Individual differences in the production of prosodic boundaries in American English

Jiseung Kim

University of Michigan
jiseungk@umich.edu

ABSTRACT

This study investigates individual differences in the weighting of phonetic properties in the production of prosodic boundaries in American English. The motivation of the study is to inform understanding of individual speaker variation and its accommodation in the representation of prosodic structure. In an acoustic study, 32 speakers produced 16 sentence pairs differing in type of boundary (Intonational Phrase (IP) boundary vs. Word boundary). Pause duration, phrase-final lengthening (three syllables before the boundary), phrase-initial lengthening (one syllable after the boundary), and pitch reset were examined. The results showed substantial individual differences in (1) which segmental and suprasegmental properties speakers phonetically modulated to produce IP boundaries, and (2) the scope and the degree of such modulations.

Keywords: individual difference, prosodic structure

1. INTRODUCTION

A phonological contrast is typically realized through multiple phonetic characteristics. Previous research has identified a set of primary acoustic properties that are relevant for marking different types of prosodic boundaries, such as pause duration, the lengthening of boundary-adjacent acoustic segments or articulatory gestures, and pitch reset ([7], [9], [10]). However, it is unclear how these different properties combine in the production of prosodic boundaries. This is in part due to large, and relatively unexplored, individual variation in the production of prosodic boundaries (e.g., [4], [5], [8]).

For example, [5] showed that speakers overall had greater linguopalatal contact for syllable-initial /n/s at larger prosodic phrases than those at smaller phrases, but individual speakers differed in how they distinguished the prosodic units by the degree of linguopalatal contact. [4] examined temporal and spatial dimensions of boundary-adjacent articulatory movements as well as the temporal scope of boundary effects, and their articulatory and acoustic results revealed large variations among the speakers.

The goal of this study is to delineate individual differences in the production of Intonational Phrase (IP) boundaries in American English, in order to understand how such variation is accommodated in the representation of prosodic structure. The current work is based on the understanding that individual speakers differ systematically from each other in how they convey prosodic structure, and focuses on how these individual speaker differences are manifested in the production of prosodic boundaries of American English. The main hypothesis is that individual speakers will show substantial variation in the phonetic features used and in the degree to which those features are used to express the IP boundary. An acoustic study involving 32 speakers of American English was conducted to investigate this hypothesis.

2. METHODS

2.1. Stimuli

Eight sentence pairs were constructed, differing in type of prosodic boundary (IP and Word boundary) and consonants used in the target words (/m/ in ‘maMIma’ and /n/ in ‘naNIna’) creating four conditions. To increase the variability, each of the four conditions included two different sentences, that systematically varied in the post-boundary target word (‘Melinda’ or ‘Belinda’ after ‘maMIma’, ‘Navarro’ or ‘Delilah’ after ‘naNIna’). The pre-boundary target words were presented to participants as ‘maMIma’ and ‘naNIna’ (the upper case signalled location of lexical stress) and described as novel names.

In the example below, the acoustic properties of the target word ‘maMIma’ and the first syllable of the post-boundary target word ‘Melinda’ were under investigation. The first part (A) of the dialogues provided the context for the target sentence pairs (B). The boldfacing in (B) signalled location of contrastive focus.

(a) A: The paramedic called maMIma. # Melinda and Peter said no one got hurt.
B: No, the **police** called maMIma. # Melinda and **Danny** said no one got hurt. (# denotes IP boundary)

(b) A: The paramedic called maMIma # Melinda. And Peter said no one got hurt.
B: No, the **police** called maMIma # Melinda. And **Danny** said no one got hurt. (# denotes Word boundary)

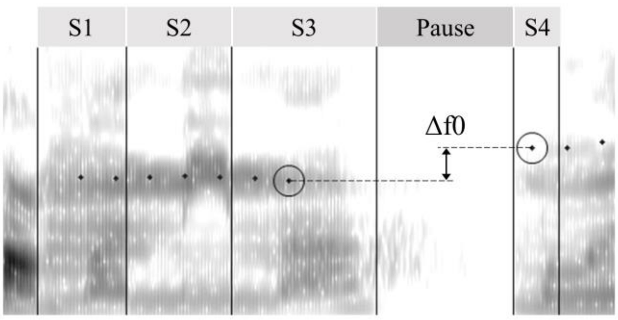
2.2. Participants and experimental procedure

Acoustic recordings were made of the production of 32 native speakers of American English. Participants were given the same verbal instructions about the target sentences. They were asked to silently read the context sentence first, and then read aloud the test sentence (B). The sentence pairs appeared on the monitor one at a time in a pseudo-randomized order in blocks of 16 sentence pairs (with half of the data collected as part of a separate experiment). Each pair was repeated nine times, for a total of 2,304 utterances (8 sentence pairs * 9 repetitions * 32 speakers).

