White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/144545/

Version: Accepted Version

Proceedings Paper:
Chodroff, Eleanor Rosalie and Cole, Jennifer (2018) Information structure, affect, and prenuclear prominence in American English. In: Proceedings of INTERSPEECH 2018. INTERSPEECH 2018, 02-06 Sep 2018 Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. International Speech Communication Association, IND, pp. 1848-1852.

https://doi.org/10.21437/Interspeech.2018-1529

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Information Structure, Affect, and Prenuclear Prominence in American English

Eleanor Chodroff and Jennifer Cole

Northwestern University, Department of Linguistics, Evanston, IL, USA

eleanor.chodroff@northwestern.edu, jennifer.cole1@northwestern.edu

Abstract

The influence of information structure (IS: givenness, accessibility, newness and focus) on pitch accent assignment and acoustic prominence measures of prenuclear words was investigated for American English speech elicited through read production of mini-stories. Results showed a consistent pattern of accenting the initial content word in the sentence, supporting an analysis of prenuclear accent as structural, or 'rhythmic'. While no association was observed between IS and accent type (e.g., H*, L*, L+H*, L*+H), the acoustic-phonetic realization of prominence was modulated by information structure. In particular, words that carry contrastive focus generally showed more extreme f0 excursions relative to the average. In addition, there was a strong influence of speaking style or 'affect' on both pitch accent type and the acoustic-phonetic realization of prominence. Speakers were more likely to produce L+H* accents in a lively than a neutral speaking style. Differences in affect were also strongly reflected in f0 excursion, duration, and amplitude within the target word. Overall, this study indicates both linguistic (information structure) and paralinguistic (affect) influences on the phonetic implementation of prenuclear prominence, with varying influence of these two factors on the phonological assignment of prenuclear pitch accents.

Index Terms: prosody, prenuclear position, information structure, affect, speaking style, speech production

1. Introduction

The prosodic system of English, among other languages, makes a distinction between nuclear and prenuclear prominence. Nuclear prominence marks the structural head of a prosodic phrase, and is located on the basis of position in the prosodic phrase and on factors related to pragmatic meaning [1-5]. For example, in English the nuclear prominence occurs on the final content word in the prosodic phrase (1a, nuclear prominence in CAPS), or earlier on a word with contrastive focus (1b), or if the final content word is discourse-given (1c) [1-5]. The word with nuclear prominence receives an obligatory pitch accent, and the melodic type of the nuclear accent encodes pragmatic meaning related to information structure (IS; here including information status (=given/newness) and focus) [4], speaker attitudes [6] or interlocutors' mutual beliefs [7].

(1) a. Sam was afraid of the DOG.

b. {Speaker A: Were you afraid of the dog?}
Speaker B: SAM was afraid of the dog.

c. {On his way home, a dog barked at Sam}. He was AFRAID of the dog.

Prenuclear prominence and accentuation have received much less attention in prior research. In English, they have been described as optional [8], 'ornamental' [5], or rhythmic [9] (see also [10]). English prenuclear accents, especially those that are rhythmically licensed, do not reliably mark contrast (focus) or other information structural distinctions [10, 11]. Similar claims have been put forward for German [5], a language with a comparable prosodic system to English, though experimental studies show a consistent placement of prenuclear accents, even on textually given information in contrastive contexts [8, 12] or on (non-contrastive) topics in topic-comment structures [13]. However, the accents displayed subtle changes in peak scaling [8, 13] or peak alignment [13] which expressed differences in information structure.

Beyond some acoustic differences found in German, additional evidence that prenuclear prominence is not necessarily devoid of linguistic meaning comes from pronoun resolution [14]. Prominence on a prenuclear subject pronoun can signal a change in the interpretation of the referent. For example, in sentence (2a), the subject pronoun 'he' is likely to refer to the subject in the preceding clause, 'John'. When the subject pronoun is stressed, as in (2b), the intended referent becomes the object of the preceding sentence, 'Harry' (from [14]).

(2) a. John hit Harry, and then he hit Sarah.

b. John hit Harry, and then HE hit Sarah.

