between words (0 'no word boundary' to 4 'full intonational phrase boundary') Price et al. 1991
- **Tonal tier:** different types of pitch events (pitch accents, phrase accents, boundary tones)
- **Miscellaneous tier:** dysfluencies.

Pauses and tones together define two levels of phrasing, intermediate (ip) and intonational (IP) phrase.

Break indices

- 0: no word boundary (clitic: did#you, got#it)
- 1: most phrase-medial word boundaries
- 2: a strong disjuncture marked by a pause
- 3: marked by a single phrase tone
- 4: marked by a final phrase tone after the last phrase tone

Issues with conventional prosodic domain marking using ToBI

ToBI labeling is a slow manual process. Even when done by highly trained and experienced labelers, AmE-TOBI labeling takes from 100-200 times real time. (10s utterance -> 17-33minutes)

Difficult to analyze non-native speakers may not show the same use of tones to mark prosodic domains.

Pause durations as proxy measure for the prosodic domain

- Pause durations are proportionally related to prosodic juncture; the pause duration is likely to increase with prosodic boundary strength (higher the prosodic domains, greater the boundary effects) (Ramanarayanan et al. 2009; Choi, 2003; Horne, Strangert, & Heldner, 1995).
- The evidence of relations between pause durations and prosodic domains provides a way to examine L2 learners' domain-induced temporal modifications without having to rely on tonal sequences to define prosodic domains.
 - longer pause durations = higher prosodic domains = greater boundary effects
 - smaller pause durations = lower prosodic domains = smaller boundary effects

Current Study

R1. How are prosodic gestures realized in L2 speakers' semi-spontaneous speech?

R2. How are pause durations related to prosodic domain encodings?

R3. How does temporal modification change as the L2 speakers' proficiency level develops?

Hypothesis:

1. phonetic segments will be more lengthened before/after greater pause durations.
2. Pause durations will correspond to size of the boundary effects, encoding prosodic domains hierarchically.

Methods: Participants & Data

Participants

Four Korean students in an Intensive English Program at the University of Pittsburgh with age ranging between 18 to 35.

Speech data

Semi-spontaneous monologues collected as part of a Recorded Speaking Activity during regular class time (McCormick & Vercellotti, 2013).

“Choose a custom in your country. Describe what is done for this custom and why “

Initial data: two recordings from each student in the first (low-intermediate level) and third semester (advanced level) ($4 \times 2 \times 2 = 16$). More to be added.

**one student's recordings from first semester were unavailable and therefore replaced with the recordings from the second semester*

Data Processing Part I

Identify speech files for the analysis from the corpus index.

- L1: Korean
- Proficiency level: 3 and 5
- File type: 2min monologue

Manually annotate speech files using Praat.

Manually creating word list that specifies...

- The number of the syllables in each word
- Syllable structure (ex: CV, CVC)
- Whether the stress is lexically stressed or not.

***The above information was based on the CMU pronunciation dictionary*

Import annotated information from Praat textgrid to csv file

Methods: Annotation

Measure

- Pause durations
 - level 0 = silent pause < 0.15 sec
 - level 1 = silent pause 0.15 - 1 sec
 - level 2 = silent pause 1 - 2 sec
 - level 3 = silent pause > 2 sec
- Vowel durations in open syllables
- VOT% (VOT/syllable) of voiceless stops in syllable onset

Annotation

- Phrase initial (level 1 or higher pause to its left) or non-initial
- Phrase final (level 1 or higher pause to its right) or non-final
- lexical stressed or unstressed

