Revisiting Paratone Prosodic Features with the EIIDA corpus

Nicolas Ballier[1], *Adrien Méli*[1], *Taylor Arnold*[2], *Alice Henderson*[3]

[1]Université de Paris, CLILLAC-ARP, France
[2]University of Richmond, USA
[3]Université Grenoble-Alpes, LIDILEM, France

nicolas.ballier@u-paris.fr, adrien.meli@gmail.com, tarnold2@richmond.edu, alice.henderson@univ-grenoble-alpes.fr

Abstract

This paper discusses the prosodic properties of the paratone, the oral paragraph of speech. Using manually annotated paratones and automatic prosodic labels on the EIIDA corpus, we re-examine the claims proposed by Tench (1996). We show that rhythmic cues to signal paratone boundaries seem to be more reliable than absolute pitch values.

Index Terms: paratones, INTSINT, clustering, declination

1. Introduction

The number and variety of spoken corpora in English for Specific Purposes (ESP) and English for Academic Purposes (EAP) continues to increase [1, 2, 3, 4, **?**], with university lectures having received the most attention [5]. Yet spoken corpora on EAP offering more than orthographic transcription are rare, with the prosodic transcription of the Hong Kong Corpus of Spoken English (HKCSE) being a notable exception [6] . This is problematical because treating spoken corpora like written text misses out on what Pickering and Byrd term the "acoustical realizations" of authentic spoken discourse [7]. Our goal is to analyse a prosodically annotated ESP corpus to test Tench's [8] claims about the prosodic properties of the paratone, the oral paragraph of speech. The aim of this paper is to test the validity of a proposal that paratones, or 'phonological paragraphs', have six defining features [8]:

- "The high pitch on the onset syllable of the initial intonation unit. [claim 1]
- The relatively high 'baseline' of that initial unit; this means that the low pitches are relatively high, compared to the low pitches in the final unit of the paragraph. [claim 2]
- There is a gradual lowering of that baseline until the final unit is reached. [claim 3]
- The depth of fall in the final units is the lowest in the whole paragraph. [claim 4]
- There is usually a slowing down process in the final unit. [claim 5]
- There is a longer pause than is normally allowed between intonation units." [claim 6]

Basing our analysis on the EIIDA corpus (*Études Interdisciplinaires et Interlinguistiques du Discours Académique* [1], *i.e.* Interdisciplinary and Cross-linguistic Academic Discourse), we will test these four pitch-based and two rhythm-based claims using automatic annotation of prosody and syllable peaks.

The rest of the paper is organised as follows. Section 2 defines the paratone and previous research. Section 3 presents the EIIDA corpus and how we operationalised Tench's claims. Section 4 presents our results and our final section discusses and concludes.

[1]The EIIDA corpus (Etudes Interdisciplinaires et Interlinguistiques du Discours Académique, i.e. Interdisciplinary and Cross-linguistic Academic Discourse) is one of the first multilingual spoken corpora of specialized academic language. See English: https://corpora.aiakide.net/?c=EIIDA-en, French: https://corpora.aiakide.net/?c=EIIDA-fr

2. Previous Research

Paratones, "structural units of spoken discourse which take the form of 'speech paragraphs'" [9] or "phonological paragraphing" [8], have been postulated for discourse analysis [10, 9], though not always included in the various descriptions of the prosodic hierarchy [11][2].[14] defines paratone as a discourse unit 'coextensive with a stretch of discourse presented by a speaker as forming a unit with a single topic'. Within discourse analysis, paratones have been used as a phonological pragmatic discourse marker as by [15], in the wake of [16] study. [3]

The notion of paratone has been used to investigate non-native speech. [18] tried to replicate expected boundaries in a text read by Spanish, Japanese, and Thai learners of English as compared to native realisations. The topic function of the paratone was properly identified by the Spanish and native groups : "pitch range was significantly higher ([respectively 29 Hz and] 22Hz) when the sentence appeared in paragraph-initial position, than when it appeared paragraph-medially". [19] discusses discourse units and intonation units by analysing declination in semi-tones, while [20] discuss pedagogical implications of paratone realisations by learners. In English Foreign Language teaching (EFL), several studies have tried to approach the need for recognition, and L2 realisation of boundaries for intonation units. Within Brazil's communicative framework [21], [22] discusses "Intonational paragraphs in L1 and L2 academic discourse", proposing a tentative correlation between boundary features and "co-occurring structural features", where the highest unit is a "sequence chain" defined as a "semantically coherent unit with a unifying topic, an introductory topic expression and bounded by prospective and retrospective markers; may coincide with pauses of 0.8 seconds or longer".