Nevertheless, meaningful prosodic manipulation is most notably associated with nuclear, as opposed to prenuclear position [4, 5, 7]. Prenuclear prominence is also distinct from nuclear prominence in perception. In English, listeners are less likely to rate words in prenuclear position as prominent [15], and trained transcribers are more likely to disagree on the accent status of prenuclear words [16]. In German, listeners show lower sensitivity and longer reaction times in prominence judgments of prenuclear accents compared to nuclear accents [17].

The aim of the present study is to gain greater insight into the patterning of prenuclear prominence and accentuation as produced by speakers of English. In particular, we investigated the effects of two potential influences on prenuclear prominence, namely IS and affect. While IS has been shown to modulate prominence in nuclear words, its effect on prenuclear words is relatively unclear. In addition, we also examined how speaking style or affect influenced prenuclear prominence. In the present study, affect corresponded to a neutral or a lively speaking style.

Previous research on German has identified some degree of correspondence between information structure and pitch accent type and scaling in both production and perception. This was noted above for prenuclear accents [8, 12, 13], but is also shown for nuclear accents in the same studies, and in [18]. In German, givenness has been found to correspond to a decrease in f0 within a pitch accent in prenuclear position, and deaccentuation of phrase-final words that would otherwise be assigned the nuclear accent. Focused words have also been found to be

longer than non-focused words [8]. The present study also examined these factors, but in English, and along a more graded scale of information structure [19].

The relation between affect and prosody has also investigated to some extent with the primary focus on intonation [20]. The prenuclear region has been implicated as a potentially relevant area for conveying differences in emotions like joy, anger, fear, among others. For instance, "happy" expressions tended to reach an f0 peak within what we might consider the prenuclear region, whereas utterances conveyed with "cold anger" or "panic fear" tended to have two local f0 maxima in the utterance, where the second f0 maximum exceeded the first. The current study investigated if and how the prenuclear position in particular can convey basic affective states, namely neutrality or liveliness.

If information structural factors exert a similar influence on the prominence and accent assignment of prenuclear words as they do with nuclear words, we expect a relationship between the IS of a word and its accent status. Specifically, we expect given words to be unaccented, and words that are not given to be accented, with a correspondence between the type of pitch accent and the level of newness/informativeness. We also expect a positive correlation between the level of newness/informativeness of a referent and its acoustic prominence. On the other hand, if prenuclear prominence is primarily ornamental, or a realization of phrase-level rhythmic structure, we expect no such correspondences.

2. Methods

2.1. Participants

Twenty native speakers of American English (12 female, ages 18-25) completed the experiment.

2.2. Materials

Participants were presented with four different mini-stories in text format on a computer screen. Each story consisted of two context sentences followed by a target sentence. The context sentences were prerecorded and played to the participant, to ensure that all participants would interpret the context sentences under the same prosodic conditions. Participants were asked to listen to the first two sentences and then read out only the last sentence of the story (the target sentence) in one of two styles: a natural, conversational style or a lively, storytelling style. In each story, the IS of the subject noun phrase (NP) in the target sentence was constructed to be 'given', 'accessible', 'new', or 'contrastive' by manipulating the content of the second context sentence (see Table 1) [19]. The target sentences of the four stories were as follows (the targeted subject NP and location of the expected prenuclear pitch accent is in bold): *'My **nephew** brought his famous lemonade'*, '***Chameleons** are a source of fascination for her*', '*The **bananas** were sold to the zoo'*, and '*The **superhero** turned out to be an idiot'*. In total, there were 16 unique story and IS condition pairings.

2.3. Experimental procedure

Each participant read all 16 unique story-condition pairings over four blocks. Each block contained two repetitions of four story-condition pairings for a total of 32 trials. In the first presentation, participants were instructed to read the stories in a neutral, conversational style. In the second presentation, participants were instructed to read the stories in a lively, storytelling style. The set of four trials consisted of one instance of each story and one instance of each IS condition, without repetition of the story-condition pairing over the course of the experiment. The presentation order of the sentences was quasi-randomized with the primary constraint that the same target sentence could not appear in succession between two blocks.

2.4. Measurements and statistical models

The subject noun in each target sentence was coded for accent type using Mainstream American English ToBI pitch accent labels [20]. The accent types encountered were: H*, L*, L+H*, and L*+H. Downstepped high tone accents were not considered since the target word was always the initial accentable word in the intonational phrase.

