Fine gradations of prosodic boundary strength can drive the assignment of prominence

Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Fine Gradations of Prosodic Boundary Strength can Drive the Assignment of Prominence

Soundess Azzabou – Kacem [1] & Alice Turk [2]

[1]The University of Edinburgh, UK & The University of Sousse, Tunisia
[2] The University of Edinburgh, UK
Sss_A_K@yahoo.com, a.turk@ed.ac.uk

Abstract

We know that post-lexical stress shift is blocked when two words with abutting main prominences (e.g., *afternoon* and *hike*) are separated by Intonational Phrase (IP) boundaries. However, when the clashing words belong to the same IP (e.g., *afternoon hike*), we do not know whether the shift can also be obstructed by finer gradations of prosodic boundary strength, below the IP level. This paper reports on an experiment manipulating prosodic boundary strength via constituent length variations. Twenty-nine speakers produced in context 28 pairs of potentially clashing sequences (e.g., *canteen soup* vs. *canteen supervisor*), where Word 2 was one syllable (e.g., *soup*, Short condition designed to elicit weaker prosodic boundaries) or four syllables long (e.g., *supervisor*, Long condition to elicit stronger boundaries). The length manipulation affected the stress shift rate, which was higher in the Short condition. The acoustic analyses show significant changes of syllable duration, f0 and SPL in both syllables of the target. The results provide evidence that the grouping within one IP of clashing words is not sufficient condition for stress shift to take place, as the shift can be obstructed by gradations in prosodic boundary strength between the clashing words, even if they belong to the same IP.

Index Terms: Post-lexical prominence, Early prominence, Stress shift, Prosodic boundary strength, Constituent length, Word length, the Rhythm Rule.

1. Introduction

Most authors agree that the presence of a major prosodic boundary between two potentially clashing words (e.g., *afternoon* and *hike*) prevents the operation of the Rhythm Rule, or stress shift from, e.g., *–noon* to *af-* in *afternoon* [1-7]. However, at least one aspect of this proposal remains unclear; When the potentially clashing words are not separated by a major prosodic boundary and are instead grouped within the same Intonational Phrase (IP) prosodic domain (e.g., IP *afternoon* # *HIKE* IP), does the strength of the sub-IP internal boundary # also matter in blocking stress shift?

The present paper reports on an experiment designed to address this question. We manipulated the strength of the phrase internal sub-IP prosodic boundary while controlling syntactic structure and prosodic prominence/ metrical structure.

1.1. Effects of constituent length

Because constituent length has been proposed to influence the strength of prosodic boundaries [8-13 inter alia] longer constituents (including words longer by virtue of greater syllable count) are predicted to result in stronger prosodic boundaries. In turn, prosodic boundaries are predicted to affect the assignment of prominence and the likelihood of stress shift. [12] found for Greek that constituent length manipulations produced effects on the strength of IP prosodic boundaries that are comparable to those induced by syntax. At the sub-IP prosodic levels, where the length effects are expected to be more subtle, there is now evidence at least for French [10, 14] that prosodic boundary strength effects induced by longer words have increased the incidence of the optional Initial Accent in longer words e.g., *somnambuliques* (‘*sleep-walking’*) as, compared to shorter words, e.g., the monosyllabic word *sottes* (‘*silly*’).

1.2. Experimental paradigm

In the present paper, we test whether word-length-induced differences in the strength of planned prosodic boundaries affect the location of pre-nuclear prominence in Word 1 in clashing Word 1 Word 2 phrases, keeping both syntactic and metrical structure under control. We present results from 1) perceptual tests of prominence as well as 2) measures of duration, F0, and SPL on matched pairs of words spoken by the same speaker, with and without perceived early pre-nuclear prominence.

2. Materials & Methods

2.1. Experimental design

2.1.1. Clash sequences

We designed a Word 1 Word 2 clash sequence (e.g., *canteen SOUP*) with a minimal number of syllables: a disyllabic Word 1 (e.g., *canteen*) followed by a monosyllabic Word 2 (e.g., *SOUP*), where Word 2 is designed to carry the main phrasal stress and Word 1 the pre-nuclear phrasal stress.