To the best of our knowledge, few experiments have been carried out on automatic detection of paratones. The spoken content-based audio navigation (SCAN) as detailed in [23] allowed for paratone detection, but the system does not seem to be accessible any more. Automatic detection of prosodic

[2]For an alternative representation of a prosodic hierarchy inspired by [12] see [13].

[3]See [17] for an extensive discussion of the conceptions of paratones as opposed to 'chain sequences'.

units does not necessarily imply the recognition of paratone as a unit. Investigating macrostructures in business phone calls [24] show the relevance of discourse units such as openings, transactions, pre-closures and closures. [25] in their multidimensional analysis of L2 suprasegmental proficiency include paratone measures such as paratone pause length, “average F0 level of paratone-initial pitch choices (usually high pitch) and the average F0 of paratone termination choices (usually low pitch)”; paratone boundaries are essentially defined as “low terminations followed by high-key resets”, which provides the number of low termination tones. More recently, prosodic detection of discourse segments has moved to ‘topic shifting’, a possibly more elusive concept but more crucial for information retrieval [26]. [27] investigated the repetition of similar sequences in different topic transition types using pseudo-normalised duration.

3. Corpus and Methods

3.1. Data

The EIIDA corpus is one of the first multilingual spoken corpora of specialized academic language. It was the main deliverable of a 2012-17 project coordinated by Shirley Carter-Thomas and Jeanne-Marie Debaisieux, from the LATTICE (UMR 8094) research group at the Ecole Normale Supérieure and Université Paris 3 Sorbonne Nouvelle.

EIIDA can be used to carry out comparative linguistic analyses on written and spoken academic discourse (research articles vs. conference presentations). It is in three languages – English, French, Spanish - thus facilitating the analysis of the impact of writers’/speakers’ linguistic culture on these two modes of communication. Furthermore, as the corpus has written and spoken forms of discourse from geochemistry and linguistics, disciplinary comparisons can be established. The spoken corpus corresponds to roughly 20 hours of audio recordings (300,000 tokens). Only recordings by native English speakers were used in the current study. The two manually annotated talks of 24 and 28 minutes are taken from a conference in linguistics.

3.2. Methods

This subsection describes the different methods followed to assess pitch analysis, paratone duration and rhythm measures in order to test Tench’s claims.

3.2.1. Automatic Prosodic Labels

The automatic prosodic labels for pitch were obtained using INTSINT (International Transcription System for Intonation), [28] using SPPAS [29] and a routine to split the talk recordings into smaller files. INTSINT consists in 8 labels representing the annotation: T (Top), M (mid), B (bottom), H (Higher), L (Lower), S (Same), U (Upstepped) and D (Downstepped). These labels are likely to operationalise the first four claims by checking the pitch value (in Hertz) associated with the pitch targets. To operationalise the detection of rhythmic changes, following DeJong & Wempe’s algorithm [30], we extracted syllable peaks and used their algorithm to count syllable peaks for 2 seconds in initial or in final position of the paratone. To normalise duration for rhythm, we compared raw measurements for 2 seconds and 15% of normalised initial/final duration.

3.2.2. Manual Annotation of Paratones

Our next step was to manually analyse plausible candidates for paratone boundaries in the English recordings, with respect to the SPPAS-generated annotations. Catering minimally for gender variation, the files of a male and a female speaker were annotated for paratone boundaries. Our hypothesis is that the initial pitch reset of paratones is likely to be expressed by a T (top value) followed by a lower value (L, B or D). The sequence should also correspond to final declination at the end of the preceding paratone (L, D or B being the expected labels). The underlying question is whether it is possible to detect paratone boundaries based on the observed MOMEL/INTSINT features produced by the Praat plugin [31]. Once agreement was reached between our two annotators [4] , we listed the candidates and searched the other recording’s INTSINT vectors (tiers) for the corresponding patterns.