Acoustic measures of accent and prominence were taken from each speaker's target sentence productions, which were counterbalanced for target word, IS condition, and affect. F0 slope and range measures in semitones were calculated from f0 minima and maxima, manually located at the visible turning points of an f0 rise or fall from the accented syllable, or at the left edge of the accented syllable and right edge of the post-accentual syllable if no turning points were visible. For H* accents, the f0 maximum was subtracted from the following f0 minimum. For L*, L+H*, and L*+H accents, the preceding f0 minimum was subtracted from the f0 maximum. Acoustic correlates of prominence were measured in the duration (ms) and RMS amplitude (Pa) of the trochaic stress foot in the target word (i.e., the stressed syllable and the immediately following syllable).

The relationship between the categorical variables of accent type (ToBI labels: H*, L*, L+H*, L*+H), IS condition (given, accessible, new, contrastive), and affect (neutral, lively) was tested using Fisher's exact test. A linear mixed-effects regression model was fit to each acoustic correlate, centered on the grand mean. Each model contained fixed effects of accent type, IS, and affect, the full set of interactions between these effects. Random intercepts and slopes for affect were included for participant and target word; richer random effect structures led to non-convergence. Due to data sparsity, L* pitch accents were removed from the regression analyses. Accent type, IS, and affect were sum-coded with 'H*', 'given', and 'neutral' as the lowest level for their respective factors.

Table 1: *Example story. The contextual triggers are italicized, the target word is bolded, and the expected prenuclear and nuclear accents are underlined.*

Context	**Sentence**
Context 1	Last week we had our big annual picnic at the park.
Context 2a: *given*	My *nephew* came with his wife and kids, and we had a lot of fun despite the heat.
Context 2b: *accessible*	All of *my siblings* came, with their families, and we had a lot of fun despite the heat.
Context 2c: *new*	There were many hot dishes and desserts.
Context 2d: *contrastive*	*My sister* made a delicious chocolate cake.
Target	My **nephew** brought his famous lemonade.

3. Results

As our research questions concerned the status of prenuclear pitch accents, trials were excluded in which a prosodic phrase boundary was produced immediately following the subject NP, which placed the target word in nuclear position. 28 out of the 640 collected trials were excluded for this reason. Approximately equal numbers of trials from each IS condition and affect were excluded (6 given, 7 accessible, 8 contrastive, 7 new; 16 neutral and 12 lively).

3.1. IS, affect, and accent type

Figure 1 shows the number of target words with each accent type by IS condition and affect. One striking finding was that the prenuclear target word was accented in every trial, even in the given IS condition where deaccentuation was considered most likely. A further finding was that apart from L*, each accent type was observed in each condition. A Fisher's exact test (two-sided) revealed no significant association between accent type and IS condition ($p = 0.69$). The variation observed in the association of accent types and IS conditions for the aggregated data was also reflected in variation among speakers. Most speakers produced two or more accent types in the experiment, but no individual speaker showed a clear pattern in the association of accent type and IS.

There was a strong association between affect and accent type, notably in the increased production of L+H* for the lively affect. The overall association between affect and accent type was significant ($p < 0.001$). For all but two speakers, L+H* was more frequent in the lively affect than the neutral affect. While this ratio varied across speakers, L+H* was, in the aggregate, approximately 2 times more frequent in the lively speaking style than in the neutral style, whereas speakers were more likely to produce H* and L*+H in the neutral style.