2.3. Analyses

The pre-boundary target word and the first syllable of the post-boundary target word were segmented and labelled using Praat ([1]). The durations of these segments were measured in order to examine boundary-adjacent lengthening. The silent interval following the target words in the IP boundary condition was also measured. For pitch reset, the minimum f0 values in the syllables before and after the boundary were extracted in Hz. The difference between these two f0 values in each utterance was calculated (Δf0 = post-boundary f0 (Hz) – pre-boundary f0 (Hz)). A positive Δf0 indicated presence of pitch reset across boundary, whereas negative or zero Δf0 indicated absence of pitch reset. An example of the measurements is shown in Figure 1.

Figure 1. Example of the temporal and pitch measurements for ‘maMIma # Melinda’, in which # denotes IP boundary.



A set of Linear Mixed-effects (LM) models tested whether the duration of each of the four syllables depends on type of boundary across all speakers. In all LM models, BOUNDARY (IP vs. Word) was included as a fixed effect, while SPEAKER was included as a random effect. CONSONANT TYPE (C-TYPE) was included as an additional fixed effect if including the variable was found to significantly improve the fit of the model based on the results of a series of Chi-square tests. C-TYPE was included as a fixed effect in the models for Syllable 3 (S3) and Syllable 4 (S4), but not for Syllable 1 (S1) and Δf0. For S2, both C-TYPE and the interaction between the two fixed effects were included.

To assess the boundary effect on syllable durations and Δf0 within individual speakers, Linear Regression (LR) models were used. For syllable durations, two explanatory variables (BOUNDARY and C-TYPE) were included in the model that tested whether the mean syllable duration is predictive of the variable. For Δf0, there was one explanatory variable (BOUNDARY). All statistical models were fitted and analysed using R.

3. RESULTS

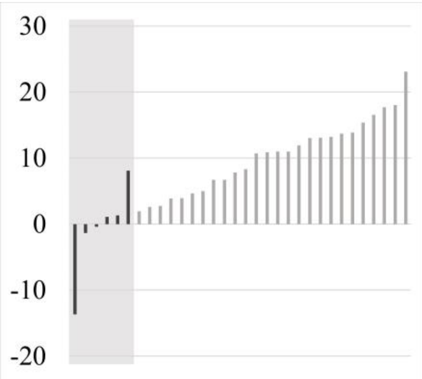
3.1. Across all speakers

The output of the LM models showed that, across speakers, S1 duration did not significantly differ depending on the type of boundary ($p=.142$), while the durations of S2, S3, S4 showed a boundary effect. S2 and S3 were lengthened in the IP boundary condition compared to the Word boundary condition ($p<.001$ for both), whereas S4 was significantly shorter in the IP boundary condition ($p<.001$). The effect of C-TYPE was significant in S3 and S4, such that S3 duration was longer when the pre-boundary target word was ‘maMIma’ compared to ‘naNIna’ ($p<.001$), and S4 duration was shorter when the pre-boundary word was ‘maMIma’ than when it was ‘naNIna’ ($p<.001$). On the other hand, the analysis of the LM model for Δf0 showed that, across speakers, there was larger pitch reset across the IP boundary than across the Word boundary ($p<.001$).

3.2. Individual speakers

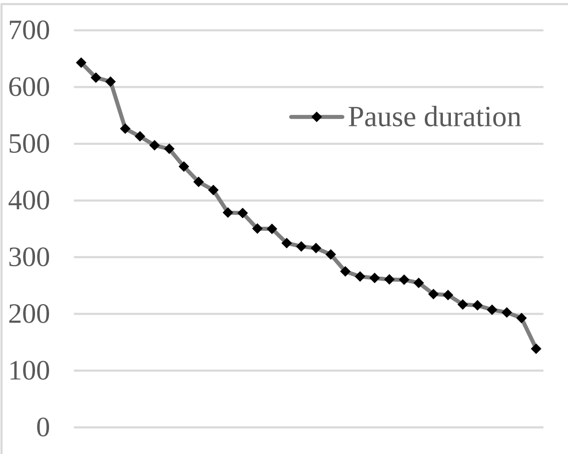
The output of the LM models performed for each participant revealed substantial differences among individual speakers. First, the analysis of the LR model for Δf0 found that, in three out of 32 participants, the pitch reset across IP and Word boundaries did not significantly differ. Out of the remaining 29 participants who showed significant differences on the size of reset across the two boundary types, three participants had negative values of Δf0 across IP boundary, meaning that the f0 extracted in the pre-boundary syllable (S3) was on average greater than the f0 in the post-boundary syllable (S4). A gradient distribution of mean Δf0 (Hz) of the 32 speakers in the IP boundary condition is represented in Figure 2.