4. Results

4.1. Capturing Pitch Resets

Regarding Tench’s claims, the highest pitch targets for the label T (Top) did not correspond to paratone initial boundaries (claim 1). Nor was the final B (bottom) the lowest of the B pitch targets (claims 2 and 4). The expected gradual baseline was not observed (claim 3), as can be seen on Figure 1.

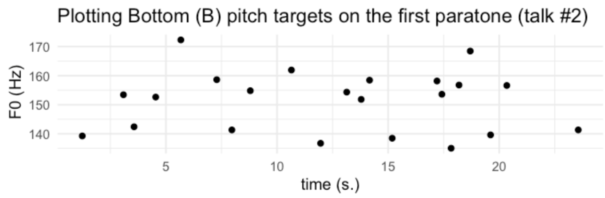

Figure 1: *F0 values of pitch targets B in the first paratone of talk #2*

Although detection based on absolute levels of pitch has been fruitless, [32] reports success in capturing the detection of pitch resets as the prediction of recurrent INTSINT patterns, finding that 3gram clusters of INTSINT pitch targets correlate with paratone boundaries. Figures 2 and 3 show that typical INSTINT patterns can be observed in initial and final position of our manually annotated paratones. The four panes feature the probability density functions of the initial and final patterns along the normalized duration of all paratones. The y-axis measures the probability density, while the x-axis continuously measures how far into the paratones the patterns can be found, with 0 marking the beginning, 0.5, half of the duration, and 1, the end, of the paratones. In order to clarify the findings, the bandwidth used to generate those figures was twice the default one in ggplot2 [33]. Figure 2 in particular shows the probability distribution across time of all the initial patterns, i.e. “TTL”, “MTL”, “DTL” and “TLS”. The higher probability density towards the beginning of paratones is clearly visible, and can be seen as evidence of the higher relative likelihood that these patterns occur predominantly towards the initial boundaries of paratones. Looking at each pattern more specifically, however, noticeable differences in probability density appear : “DTL” and “TLS” feature higher relative likelihoods to occur halfway through paratones, while “TTL” and “MTL” have higher rela-

[4]Since the two annotators did not find the same number of paratones, we cannot report a kappa agreement but we report an evaluation of mutual precision . Annotator1 agreed with 82% of Annotator2’s 90 boundaries. Annotator2 agreed with 63% of annotator1’s 113 boundaries.