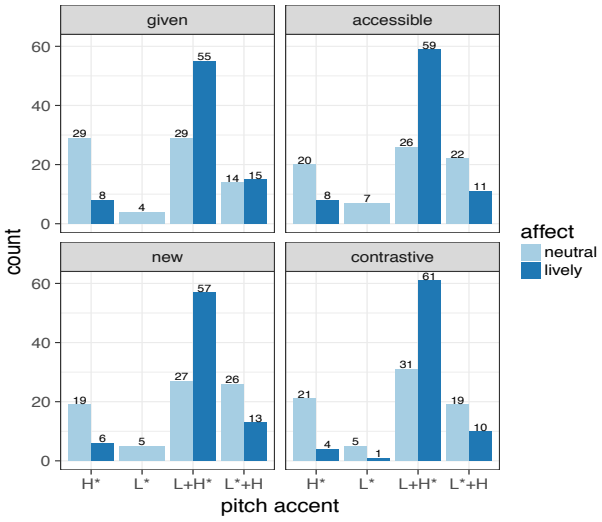

Figure 1: *Count of pitch accent type by IS condition and affect.*

3.2. IS, affect, and acoustic prominence

The acoustic-phonetic realization of prominence was analyzed in a series of linear mixed-effects models described in section 2.4. The f0 analyses included data from the first and last block of trials for each speaker,[1] whereas the duration and amplitude analyses were conducted on the full dataset. As in the previous analyses, target words were excluded if produced in nuclear position. In addition, L* accents were excluded due to their relative sparsity. This resulted in 288 trials for the f0 analyses and 590 trials for the duration and amplitude analyses.

For f0 slope, we observed significant effects of accent type, as expected ($\beta_{L+H*} = 0.018$, $t = 4.22$; $\beta_{L*+H} = 0.011$, $t = 3.85$).[2] In addition, the f0 slope of L+H* was significantly modulated depending on the information structure. Relative to the average, L+H* in the accessible IS condition had a significantly lower f0 slope, whereas L+H* in the contrastive IS condition had a significantly higher f0 slope ($\beta_{L*+H \times accessible} = -0.010$., $t = -3.50$; $\beta_{L*+H \times contrastive} = 0.009$, $t = 2.93$). In addition, there was a significant interaction between L+H* and affect, such that lively productions were produced with a higher f0 slope than the average ($\beta_{L*+H \times affect} = 0.009$, $t = 4.04$). The model for f0 range yielded the same pattern of significance and direction of effects as the model reported for f0 slope.

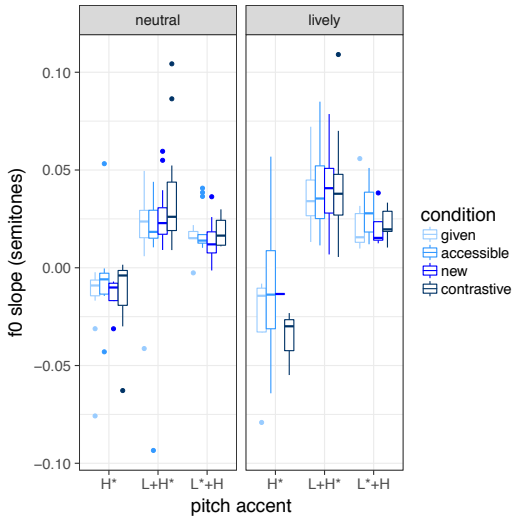

Figure 2: *F0 slope (semitones) by accent type across IS conditions within affect (panel)*

For both duration and amplitude of the trochaic foot (disyllable), we observed significant effects of affect (duration: $\beta_{affect} = 8.89$, $t = 2.17$; amplitude: $\beta_{affect} = 0.003$, $t = 4.63$), but no significant effects of accent type, information structure, or any interactions. Disyllables were both longer (Figure 3) and had increased amplitude (Figure 4) when produced in a lively speaking style compared to average.

[1] ToBI annotation and labeling of f0 maxima and minima were restricted to half of the data for this initial analysis.

[2] Effects and interactions with *t*-values greater than 2.00 were considered significant. The calculation of *p*-values for linear mixed-effects models is not trivial, and therefore not included in the standard lme4 implementation of these models [22]. Using an additional R package, lmerTest, we verified that under certain assumptions, the *p*-values of all reported results were less than 0.001 [23].

4. Discussion

The present study investigated the influence of IS and affect on prenuclear prominence in American English. We found no association between pitch accent type and IS, but found moderate influences of IS on measures of f0 excursion, particularly in the scaling of L+H* accents. Strong influences of affect were also observed on prenuclear pitch accent type and the phonetic implementation of pitch accents.

