9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Cue-based annotation and analysis of prosodic boundary events

Alejna Brugos[1], *Mara Breen*[2], *Nanette Veilleux*[3], *Jonathan Barnes*[1], *Stefanie Shattuck-Hufnagel*[4]

[1] Boston University, USA, [2] Mount Holyoke College, USA
[3] Simmons College, USA, [4] Massachusetts Institute of Technology, USA

abrugos@bu.edu, mbreen@mtholyoke.edu, veilleux@simmons.edu, jabarnes@bu.edu, sshuf@mit.edu

Abstract

Prosodic categories, like other grammatical categories, are realized with wide variability, yet listeners interpret linguistic meaning with apparent ease. ToBI aims to capture the linguistically meaningful prosodic elements of utterances, but does not capture the variability in acoustic cues that the labeller (and listener) must interpret in order to assign distinctive categories. Despite a long history of empirical work demonstrating the importance of individual cues in signalling prosodic events, little previous work has explored how these cues might systematically combine, trade off and vary in magnitude across ToBI categories. Following recent proposals of cue-based annotation of segments, and disfluent and stuttered speech prosody, the current paper assesses the role of six timing, pitch, and voice quality cues in signalling boundary events in a corpus of independently ToBI-labelled fluent American English speech. Results demonstrate that each cue accounts for unique variance in break index level, and that break index levels increase with the number of cues present per token. The data further suggest that labeller uncertainty is more frequent for tokens that exhibit only a subset of the cues. Future work will investigate annotation of cues to prominence, as well as additional cues to boundaries and grouping.

Index Terms: prosodic boundaries, ToBI, phonetic cues, annotation, cue integration

1. Introduction

Phonological categories are realized with a high degree of variability in the signal, as has been well-documented for both segmental and prosodic elements. In the domain of segmental analysis, growing evidence shows that speakers systematically control aspects of speech below the level of contrastive linguistic constituents, and that listeners are sensitive to this variation. Likewise, in the study of prosodic events associated with prominence and boundary, there is well-documented variation of the cues that signal a prosodic category (see [1] for overview).

Phonological prosodic annotation systems such as ToBI (for **To**nes and **B**reak **I**ndices) [2,3] require labellers to interpret the signal to indicate the occurrence of prosodic categories of prominence (pitch accents) and boundary-related phenomena (phrase tones and boundary tones, and their associated level of disjuncture). While many individual prosodic events are straightforward to assign to a category, many more may leave the labeller uncertain as to which category was intended by the speaker. In boundary annotation, in particular, a labeller may be certain that there is a phrase-level boundary, but have uncertainty about the level of boundary. It is our observation, from years of annotation and prosodic annotation training, that the clearer tokens of prosodic categories show more and stronger cues in the signal. In contrast, harder-to-categorize tokens of prosodic events tend to show only a subset of the cues that are associated with established categories.

While trained labellers are generally able to make a category decision, the potentially systematic and informative variability in the signal is not captured by the use of a categorical label alone. Cole and Shattuck-Hufnagel (2016) [4] advocated for "the identification of individual cues to the contrastive prosodic elements of an utterance:"

> Cue specification provides a link between the contrastive symbolic categories of prosodic structures and the continuous-valued parameters in the acoustic signal, and offers a framework for investigating how factors related to the grammatical and situational context influence the phonetic form of spoken words and phrases. [4, p. 1]

Towards this goal, we propose the beginnings of an extension to prosodic annotation to explicitly label cues to prosodic categories. We adapt proposed systems for cue-based annotation of disfluent [5] and stuttered [6] speech, and expand this type of annotation to fluent speech. While we plan to address cue specification for prosodic prominence in future work, we begin our proposal with a set of boundary cues. Using a small corpus of American English speech, we compare a set of new cue annotations to independently produced ToBI labels, to illustrate how such an annotation system might shed light on the realization and perception of prosodic categories.

While phonetic cues to boundaries have been extensively discussed in the literature, and many studies have examined the role of cues for boundary perception (e.g. [7,8,9]) and automatic boundary detection (e.g. [10,11,12]), we do not yet know of a system designed for explicit prosodic cue annotation of corpora. This system is not intended to supplant categorical prosodic annotation systems, but to enhance them by showing the concrete ways in which speakers signal categories, and how these cues are interpreted by listeners. Indeed, in agreement with previous proposals [4, 13], we believe that the essential goals of prosodic transcription are to identify the contrastive linguistic categories of spoken utterances, and not merely the salient aspects of the surface form. On this cue-based view, the transcription of an utterance has two important levels: 1) the observable acoustic cues to the features of its constituents, and 2) the abstract sound categories that are signalled by those cues and that define the contrastive categories that underlie the utterance.