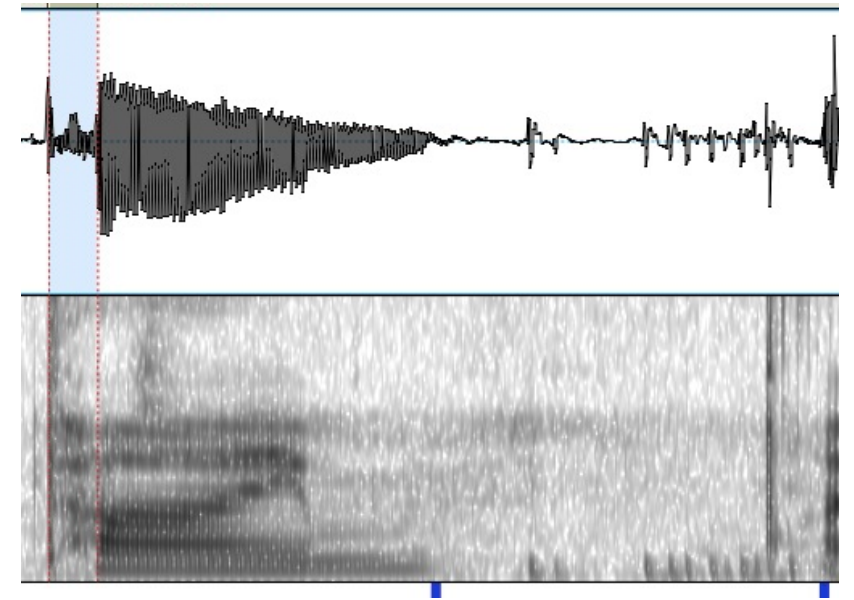


Figure 4

Waveforms and spectrogram of 'time' with a following pause

Methods: Analysis

Four data sets

| | |
|---|--|
| V durations in final, non-final syllables | V durations in initial, non-initial |
| VOT% in final, non-final syllables | VOT% in initial, non-initial syllables |

- Linear mixed-effects models with lme4 package in R (Bates, Mächler, Bolker, & Walker, 2014).
- Dependent variable: vowel duration, VOT% duration
- Fixed effects
 - a. stress (none vs. primary)
 - b. pause level (0 vs. 1-3)
 - c. semester: (1st vs. 3rd)