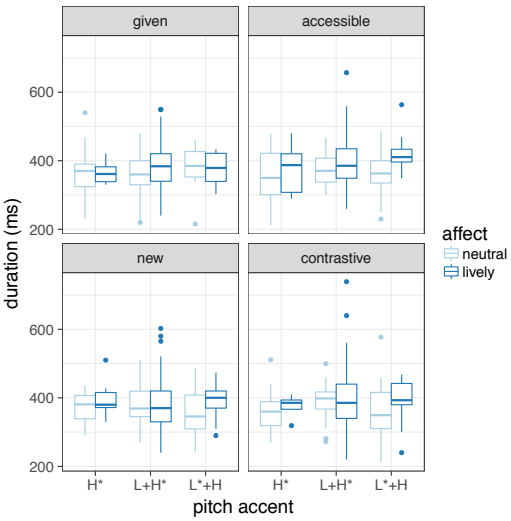

Figure 3: *Duration of the stressed and post-tonic syllables (milliseconds) by accent type across affect within IS condition (panel)*

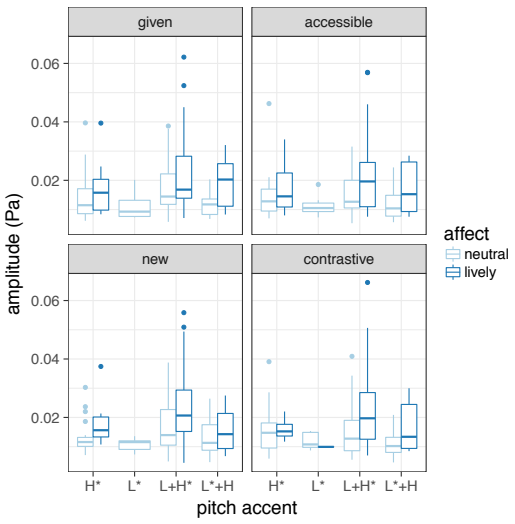

Figure 4: *RMS amplitude of the stressed and post-tonic syllables (pascals) by accent type across affect within IS condition (panel)*

The overall results lend support for a structural or rhythmic usage of prenuclear prominences in read American English. We found an exceptionless placement of prenuclear pitch accents on the initial content word of a sentence. This means that even textually given referents (which are usually deaccented if they occur sentence-finally) were consistently marked by a pitch accent in sentence-initial position. Consequently, differences in the prosodic marking of information status can at best be subtle.

The phonological assignment of pitch accent type on the prenuclear subject NP was not significantly influenced by IS, nor were any strong trends observed in the data. Interestingly, pitch accent assignment in prenuclear position varied significantly by the paralinguistic factor of affect. Namely, a lively affect corresponded to a significant increase in the use of L+H* on the prenuclear subject NP.

The phonetic implementation of prenuclear prominence was modulated both by IS and affect. From these data, it seems that the role of IS is limited to f0 effects, as no significant influence of IS was observed on the duration or amplitude of the disyllable stress foot in the target word. These findings are largely consistent with previous studies in German for IS [8], but counter to [8], we did not find a relation between focus and duration. The present findings should however be considered preliminary until confirmed in a larger dataset with greater diversity in the stories conveying IS, as well as segmentally and rhythmically more consistent target words. Differences in the syllable structure and sonority of the target words limited the extent to which more complex analyses could be conducted over the full f0 contour. Furthermore, the present study contained only four sets of stories. A follow-up study that addresses these limitations is currently underway.

The influence of affect on the phonetic implementation of prenuclear prominence was quite strong. Lively affect corresponded to more extreme f0 slopes, longer duration, and higher amplitude within the target words. The prenuclear position may be a prime location for conveying such paralinguistic cues, given the marginal role of the prenuclear region in encoding IS. It would also be beneficial for future work to examine whether the prenuclear position is privileged relative to the nuclear position for paralinguistic modulations of prominence. Our study only confirms that the prenuclear position is available for this purpose.

In addition to increasing the diversity in the stories and control over the phonological structure in the target sentences, we plan future studies to examine how phonological and phonetic variation in prenuclear accents influences listeners' perception of prominence and referential meaning.