10.21437/SpeechProsody.2018-50

2. Background

2.1 Feature Cues to Segmental categories

In the segmental domain, there is evidence that language users systematically control aspects of speech below the level of contrastive linguistic constituents. Evidence comes from many phenomena, and a wide range of analyses (see [1] for an overview). Patterns of sub-allophonic variation often occur in prosodic boundary and accent contexts (e.g. [14]). The range and pervasiveness of phonetic variability across contexts is challenging to deal with using traditional units (phonemes, features or allophones), which suggests the usefulness of a lower level of description, i.e. the acoustic cues that signal distinctive features, and the quantitative values these cues can take on.

2.2 Annotation of prosodic categories

ToBI [2,3] is a system of prosodic transcription with which users aim to identify and label the categories of prosodic events. In ToBI conventions, based in large part on the autosegmental-metrical model of prosodic phonology [15], tonal events are considered to indicate prominence or boundary, and are categorized as either pitch accents or two levels of edge tone.

ToBI uses a tier-based system of annotation, time-aligned to the speech signal. ToBI's two main analysis tiers are the *tones* tier, to label pitch-related events of prominence and phrase-level boundaries, and the *breaks* tier, for an index of between-word disjuncture. These are used in conjunction with tiers for *words* (orthographic text), and *misc* (miscellaneous other signal information). In the process of ToBI annotation, a trained labeller interprets the acoustic cues (through listening and visual inspection) to identify the phonological categories represented by the prosodic events of the utterance. When cues to an event are ambiguous, however, the labeller may struggle to identify a category. Use of uncertainty markers and the *alternatives* tier [16] help to capture such cases, but do not directly identify the aspects of the signal that make category identification difficult.

ToBI labellers are required to determine a level of perceived disjuncture between each pair of words, labelled by a break index between 0 and 4. A break index of 4 is the largest labelled disjuncture, and indicates the end of a Full Intonational Phrase (IP). A 3 break indicates the end of an intermediate intonational phrase (ip). The standard phrase-medial word boundary is a 1 break. A 0 break indicates that two words have been tightly grouped via connected speech processes, such that the labeller does not perceive a word boundary. A break index of 2 was devised, "to mark cases of these two types of 'mismatch' between the subjective boundary strength and the intonational constituency." [2, p. 35] In practice, 2 breaks are used where a labeller perceives disjuncture that is greater than a word-level boundary, but not strong enough to merit a break index 3.

2.3 Cues to prosodic boundaries

In the area of prosodic boundary perception in American English, boundaries and their relative strength are marked by a variety of cues in the signal, including pre-boundary segmental lengthening (e.g. [17,18,19]), pauses (e.g. [20, 21,22]), edge tones (e.g. [2,7,15]), and cross-phrase pitch reset (e.g. [23]). (See [24] for a recent overview of pitch and timing boundary cue literature.) Other cues include voice quality cues, such as irregular pitch periods word-initially at the beginning of a phrase [25], or phrase finally [26].

While studies have shown that certain boundary cues (e.g. preboundary lengthening [19]) have a high degree of correlation with boundary strength, no single cue has been found to account for all of the variation in perceived level of grouping. When pitch cues are controlled, timing cues alone can signal boundary strength (e.g. [8,27]). However, when timing cues are controlled and weaker, pitch can also strongly cue boundaries [23,24,28,29]. Explicit labelling of these cues as they occur in naturally produced speech has the potential to guide research into this type of interaction.

3. Methods

In the current study, we begin the foray into explicit labelling of cues to prosodic structure with proposed labels for a constrained set of cues that are well established as boundary cues for American English. We examine the feasibility of using these labels to explicitly indicate the presence of boundary-related acoustic cues through comparison to independently-produced ToBI break index labels. We hypothesize that where boundary cues converge (i.e. tokens with multiple cues to the presence of a phrasal boundary), there will be greater perceived disjuncture, as operationalized by ToBI break index labels. In cases where only a subset of boundary cues are present, we predict that labellers are not only likely to perceive weaker boundaries, but also to have uncertainty about the level of boundary.

The corpus used to test this cue labelling was used in study 2 of Breen *et al.,* 2012 [30]. It consists of eight files, comprising roughly six minutes of speech: 178 seconds professionally read and 181 seconds of spontaneous speech, from the Boston Radio News and CallHome corpora, respectively. The files were produced by seven individual speakers, roughly balanced between male and female speakers, and contained a total of 1076 words. All files were independently labelled by 4 expert labellers with Mainstream American English ToBI conventions, as described in [30].