The experiment used 28 target critical phrases (clash sequences) where Word 1 (the site of potential stress shift) was kept the same across conditions. The length manipulation was performed by increasing the number of syllables in Word 2, from one to four syllables, e.g., *dams* (**Short condition**) and *dandelion* (**Long condition**) each appearing after, e.g., *Bombay*. All the quadrisyllabic Word 2s in the Long condition were designed with primary lexical stress on their initial syllables which start with the same onset and whenever possible the same following vowel as in the Short condition; e.g., /dæ/ in *dams* and *dandelion* or /su/ in *soup* and *supervisor.*

2.1.2. Targets

The critical phrases for both structures in a pair were embedded in **carrying target** sentences matched for number of syllables, and were preceded by **leading** sentences designed to elicit pre-nuclear prominence on Word 1 (the target), and nuclear prominence on Word 2; see example (1):

 10.21437/SpeechProsody.2022-56

(1) **Critical sentence trio (Short Condition)**

Context:	*We've finalised the curriculum for the first year architecture course.*
Leading sentence:	*We'll study the Tokyo bridges again.*
Target sentence:	*We'll study the **Bombay dams** again.*

Critical sentence trio (Long Condition)

Context:	*Our lab runs seasonal tests of medicinal plants.*
Leading sentence:	*We'll study the Tokyo ginger again.*
Target sentence:	*We'll study the **Bombay dandelion** again.*

All the material in the carrying target sentences were identical across conditions except for the shorter or longer Word 2 for each item pair across conditions. As example (1) shows, the sentence design used the same stretch of speech before and after the critical phrase, e.g., *We'll study the Bombay* [Word 2] *again*, where Word 2 is alternatively the shorter or longer version, e.g., *dams* or *dandelion*.

2.1.3. Fillers

Fillers used the same sentence-trio structure as the targets. Filler-Word-1s were tri-syllabic with stress on the second syllable, *e.g. potato* in *potato omelette.*

2.1.4. Design

Experimental sessions were divided into two parts of 28 target- and 30 filler-trios each. Each part contained two blocks of seven targets (a block of 7 items from the Short condition and a block of 7 different items from the Long condition) plus 15 fillers. Targets and fillers alternated, and the order of targets alternated between the Long and Short conditions within each part. The two parts started and ended with fillers and each speaker read each part and the material within each block in a different random order.

2.1.5. Speakers

The experiment used 29 native Scottish English speakers (19 female) who reported no language, speaking or hearing impairment. All speakers were students or recent alumni of one of the universities of the city of Edinburgh, UK. Data from 23 speakers (15 female) producing the 28 targets in the Short and Long conditions were analysed, totalling 1288 tokens.

2.2. Experimental procedure

Participants were asked first to read each sentence trio silently and understand the relationships among sentences; then, to produce the sentences so as to reflect those relationships. Sound recordings were made using a high-quality microphone, and digitised at a sampling rate of 48 KHz, at a bit depth of 16. The leading and the carrying target sentences were manually extracted from the main recording using *Audacity* and appropriately labelled, in preparation for the perception tests.

2.3. Perception materials and method

A seven-step scale with options ranging from 'Early sure' to 'Late sure' was designed to collect the judgements of prominence location (cf. the top part of Table 1, below). The expectation was that the use of this fine-grained 7-step scale would produce a more precise categorisation of prominence location than, for example, a forced choice scale with only two options ('Early' or 'Late'). Three expert judges, familiar with the material used, were asked to listen to the leading and the carrying target sentences and to determine if prominence was perceived earlier or later in the target according to the 7-step scale. The perception tests were run in two successive stages. In Stage 1, Judge 1 and Judge 2 provided judgements for the full set of 1208 tokens (604 in each condition). Then in Stage 2, Judge 3 provided judgements for a subset of the material (N = 296) divided equally between all those with prominence locations agreed by Judge 1 and Judge 2 and those without.

2.4. Categorisation of early prominence

To determine whether stress shift was perceived, the next step was to identify the tokens paired across conditions (henceforth token-pairs), where at least two judges shared the same percepts of prominence location. Notwithstanding the precision of this fine-grained scale, it may result in scattering the encoded locations of prominence into different categories. To maximize the number of 'Early' vs 'Late' token-pairs used for the analysis, the results of the perceptual categories that were used in the original collection of data (i.e., the 7-step scale) were aggregated by applying the step aggregation method shown in Table 1. The cut-off point is any judgement of 'Early' (categories 1-2), not including any 'Both' category (categories 3-5) for a **Stringent 3-step scale** (cf. Table 1**).**

Table 2. *Seven-step judgement collection scale and method for the categorisation of perceived early prominence*