Figure 2. Individual speakers' mean Δf0 (Hz) in the IP boundary condition. X axis represents 32 individual speakers. The leftmost six speakers in the grey box did not show significantly different pitch reset sizes between the IP boundary and the Word boundary conditions.



Pause durations averaged across individual speakers are distributed in a gradient manner, as represented in Figure 3.

Figure 3. Individual speakers' mean pause duration (ms) in the IP boundary condition. X axis represents 32 individual speakers.



The individual results for syllable duration showed that the scope of the phrase-final lengthening effect substantially varied across participants. For all participants, the phrase-final syllable (S3) was significantly longer in the IP boundary than in the Word boundary condition. However, the scope of the lengthening effect substantially varied across participants. For 15 participants, both S2 and S3 were subject to the lengthening effect. For nine participants, lengthening did not extend leftwards beyond S3. In addition, three participants showed lengthening in S3 and shortening in S1 in the IP compared to the Word boundary condition. Three other participants also showed shortening; one participant had shortening in S1 and lengthening on S2 and S3, while the other two participants had shortening in S2 and lengthening in S3. The remaining two participants showed lengthening in all three pre-boundary syllables.

The first syllable of the post-boundary target word (S4) also showed more than a single pattern. Twenty-three participants significantly shortened S4 duration in IP than Word boundary condition, while one participant significantly lengthened S4 in IP than in Word boundary condition. The remaining 9 participants did not show boundary effect on S4 duration.

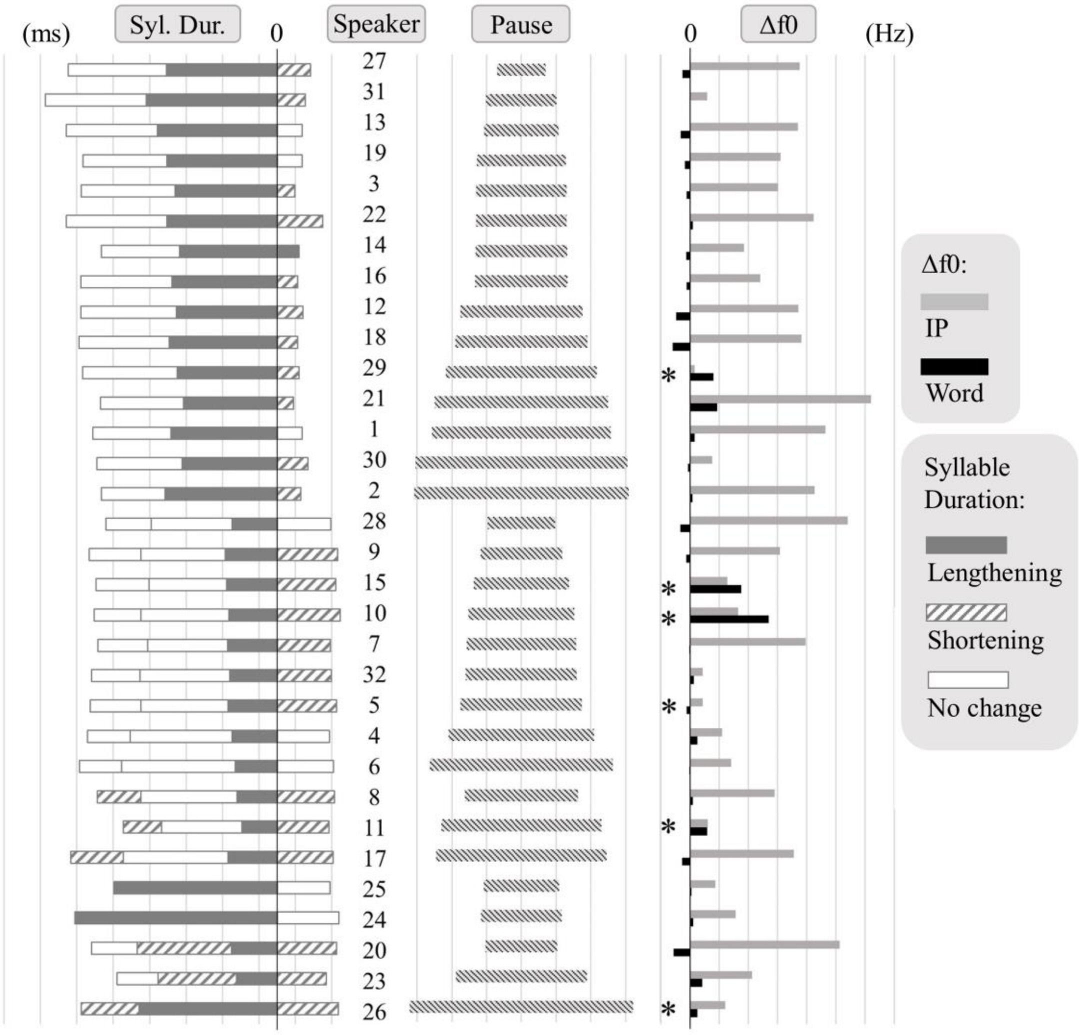
Figure 4 shows all 32 participants' mean values of the three acoustic measurements in three bar graphs. In the left graph, the horizontal bars represent S1-S3 durations of the pre-boundary target word and S4 duration of the post-boundary target word. The interval between the vertical guidelines in light grey is 100ms. The graph in the middle represents mean pause duration in the IP boundary condition. Again, the interval between the vertical guidelines is 100ms. The graph on the right represents mean Δf0 in the IP boundary condition (light grey bars) and in the Word boundary condition (black bars). The six participants with asterisk (*) are those who did not use Δf0 to differentiate IP boundary and Word boundary. The interval between the vertical guidelines is 25Hz.