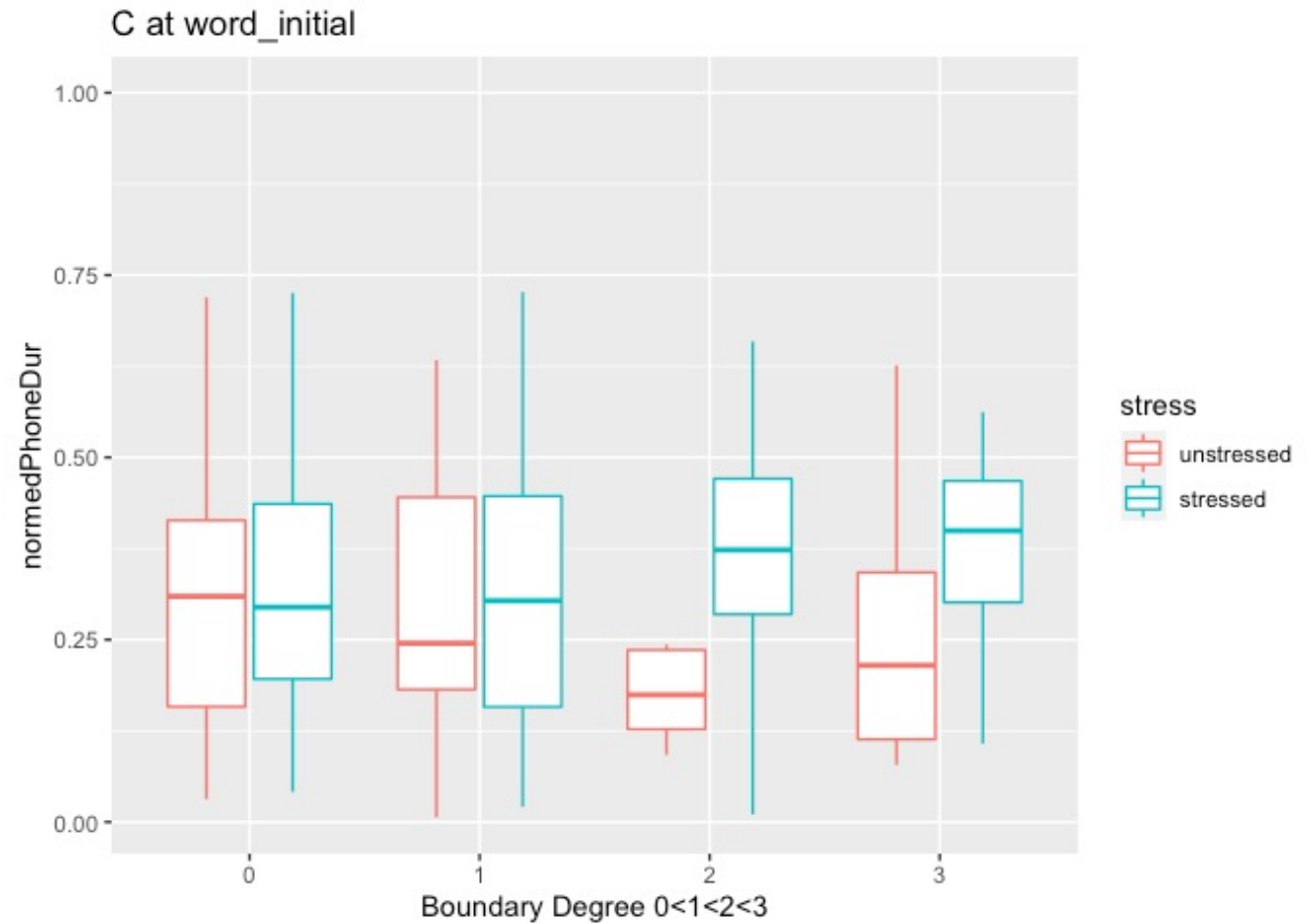
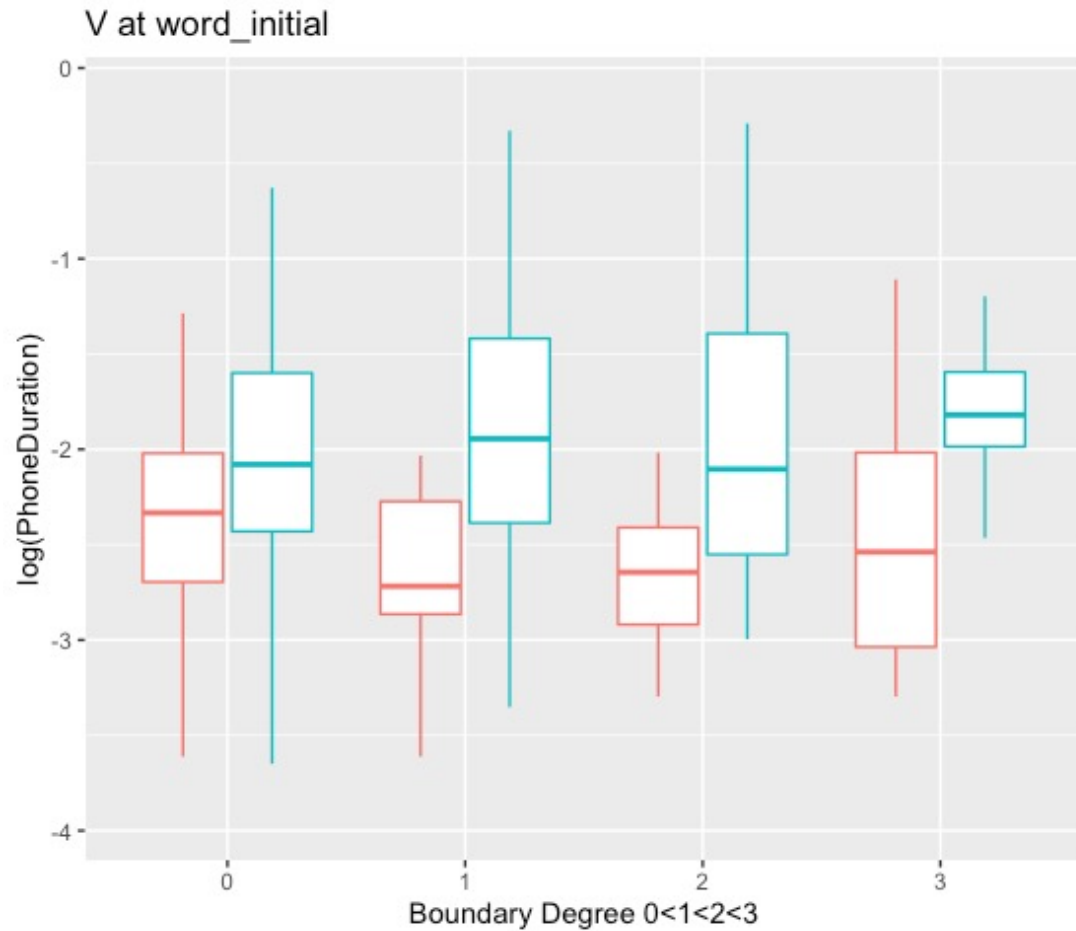
Analysis on DIS effect

| | |
|---|--|
| V durations in final, non-final syllables | V durations in initial, non-initial |
| VOT% in final, non-final syllables | VOT% in initial, non-initial syllables |

[Say 'ice']_{ip}, # ['can again' with me]_{ip}.