Category	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Original 7-step scale	Early Sure	Early not so sure	Both or early	Both sure	Both or late	Late not so sure	Late Sure
3-step scale	Early prominence		Both			Late prominence	

2.5. Variables: Perceived 'Early' prominence rates and scores

Two types of perceptual results are presented. First, the overall ***rates*** of 'Early', 'Both' and 'Late' judgements are described for each judgment sample. The perceived Early prominence ***rates*** are the proportion of stress shift in each experimental condition. To test whether the overall rate of perceived 'Early' prominence is different across conditions (e.g., Short vs. Long in the Length Experiment), a prominence perception ***score*** consisting of the number of 'Early' prominence judgements (0-2) collected in Stage 1 (cf. 2.3) was calculated for each target token. The ***scores*** are also used as the dependent variable in the statistical tests.

2.6. Preparing files for the acoustic analysis

2.6.1. Item pairing

For the purpose of valid comparisons across the Short vs. Long conditions, we analysed paired items. Every item produced by one participant in one condition; e.g., speaker P227's rendition of *bamboo decks* (Short condition), was paired with its 'mate' from the other condition, e.g., *bamboo decorators* (Long Condition) as produced by the same speaker P227. The acoustic analyses will be performed in order to determine which acoustic correlate made the judges hear a particular pattern of prominence location across token-pairs.

2.6.2. Classification of intonational contours

Items were included in the analysis only if they had the phrase nucleus on Word 2, *e.g.* SOUP in *canteen SOUP*. Renditions of critical phrases involving a low accent, i.e., either 1) a low upstepping tone in the critical phrase, or 2) with low pitch accent on syllable 1 in Word 1 (e.g., on *can-* in *canteen*) were not included in the analysis. Therefore, all remaining pitch accents used for the acoustic analysis of the correlates of prominence included either H* or a downstepped H* _but crucially no L.

2.7. Acoustic variables and segmentation methodology

Carrying target sentences were manually segmented and annotated using *Praat* [15]. Overall, the segmentation protocol followed the methodology proposed in [16].

2.7.1. Durations of syllable 1 (s1) and syllable 2 (s2)

Syllabic duration was automatically measured for syllable 1 (e.g., *can-* in *canteen*) and syllable 2, (e.g., *-teen*) following manual segmentation. Syllabification normally followed the maximal onset principle. But, the main consideration in determining the onset and offset of the intervals required that the same reliable segmentation criterion be applicable to each member of a token pair for each speaker across conditions.

2.7.2. Fundamental frequency (f0) measures

As explained above, critical phrases with low pitch accent on the initial syllable were excluded from the main f0 analysis, as were their matching tokens uttered by the same speakers in the other length condition. Fundamental frequency (f0) was measured for the vowels of syllable 1 (e.g., *can-* in *canteen*) and syllable 2, (e.g., *-teen*), trimmed to exclude the first and last 15 milliseconds (ms) of each vowel interval, where most of the micro-prosodic effects [17-19] are likely to take place. The value of 15 ms was determined after visual inspection and hand measurements of all the vowels. The voiced vocalic interval less the first and final 15 ms will be referred to as the ***central portion of the vowel***. The segmentation of the vowel interval was determined by spectral discontinuity associated with constriction onset/release, under auditory control. The f0 variables were: f01, f02 and f01Max, f02Max: respectively mean f0 and f0 peaks/ maxima in the central portion of the vowels of syllables 1 and 2. The f0 measurements were automatically extracted for each speaker under visual and auditory inspection. Errors in f0 tracking were corrected by manually calculating the frequency from the waveform.

2.7.3. Sound Pressure Level (SPL) measures

Sound Pressure Level (SPL) means: the calculation of the SPL means uses the averaging method that convolves a Gaussian analysis window [15, p. 10ff], the duration of which is computed from the minimum pitch value for each speaker (the pitch floor). The step size is 0.8 divided by the pitch floor; this is the default method used in *Praat* [15]. SPL peaks: Maximum intensity in the central portion of the vowel was also measured for syllable 1 and syllable 2.

2.8. Statistical procedure

Within-subject designs were implemented. For the **perceptual tests**, the test statistics used for perceptual analyses were paired samples *t*-tests run on the perceived 'Early' prominence scores for two experimental conditions (e.g., Short vs. Long). Within-subject paired samples *t*-tests were also used for the **acoustical tests**.