4. DISCUSSION

Previous studies have documented significant inter-speaker variation in the effects of prosodic boundary. However, the results of such studies have tended to focus on the differences between group averages. The current study systematically investigated how individual speakers vary in the effects of prosodic boundary in the acoustic dimension.

The results showed that there is robust individual variation in terms of the type of acoustic correlates used to differentiate prosodic boundaries, suggesting that individuals encode prosodic structure differently. All speakers produced pauses at IP boundaries, but the pause durations varied across speakers in a gradient manner. Six out of 32 speakers did not produce IP boundary with a positive pitch reset, unlike the other 26 speakers. As for phrase-final lengthening, while all speakers employed phrase-final lengthening to some extent, they varied in how far the effect extended leftwards from IP boundary, indicating that the scope of lengthening is not uniform across speakers.

Figure 4. Mean values for syllable durations (Syl.Dur.), pause duration, and Δf0 for all 32 speakers.



For a subset of participants, phrase-final lengthening was accompanied by shortening in pre-boundary syllables (S1 and S2) and/or a post-boundary syllable (S4). Again, there was variation within these participants regarding which syllable(s) was shortened. The shortening is likely to be a compensatory consequence, rather than a process independent from the lengthening effect ([4], [6]).

There are a number of implications for the current study. First, the mixed results of previous studies on the production of prosodic boundaries might be due to systematic inter-speaker variation that needed to be taken into account. The current study showed that there seems to be no apparent relationship between how speakers modulated boundary-adjacent syllable durations and whether and how they used other acoustic correlates for IP boundary, such as pause duration and pitch reset (Figure 4). Moreover, the results of the current study suggested continuous extension of the lengthening effect of the IP boundary over a certain interval without skipping a syllable, as expected under the π-gesture model ([2]). In addition, the study provides evidence for (compensatory) shortening at prosodic boundaries, adding to the small body of research that has identified this effect. Lastly, current models of prosodic structure need to accommodate the fact that individuals may vary significantly while systematically modulating the acoustic correlates relevant for encoding a prosodic contrast.

6. CONCLUSION

The current study investigated the effect of two types of prosodic boundaries on a set of acoustic correlates. The analysis of the data at the level of individual speakers showed substantial variation among speakers, and revealed patterns of phrase-final lengthening that were not observed in the group-level analysis. The results of the study highlighted individual differences that need to be accounted for in different models of prosodic structure.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

[2] Byrd, D., Saltzman, E. 1998. Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics* 26, 173-199.

[3] Byrd, D., Saltzman, E. 2003. The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149-180.

[4] Byrd, D., Krivokapić, J., Lee, S. 2006. How far, how long: On the temporal scope of phrase boundary effects. *Journal of the Acoustical Society of America* 120, 1589–1599.

[5] Fougeron, C., Keating, P. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101,3728-3740.

[6] Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., Goldstein, L. 2014. The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics* 44, 61-82.

[7] Ladd, D. R. 2008. *Intonational phonology*. 2nd edition. Cambridge: Cambridge University Press.

[8] Mo, Y., Cole, J. 2010. Modeling perceived prosody – speaker-dependent vs. speaker-independent models. Poster presented at the LabPhon 12 in Albuquerque, NM.

[9] Swerts, M., Collier, R., Terken, J., 1994. Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication* 15, 79–90.

[10] 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* 91, 1707-1717.