An expert prosodic labeller (the first author, who was not among the ToBI labellers for this study) independently labelled the eight files for boundary cues using the 6 labels described below. This proposed system takes as its base the prosodic disfluency labelling system of Brugos & Shattuck-Hufnagel (2012) [5], which in turn is modelled after Arbisi-Kelm's 2006 [6] system for labelling stuttered speech. Two labels (pr and ps) are based on timing cues, and are adapted from the disfluency and stuttering annotations [5,6]. Four new labels are proposed for pitch and voice quality cues.

pr: pre-boundary lengthening, noticeable prolongation of the segments of the final syllable of a word
ps: pause, a silent interval following a word
et: a local change in f0 across the final syllables of a word that leads a labeller to consider a boundary (edge tone)
jp: a change in f0 across word boundaries, whether a step up or a step down (reset, or pitch "jump")
gf: irregular pitch periods in the final syllable of a word (final glottalization)
gi: irregular pitch periods at the beginning of a vowel-initial word (initial glottalization)

Cue labels were placed on a point tier, time-aligned to the end of each orthographic word. While several cues may be following that word (namely pause, pitch jump, and word-initial glottalization), for ease of processing, all cues were labelled at word end. Labels are used in sequences delimited by a period. Therefore, each word was labelled with between 0 and 6 cues from this identified set (e.g. only **pr** or **pr.et.gf.ps.jp.gi**). No judgement was made about the presence or size of boundaries, nor of perceived prominence.

Of the 1076 words in the data set, 649 words were labelled with 1 or more cues. All files were annotated using Praat [31]. Labels from all of the independently labelled TextGrids were automatically extracted for analysis.

4. Results & Analysis

Our first analysis addresses the question of whether the proposed cues map reasonably onto the break index categories used by coders. A first summary of the data, in Table 1. shows correspondence between the number of cues labelled for a given word, and disjuncture, operationalized here as break index agreed upon by 2 or more of the 4 ToBI labelers. The table shows that tokens with an increased number of cues tend to be labelled with higher break indices.

Table 1: *Break indices are where 2 or more labellers agreed*

	Common Break Index				
	0	1	2	3	4
Number of Cues					
0	0	428	2	1	1
1	0	195	9	19	4
2	0	85	12	49	15
3	1	39	6	36	31
4	0	8	4	18	52
5	0	0	2	4	44
6	0	0	0	0	11

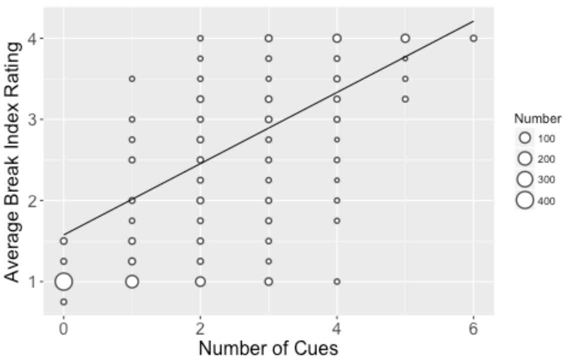

Fig. 1. *Average break index rating by number of cues per token. Size of dot indicates number of tokens.*

We used linear mixed effects regression to determine whether break index ratings increased with increases in the number of cues to a boundary. Here, we assigned each word the average of the break indices of the 4 labellers for that token, giving a continuous disjuncture score between 0 and 4. To account for possible differences across speakers, we included Speaker as a random effect. We also tested whether a random slope of Cue Number over Speaker improved model fit, but this model failed to converge, so we removed this parameter. In the final model, Cue Number was a significant predictor of average break index rating, B = 0.53, SE = 0.01, t = 42.84, $p < .05$ (Figure 1).

To determine whether the individual cues account for unique variance in signaling boundaries, we used linear mixed effect regression to predict the average break index from the six cues. Each cue was entered as a fixed effect in the model. In order to account for potential differences across speakers, we included Speaker as a random effect. We also tested whether random slopes for each cue over Speaker improved model fit. The full model did not converge, so we removed the random slope of the cue that accounted for the least Speaker variance (**jp**), and the model converged. The fixed effects of the final model are presented in Table 2 and Figure 2.