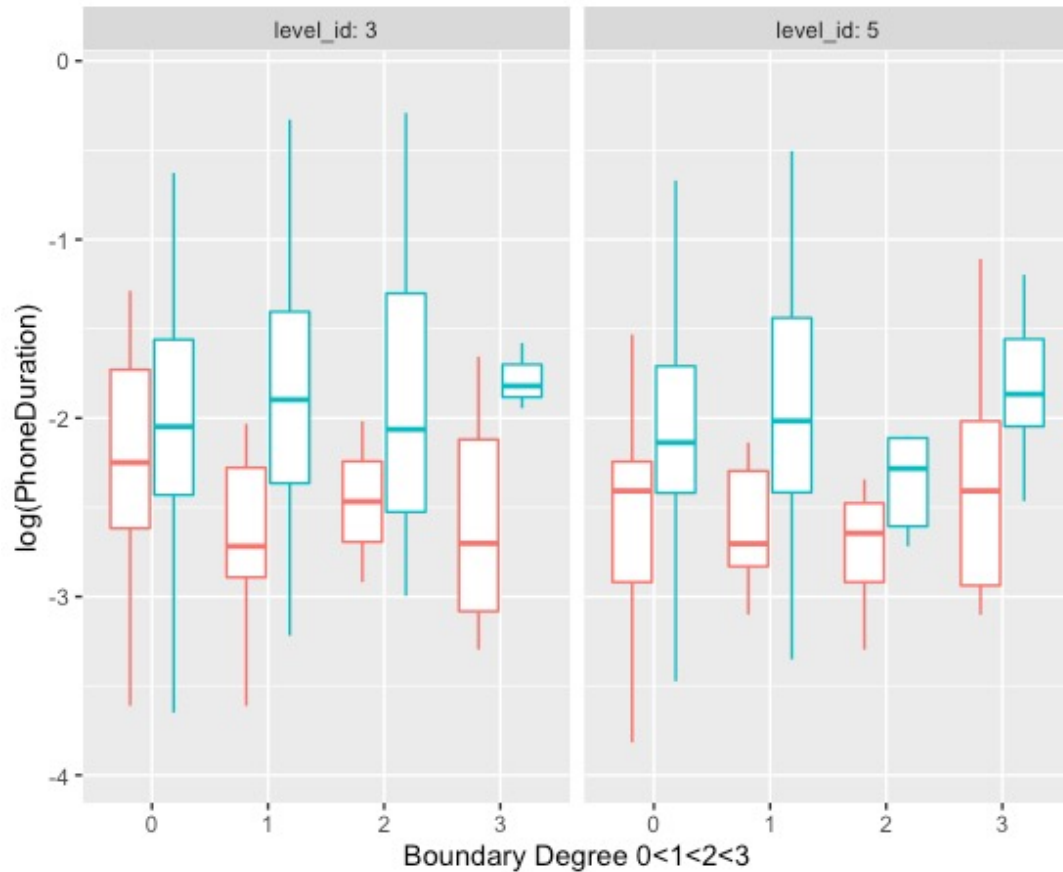


Stress and boundary effects on #CV,#CVC syllables.