5. Conclusions

Factors governing the phonological assignment of prenuclear pitch accents, as well as their phonetic implementation have received relatively less attention than those governing nuclear prominence. This study investigated the effects of IS and affect on pitch accent type and acoustic-phonetic measures of prominence in prenuclear subject noun phrases. Counter to claims that prenuclear accents may be 'ornamental', our findings suggest that affect significantly influences phonological assignment of pitch accents, and both IS and affect modulate acoustic-phonetic factors associated with prenuclear prominence.

6. Acknowledgements

We thank Stefan Baumann for inspiring many of the research questions and for his contribution to the experiment design. We also thank José Hualde, Suyeon Im, Janina Kalbertodt, and Jane Mertens for their contributions to the experiment design. We also thank Alaina Arthurs for her contribution to data processing.

7. References

[1] M.A.K. Halliday, *Intonation and Grammar in British English*. The Hague: Mouton, 1967.

[2] D. Bolinger, "Accent is predictable (if you're a mind reader)," *Language*, vol. 48, no. 3, pp. 633-644, 1972.

[3] D.R. Ladd, *The structure of intonational meaning: Evidence from English.* Bloomington: Indiana University Press, 1980.

[4] C. Gussenhoven, *On the grammar and semantics of sentence accents.* Dordrecht: Foris, 1984.

[5] D. Büring, "Intonation, semantics, and information structure," in Ramchand, G.; Reiss, C., editors. *The Oxford handbook of linguistic interfaces*. Oxford: Oxford University Press, 2007, pp. 445-473.

[6] J. Hirschberg and G. Ward, "The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English," *Journal of Phonetics*, vol. 20, pp. 241–251, 1992.

[7] J. Pierrehumbert and J. B. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," *Intentions in Communication*, pp. 271-311, 1990.

[8] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics*, vol. 36, no. 4, pp. 680-703, 2008.

[9] S. Calhoun, "The Centrality of Metrical Structure in Signaling Information Structure: A Probabilistic Perspective," *Language*, vol. 86, no. 1, pp. 1–42, 2010.

[10] C. Gussenhoven, "Does phonological prominence exist?" *Lingue E Linguaggio*, vol. 14, no. 1, pp. 7–24, 2015.

[11] C. Gussenhoven, "Sentential Prominence in English," *The Blackwell Companion to Phonology*, (September), pp. 2778–2806, 2011.

[12] S. Baumann, J. Becker, M. Grice, and D. Mücke, "Tonal and Articulatory Marking of Focus in German," Proceedings of the 16th International Congress of Phonetic Sciences, (Saarbrücken), pp. 1029-1032, 2007.

[13] B. Braun, "Phonetics and phonology of thematic contrast in German," *Language and Speech*, vol. 49, no. 4, pp. 451-493, 2006.

[14] A. Akmajian and R. Jackendoff, "Coreferentiality and Stress," *Linguistic Inquiry*, vol. 1, no. 1, pp. 124-126, 1970.

[15] J. Cole, Y. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence," *Laboratory Phonology,* vol. 1, pp. 425–452, 2010.

[16] D.R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.

[17] N. Jagdfeld and S. Baumann, "Order Effects on the Perception of Relative Prominence," *Proceedings 17th ICPhS,* Hong Kong, China, pp. 958-961, 2011.

[18] S. Baumann and M. Grice, "The intonation of accessibility," *Journal of Pragmatics*, vol. 38, pp. 1636-1657, 2006.

[19] A. Riester and S. Baumann, "The RefLex-scheme annotation guidelines," 2017.

[20] T. Bänziger and K.R. Scherer, "The role of intonation in emotional expressions," *Speech Communication*, vol. 46, pp. 252-267, 2005.

[21] M.E. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, "The original ToBI system and the evolution of the ToBI framework," in Jun, S.-A. (ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing.* Oxford: Oxford University Press, pp. 9–54, 2005.

[22] D. Bates, M. Maechler, B. Bolker, and S. Walker, "lme4: Linear mixed-effects models using Eigen and S4," *R package*, version 1, no. 7, pp. 1-23, 2014.

[23] A. Kuznetsova, P.B. Brockhoff, and R.H.B. Christensen, "Package 'lmerTest'," *R package*, version 2.0, 2015.