Table 2. *Fixed effects in the final model predicting average break index rating from each cue.*

	Estimate	SE	t	p-value
(Intercept)	0.96	0.03	37.30	0.00
pr	0.62	0.10	5.95	0.00
ps	1.23	0.12	10.46	0.00
et	0.45	0.11	4.06	0.01
jp	0.13	0.05	2.50	0.01
gf	0.37	0.08	4.35	0.01
gi	0.26	0.13	1.96	0.10

Five of the six cues independently and significantly increased the average break index rating. The sixth cue was a marginally significant predictor of increased break index rating. This result demonstrates that each of the cues in the current study is being perceived by coders as a cue to boundaries. For example, pause has a large effect on increasing average break index. This also demonstrates that the cues are weighted differently. Although the presence of initial glottalization does not significantly increase break index rating, it should be noted that the presence of this particular cue is limited to cases where a word following the labelled disjuncture begins with a vowel.

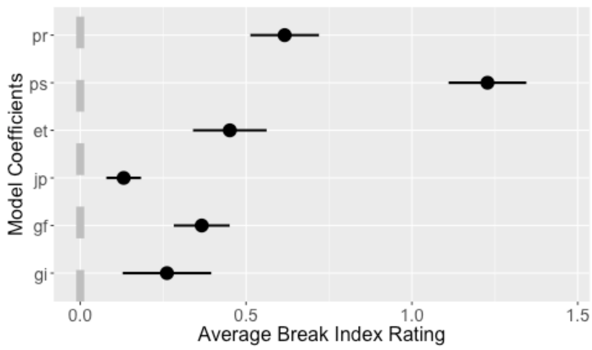

Figure 2. *Coefficients for the fixed effects in the model assessing the influence of each cue on average break index.*

Finally, we assessed the relationship between number of boundary cues and labeller uncertainty. A break index was considered uncertain if any one of the four ToBI labellers included the '?' or '-' diacritic in their label. Table 3 shows the frequency with which any labeller indicated uncertainty by the number of cues labelled for each token. The proportion of uncertain labels was higher when there was a smaller number of cues to a boundary than when there were either no cues, or many, X^2 (df = 6, N = 1076) = 67.98, p < .01. Uncertainty about level of disjuncture was more likely when there were 2, 3 or 4 cues. When there were 5 or 6 cues, there was almost no uncertainty. Similarly, when 0 or only 1 cue to potential boundary was present, uncertainty was infrequent.

Table 3: *Cases of labeller uncertainty by cue number.*

		Cue Number						
		0	1	2	3	4	5	6
Uncertain	0	427	210	134	94	70	49	11
	1	5	17	27	19	12	1	0
%		1.2	7.5	16.8	16.8	14.6	2	0

5. Discussion

As predicted, there was a strong correlation between number of annotated boundary cues, and perceived disjuncture, as operationalized as an averaging of break index label from four independent labellers.

This type of cue annotation combined with categorical labelling has the potential to offer insight into how cues interact in perception, through combination and cue trading: the absence and/or weakness of one cue to a given category may be compensated for by the presence and/or strength of a different cue. Such cases can be seen in our data, such as when ToBI labellers chose a 4 break index, but only 1 or 2 boundary cues were in evidence, but of very large magnitude (e.g. very long prolongation). This is in contrast to more straightforward example, when all 6 boundaries were present (and where all labellers agreed on a 4 break). Two such examples are shown in Figure 3.

While at this stage of labelling we focus on the presence of a cue perceived by a trained labeller, it is clear that these cues also vary in their magnitude. For example, pauses and pre-boundary lengthening vary in their degree, and edge tones and cross-phrase reset vary in the size of the f0 excursion or difference, and such differences of magnitude affect perceived grouping (e.g. [19, 23, 32]).

Future work will consider cues from amplitude [11] and phrase-initial onset strengthening [33]. We hope also to investigate labelling non-local and potentially overlapping patterns that cue grouping, such as cross-phrase global pitch patterns (e.g. "schemas" [24,34] or declination and pitch accent scaling relations [35,36]) as well as other non-local grouping cues (e.g. [37,38,39]). A similar approach labelling cues to prominence, could guide investigation into how prominence and boundary cues overlap and interact.