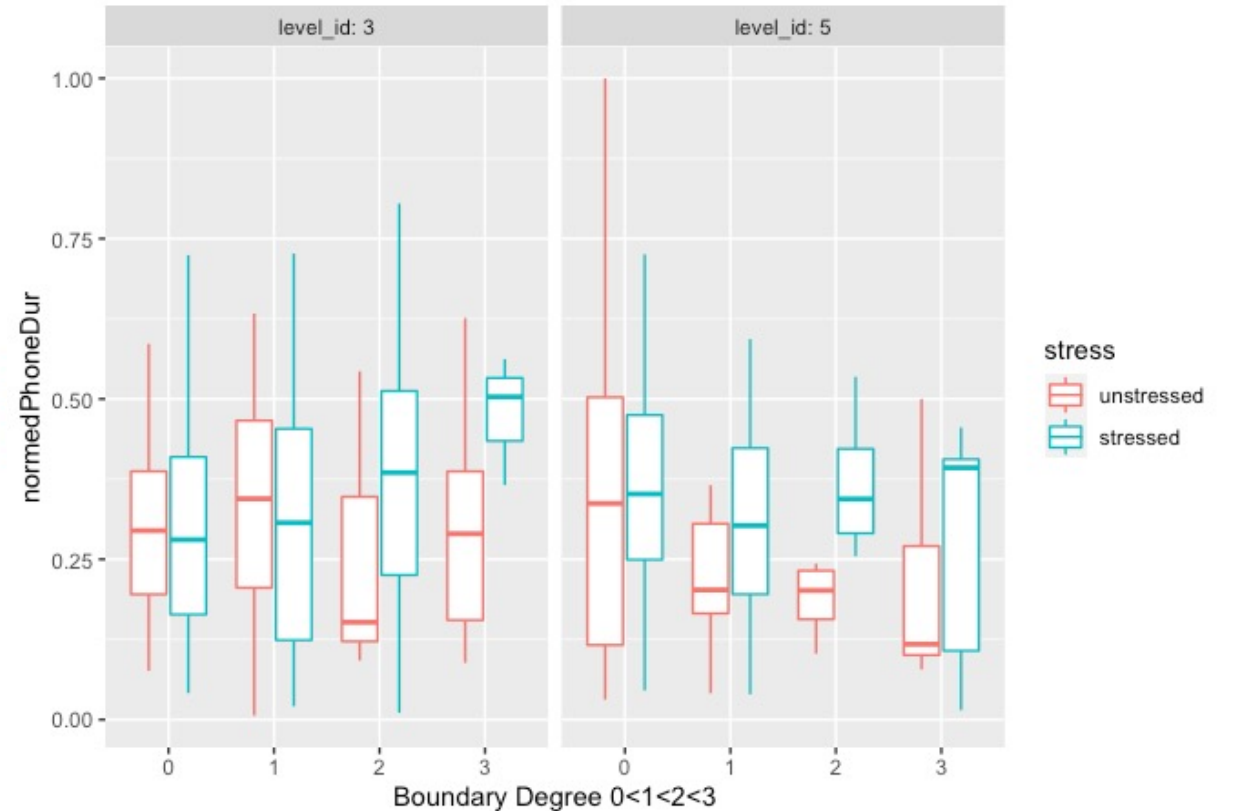


Stress and boundary effects in level 3 (1st semester) vs. 5 (3rd semester) on #CV,#CVC syllables.

V at word_initial in 1st vs. 3rd semester



C at word_initial in 1st vs. 3rd semester



Statistical results on word-initial vowels #CV

- No significant DIS effect
- Significant stress effect (increase in vowel duration)
- Marginal proficiency effect in that stress lengthening actually was reduced with DIS effect

Competitive relationship between DIS and stress.

Statistical results on word-initial consonants #CV

Changed the original plan of analyzing only voiceless VOT due to the small number of data points - > voiceless VOT and voiceless fricatives are included.

- No significant DIS effect
- Interaction between DIS and stress in that consonant% increase with DIS and stress.
- Significant interaction between DIS and proficiency in that consonant% increased with higher proficiency.

Complementary relationship between DIS and stress.

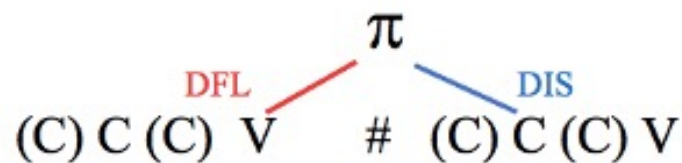
Discussion

Developmental patterns in temporal modifications of L2 speech

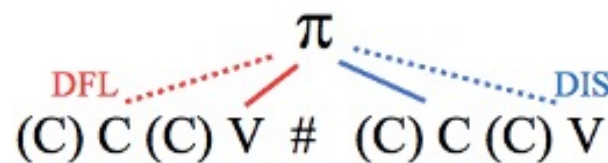
The data suggest expanded DIS pattern conditioned by the pause durations in relation to an increase in proficiency level.

L2 speakers' phonetic encoding of prosodic structures changes with more speech planning time. This suggests that L2 rhythmic structure of prosodic properties should be understood in the context of L2 speakers' cognitive load, rather than simply through L1 segmental or prosodic structure transfer.

L2 speakers' DFL & DIS



First semester



Third semester

Limitations & Future work

- Only CV syllable structure was examined at the moment – other syllable structure such as CVC needs to be analyzed.
- Larger dataset with various L1 backgrounds to examine how stress and boundary interactions are realized depending on the presence of lexical stress in a language.
- Certain part of the current data processing process (such as lexical stress specification for each syllable in a word) can be automated.

Thank you