3. Results

3.1. Perception results

3.1.1. Early prominence rates

Table 3. *Judgments of prominence location in Word 1 (e.g., canteen) in the Short (e.g., canteen soups) and Long Word 2 (e.g., canteen supervisors) conditions for the tokens where both Judge 1 and Judge 2 agreed on prominence location.*

Condition	Location	7-step scale (original)	3-step scale
SHORT	**EARLY**	**262**	**376**
(e.g., *canteen*	BOTH	1	12
soup)	LATE	43	50
LONG	EARLY	195	290
(e.g., *canteen*	BOTH	2	19
supervisor)	**LATE**	**59**	**77**
N		562	824

Results show that:

1) the number of 'Early' items in the Short condition outnumbers the number of 'Early' in the Long condition. For instance, both Judge 1 and Judge 2 agreed that 376 tokens in the Short condition were 'Early' compared to 290 tokens in the Long condition, as filtered through the 3-step scale; and that the reverse pattern is observed for the 'Late' judgements, i.e., 77 tokens heard as 'Late' in the Long condition relative to only 50 in the Short condition; cf. Table 2.

2) the incidence of early prominence is quite high regardless of condition; 75% in the Long- and 86% in the Short- condition. See Figure 1 for the proportions of perceived prominence using the 3-step scale ($N = 824$) for the tokens where Judge 1 and Judge 2 agreed. Results show that for both the Short and Long conditions, the judgments of perceived 'Early' prominence are at least three times as frequent as the judgements of 'Late'.

Figure 1. *Distribution of perceived 'Early', 'Both' and 'Late' prominence location in the target, (e.g., canteen) in the Short (e.g., canteen soup) and Long (e.g., canteen supervisor) condition for the tokens where Judge 1 and Judge 2 agreed using the stringent 3-step scale reduction method. N = 824.*

3.1.2. Perceived 'Early' prominence scores

The 'Early' prominence scores are significantly higher in the Short condition (e.g., *Bombay* DAMS; $M = 1.49$, $SD = .72$) relative to the Long condition (e.g., *Bombay* DANdelions; $M = 1.29$, $SD = .77$). The results of the paired-samples *t*-test ran on the scores for each token pair by Length condition show highly significant difference between the 'Early' scores (Short – Long; Stringent 3-step-scale; $t(603)= 6.570$, $p < .0001$).

3.2. Acoustic Results

We present acoustic results to answer the question of whether the token pairs with the perceptual stress shift pattern reliably exhibit acoustically prominent first syllables (e.g., *car-* in *cartoon*) when perceived as 'Early' and second syllables (e.g., *-toon*) when perceived as 'Late'. Paired *t*-tests were run on any token perceived as 'Early' and its mate produced by the same speaker in the other condition was perceived as 'Late', where any two of the three judges agreed on prominence location in both of the paired tokens.

3.2.1. Syllable duration

For any token perceived with 'Early' prominence, **Syllable 1** (e.g., *mar-* in *marquee*) is significantly longer (+11%) than in the token perceived as 'Late' (M_D = 29 ms, *SD* = 36; *t*(23) = 4.033; *p* < .0001, 1-tailed), while **Syllable 2** (e.g., *quee-* in *marquee*) shows a significant durational decrease (-14.6%) in the 'Early' relative to the 'Late' tokens (M_D = -34ms, *SD* = 41; *t*(23) = -4.080; *p* < .0001, 1-tailed); cf. Figure 2.

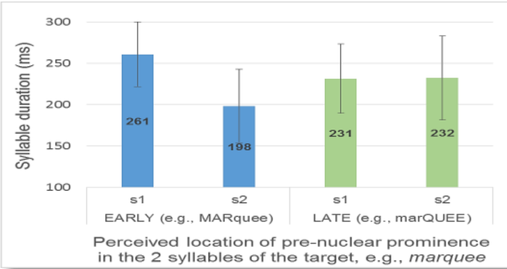

Figure 2. *Duration (and SD) of the two syllables (s1, e.g.,* mar- *and s2, e.g.,* -quee*) of the target (e.g.,* marquee*) for all the token-pairs perceived (Early vs. Late).*

The perceptual judgements for the (Early-Late) pairs are associated with a durational difference of about 30 ms in both directions. Since the items we examined were perceived as stress shifting, these results indicate two facts: 1) that syllable duration is one acoustic correlate of stress shift in this experiment, and 2) that changes in both Word 1 syllables could contribute to the perception of Early/ Late prominence.