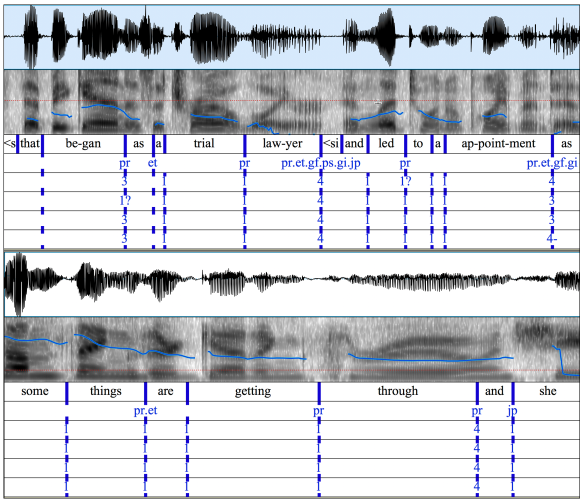

Figure 3: *Two examples of labelling. Tiers are* words, cues, *and then* break indices *from each of 4 ToBI labellers. The top shows an example where all 6 boundary cue labels were marked (on the word "law-yer", and where all ToBI labellers chose break index 4. The bottom example shows a case where only pr was marked (on the word "through",) but all ToBI labellers marked 4.*

Annotating prosodic cues, in addition to or as an intermediate step towards categorical labels, has the potential to help in the study of prosody of lesser-studied languages and varieties, an approach endorsed by those developing IPrA [40]. Boundary cues such as we have annotated could be used in conjunction with the proposed tonal markers of [40] and [41].

6. Conclusions

Using a transcription system that explicitly labels the cues to prosodic categories both captures aspects of the variability of the signal, and sheds light on the systematicity of the variability. This study provides a proof of concept that when people label boundaries, and when people label cues, those two sets of labels are usefully related.

7. Acknowledgments

Work for this study was in-part supported through NSF grant #1451663. We are grateful to the ToBI labellers of Breen *et al.*, and to Sun-Ah Jun for inspiration through many discussions.

8. References

[1] Cole, J. & Shattuck-Hufnagel, S. (Forthcoming) "Quantifying phonetic variation: landmark labelling of imitated utterances" De Gruyter Mouton, in *Rethinking Reduction*, Cangemi et. al (ed).

[2] Beckman, M., & Ayers Elam, G.. (1993/1997). *Guidelines for ToBI Labelling (Version 3, March 1997)* Copyright (1993) the Ohio State University Research Foundation.

[3] Beckman, M., Hirschberg, J. & Shattuck Hufnagel, S. (2005). "The Original ToBI System and the Evolution of the ToBI Framework." In *Prosodic Typology: the Phonology of Intonation and Phrasing,* edited by S. A. Jun: 9–54.

[4] Cole, J., & Shattuck-Hufnagel, S. 2016. "New methods for prosodic transcription: Capturing variability as a source of information." *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1): 8, pp. 1–29.

[5] Brugos, A., & Shattuck-Hufnagel, S. 2012. "A proposal for labelling prosodic disfluencies in ToBI." Poster presented at Advancing Prosodic Transcription for Spoken Language Science and Technology, July 31, 2012, Stuttgart, Germany.

[6] Arbisi-Kelm, T. R. (2006). *An intonational analysis of disfluency patterns in stuttering* (Dissertation, UCLA).

[7] Streeter, L. (1978). "Acoustic Determinants of Phrase Boundary Perception." *The Journal of the Acoustical Society of America* 64: 1582-1592.

[8] Scott, D. (1982). "Duration as a Cue to the Perception of a Phrase Boundary." *Journal of the Acoustical Society of America* 71, no. 4: 996–1007.

[9] Krivokapic, J., & Byrd, D. (2012). "Prosodic Boundary Strength an Articulatory and Perceptual Study." *Journal of Phonetics* 40, no. 3 (May 1, 2012): 430–442.

[10] Wang, M. Q., & Hirschberg, J. (1992). "Automatic classification of intonational phrase boundaries." *Computer Speech & Language*, 6(2): 175-196.

[11] Choi, J. Y., Hasegawa-Johnson, M., & Cole, J. (2005). "Finding intonational boundaries using acoustic cues related to the voice source." The *Journal of the Acoustical Society of America*, 118(4): 2579-2587.

[12] Rosenberg, A. (2010). "Autobi--A tool for automatic tobi annotation." In *Eleventh Annual Conference of the International Speech Communication Association*: 146-149.

[13] Frota, S. (2016). "Surface and Structure: Transcribing Intonation within and across Languages." *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1): 7, pp. 1–19.

[14] Turk, A., & Shattuck-Hufnagel, S. (2014). "Timing in talking: what is it used for, and how is it controlled?". *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1658), 20130395.

[15] Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation* (Unpublished Ph.D. thesis). Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

[16] Brugos, A., Veilleux, N., Breen, M. & Shattuck-Hufnagel, S. 2008. "The Alternatives (Alt) tier for ToBI: advantages of capturing prosodic ambiguity." In *Proceedings of Speech Prosody*. Campinas, Brazil.

[17] Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). "Segmental Durations in the Vicinity of Prosodic Phrase Boundaries." *The Journal of the Acoustical Society of America* 91: 1707–1717.

[18] Price, P., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991). "The Use of Prosody in Syntactic Disambiguation." In Speech and natural language: proceedings of a workshop, Pacific Grove, CA, February, 1991: 372-377.

[19] Yoon, T. J., Cole, J., & Hasegawa-Johnson, M. (2007, August). "On the edge: Acoustic cues to layered prosodic domains." In *Proceedings of ICPhS* (pp. 1264-1267).

[20] Lehiste, I. (1973). "Phonetic disambiguation of syntactic ambiguity." *The Journal of the Acoustical Society of America*, 53(1): 380-380.

[21] Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech* (No. 3). Harvard University Press.

[22] Krivokapic, J. (2007). "Prosodic planning: effects of phrasal length and complexity on pause duration." *Journal of Phonetics*, 35: 162–179.

[23] Brugos, A., & Barnes, J. (2012). "Pitch Trumps Duration in a Grouping Perception Task." *25th Annual CUNY Conference on Human Sentence Processing,* New York, NY. May, 2012.

[24] Brugos, A. (2015). *The interaction of pitch and timing in the perception of prosodic grouping* (Doctoral dissertation, Boston University).

[25] Dilley, L., Shattuck-Hufnagel, S. & Ostendorf, M. (1996). "Glottalization of Word Initial Vowels as a Function of Prosodic Structure." *Journal of Phonetics* 24 (November 5, 1996): 423–444.

[26] Redi, L. & Shattuck-Hufnagel, S. (2001) "Variation in realization of glottalization in normal speakers." *Journal of Phonetics* 29: 407- 429.

[27] Wagner, M., & Crivellaro, S. (2010). "Relative Prosodic Boundary Strength and Prior Bias in Disambiguation." In *Proceedings of Speech Prosody 5*, Chicago, IL, May, 2010.

[28] Beach, C. (1991). "The Interpretation of Prosodic Patterns at Points of Syntactic Structure Ambiguity: Evidence for Cue Trading Relations." *J. of Memory and Language* 30, no. 6: 644–663.

[29] House, D. (1990). *Tonal Perception in Speech*, Lund: Lund University Press.

[30] Breen, M., Dilley, L. C., Kraemer, J. & Gibson, E. (2012). "Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch)." *Corpus Linguistics and Linguistic Theory* 8: 277–312.

[31] Boersma, P. & Weenink, D. (2016). "Praat, a System for Doing Phonetics by Computer".

[32] Krivokapic, J. (2007). *The Planning, Production, and Perception of Prosodic Structure.* Doctoral Dissertation, University of Southern California.

[33] Fougeron, C., & Keating, P. A. (1996). "Articulatory strengthening in prosodic domain-initial position." *UCLA Working Papers in Phonetics*: 61-87.

[34] Brugos, A. & Barnes, J. (2015). "Intonational schemas, perceived grouping, and distortions of perceived duration." *Proceedings of the 18th ICPhS*. Glasgow, UK: the University of Glasgow.

[35] Ladd, D. R. (1988). "Declination 'Reset' and the Hierarchical Organization of Utterances." *The Journal of the Acoustical Society of America* 84: 530.

[36] Féry, C., & Truckenbrodt, H. (2005). "Sisterhood and Tonal Scaling." *Studia Linguistica* 59, no. 3: 223–243.

[37] Kentner, G., & Féry, C. (2013). "A New Approach to Prosodic Grouping." *The Linguistic Review* 30, no. 2.

[38] Morrill, T., Dilley, L., & McAuley, J. D. (2014). "Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure." *Journal of Phonetics*, 46: 68-85.

[39] Byrd, D., Krivokapic, J. & Lee, S. (2006). "How Far, How Long: on the Temporal Scope of Prosodic Boundary Effects." *The Journal of the Acoustical Society of America* 120: 1589 1599.

[40] Hualde, J., & Prieto, P. (2016). "Towards an International Prosodic Alphabet (IPrA)." *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1).

[41] Jun, S.-A. & Fletcher, J. (2014) "Methodology of Studying Intonation: From Data Collection to Data Analysis", in Sun-Ah Jun (ed.) *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford University Press.