3.2.2. F0

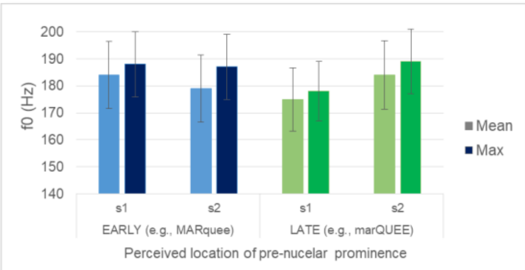

Figure 3. *Mean and Max f0 (and SE) in the two syllables (s1, e.g.,* mar- *and s2, e.g.,* –quee*) of the target (e.g.,* marquee*) for all the token-pairs perceived (Early vs. Late).*

When the prominence location in the target is reported as perceived 'Early', **Syllable 1** shows higher f0 means (*M* = 9 Hz, *SD* = 12, *t*(24)= 3.763, *p* < .001 (1-tailed)), and higher f0 maxima (*M* = 10 Hz, *SD* = 13; *t*(24)= 3.756, *p* < .001 (1- tailed)) than when the prominence location for the same items is perceived 'Late'. **Syllable 2** shows smaller (Early – Late) numerical differences (-5 Hz in f0 mean, *t*(24) = - 1.645, *p* = .0565) and -1 Hz in f0 Max, *t*(24) = -.595, *p* = .279) that do not reach statistical significance. The **f0 means** in the target show a falling trend (s1 > s2) in the 'Early' tokens and a rising trend (s1< s2) in the tokens perceived with 'Late' prominence. As for the **f0 maxima**, they do not differ in the 'Early' tokens (level), but are significantly different (rising) in the 'Late' tokens. The f0 maxima differences may not be as robust as the f0 means as they take one value only for each token.

3.2.3. SPL

Syllable 1 (e.g., *Bom-* in *Bombay*) showed significant differences between the (Early - Late) tokens (SPL **mean** M_D = 1.7 dB, *SD* = 3.7; *t*(24) = 2.33; *p* = .015; SPL **Max** *M* = 1.8 dB, *SD* = 4; *t*(24) = 2.26; *p* = .017; 1-tailed). **Syllable 2** (e.g., *–bay* in *Bombay*) SPL mean and maxima values are higher when pre-nuclear prominence was perceived as 'Late' than when it was perceived as 'Early' (mean and Max respectively M_D = -1.4 dB and -1.5 dB, *SD* = 2.5 and 2.7); SPL **mean**, *t*(24) = -2.78; *p* = .005 (1-tailed); SPL **Max**, *t*(24)= -2.74; *p* = .006; 1-tailed.

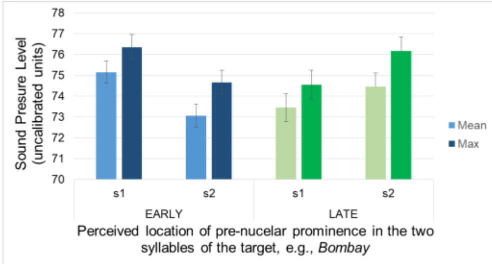

Figure 4. *Sound Pressure Level (SPL) means and maxima (and SE) in the vowels of the two syllables (s1 and s2) of the target, e.g.,* marquee *for the (Early vs. Late) token pairs.*

The above SPL results show that Syllable 1 has higher SPL levels in the Short condition and in the 'Early' tokens (+ 2 dB) while Syllable 2 has higher SPL levels in the Long condition and in the 'Late' tokens (≅ 1.5 dB). The SPL results confirm the previous duration and f0 results that the acoustic characteristics of syllable 1 (e.g., *Bom-* in *Bombay*) correlate with the perception of 'Early' pre-nuclear prominence in this experiment.

4. Discussion

The goal of this experiment was to test whether the length-induced gradations of prosodic boundary strength had an effect on the location of pre-nuclear prominence/ the incidence of stress shift. For cases with judged differences in prominence, increasing the length of Word 2 by changing the number of syllables (*cats* vs *caterpillars*) had an effect on the likelihood of early prominence in Word 1, which is higher when Word 2 is Short (e.g., *Argyle CATS*) in Word 1 Word 2 sequences. The results confirm the prediction that shorter Word 2s, e.g., *cats* (vs. *caterpillars*) can increase the likelihood of assignment of prominence in the pre-nuclear word, e.g., *Argyle*. The acoustic analyses show higher duration, f0 and Sound Pressure Level values for the first accentable syllable when the stress-shifting word is perceived as 'Early' relative to 'Late'. For the same syntactic mappings across Word 2 Length conditions (e.g., *Bombay dams* and *Bombay dandelion*) and the same stress clash configurations, Word 2 Length affected the location of pre-nuclear prominence in Word 1 (e.g., *Bombay*) indicating that prosodic constituent structure can play a role in the assignment of pre-nuclear prominence. The tokens produced in the Long condition that are perceived with Late pre-nuclear prominence (on, e.g. *-bay* in *BomBAY DANdelion*) may have a different sub-IP structure than their mates produced in the Short condition and perceived with Early prominence (*BOMbay DAMS*). The results presented here confirm that even at this lower sub-IP prosodic level, word-length-induced fine gradations of prosodic boundary strength can have a significant effect on the likelihood of assignment of early prominence.

5. References

[1] M. Liberman & A. Prince, "On stress and linguistic rhythm", *Linguistic Inquiry, 8*, pp. 249-336. 1977.

[2] A. S. Prince, "Relating to the Grid", *Linguistic Inquiry, 14*(1), pp. 19-100, 1983.

[3] E. Selkirk, "On derived domains in sentence phonology", *Phonology Yearbook 3*, 371-405, 1986.

[4] Selkirk, E., "Sentence Prosody: Intonation, stress, and phrasing", in J. A Goldsmith (Ed.), *The Handbook of Phonological Theory*, pp. 550-569, Oxford: Blackwell Publishing, 1995.

[5] M. Nespor, & I. Vogel, *Prosodic phonology: With a new foreword* (Studies in generative grammar; 28). Berlin: Mouton de Gruyter, 2007 (Original work published in 1986).

[6] S. Inkelas & D. Zec, "The Phonology/Syntax Interface", in J. Goldsmith (Ed.), *Handbook of Phonological Theory*. Basil Blackwell, 1995.

[7] E. Grabe, & P. Warren, "Stress Shift: Do speakers do it or listeners hear it?", in B. Connell (Ed.), *Laboratory Phonology IV: Phonology and Phonetic Evidence*, pp. 95-110, Oxford: Oxford University Press, 1995.

[8] E. Zvonik, & F. Cummins, "The effect of surrounding phrase lengths on pause duration", In *EUROSPEECH-2003*, 777-80, 2003.

[9] J. Krivokapić, "Prosodic planning: Effects of phrasal length and complexity on pause duration", *Journal of Phonetics, 35*, pp. 162-179, 2007.

[10] C. Astésano, E. G. Bard, & A. Turk, "Structural Influences on Initial Accent Placement in French", *Language and Speech, 50* (3), pp. 423-446, 2007.

[11] J. Krivokapić, & D. Byrd, "Prosodic boundary strength: An articulatory and perceptual study", *Journal of Phonetics, 40*(3), p. 430-442, 2012.

[12] E. Kainada, *Phonetic and phonological nature of prosodic boundaries: evidence from Modern Greek* (Doctoral dissertation), Retrieved from Edinburgh Research Archive. University of Edinburgh, UK, 2010.

[13] S. Fuchs, C. Petrone, J. Krivokapić, & P. Hoole, "Acoustic and respiratory evidence for utterance planning in German", *Journal of Phonetics*, *41*(1), pp. 29-47, 2013.

[14] L. Garnier, L. Baqué, A. Dagnac, & C. Astésano, « Perceptual investigation of prosodic phrasing in French", *Proceedings of Speech Prosody 2016*, pp. 1153-7, Boston, MA, USA, 2016.

[15] P. Boersma, & D. Weenink, *Praat*: doing phonetics by computer. (Version 5.4.08) [Computer software], 2015. Retrieved from http://www.praat.org/

[16] A. Turk, S. Nakai, & M. Sugahara, "Acoustic segment durations in prosodic research: a practical guide", in S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, & J. Schliesser (Eds.), *Methods in Empirical Prosody Research*, pp. 1-28, Berlin: De Gruyter, 2006.

[17] A. S. House, and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels", *Journal of the Acoustical Society of America*, *25*(1), pp. 105-113, 1953.

[18] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English", *Journal of the Acoustical Society of America*, *125*(1), pp. 425-441, 2009.

[19] J. P. Kirby and D. R. Ladd, "Effects of obstruent voicing on vowel F0: Evidence from 'true voicing' languages," *Journal of the Acoustical. Society of America*, 140(4), pp. 2400–2411, 2016.