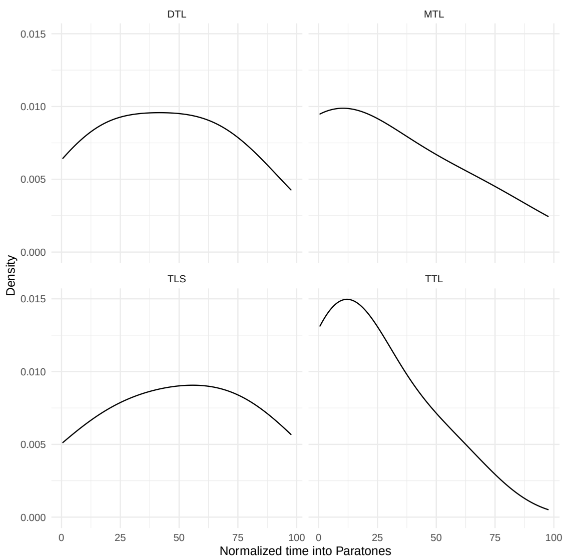

Figure 2: *Initial Candidates (normalised duration) for paratone signatures*

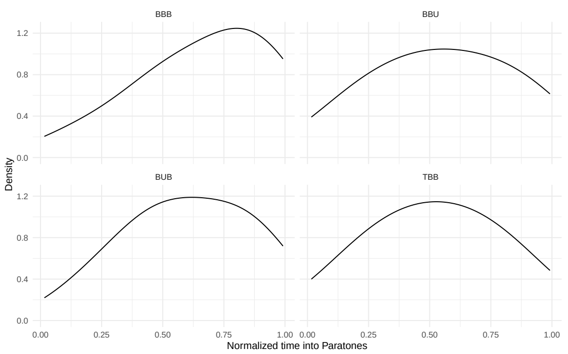

Figure 3: *Final Candidates (normalised duration) for paratone signatures*

tive likelihoods towards the beginning of paratones. "TTL" in particular seems to be more likely to occur exclusively initially.

Figure 3 shows the probability distribution of all final patterns, namely "BBB", "BBU", "BUB" and "TBB". Taken collectively, these patterns have a higher relative likelihood to occur in the third quarter of the duration of paratones - towards the end, then, but not exactly at the end. This trend can once again be explained by looking into detail at the probability distributions of each final pattern. Just like with initial patterns, final patterns seem to fall into two categories: "BBU" and "TBB" on the one hand display higher relative likelihoods to occur medially; "BBB" and "BUB" on the other have higher probability densities in final position. "BBB" shows a much clearer tendency to occur very close to the final boundary, in a way similarly exclusive to "TTL" for initial patterns. Further research is of course needed to examine the extent to which these findings are extendable, and whether other patterns display similar position-dependent behaviours. Although our results are consistent with cues of finality, they should not be construed as being absolute values, whereby the lowest pitch target would inevitably correspond to the end of a paratone. Claim 4 needs to be qualified: "The depth of fall in the final units is the lowest in the whole paragraph". As shown in Figure 4, B pitch targets can be found several times in a paratone and the last pitch target is not necessarily the lowest.

4.2. Capturing Deceleration

Overall, our investigation over the 82 paratones of our recording validates claim 5 ("There is usually a slowing down process in the final unit."). Figure 4 shows a deceleration of speech that we captured by measuring the decrease in syllable counts between the initial interval (2s) and the final interval (2s) of the paratones. The importance of the phenomenon is reflected in the number of paratones that follow this pattern (y axis). Apart from two outliers, a deceleration can be observed, whether we take into consideration normalised duration or not.

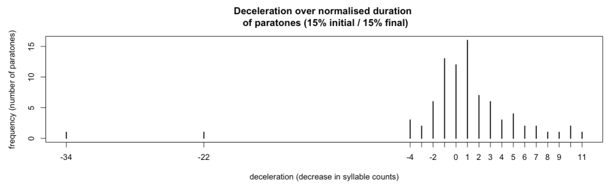

Figure 4: *Deceleration between final and initial position of paratones in talk #2*

The intuition that rhythmic cues could signal paratone boundaries seems to be validated. However, in Tench's terms, it seems that slowing down should be a cue for finality, with no specific prediction for initial boundaries. We tried reverse engineering to predict paratone boundaries on the basis of these observations. With our annotation scheme, the hypotheses are operationalised as follows: a) the lowest number of syllable peaks per a certain duration may signal right (final) boundaries; b) the highest number of syllable peaks per a certain duration may signal left (initial) boundaries. To our surprise, for talk #16, only 5 out of the 82 candidates for paratone right boundaries were detected, with a maximum of six syllable peaks over two seconds. Hypothesis b) was confirmed with 75% of paratone boundaries detected, taking as criterion the highest number (12) of syllable peaks over an interval of two seconds. The accuracy drops to 53% for intervals of seven syllable peaks.

4.3. Inter-paratone duration

Figures 5 and 6 represent the boxplots of the duration of the paratones and the boxplots of inter-paratone break duration. (Note the 8sec outlier for page turning and clapping sessions). The mean of our inter-paratone intervals is much smaller than baseline expectations (e.g. 200 ms according to [34]) so that claim 6 ("There is a longer pause than is normally allowed between intonation units.") does not quite hold. For talk #16, only 5 of the 74 longest silent intervals actually correspond to paratone boundaries. These results seem to contradict previous assumptions that "the most consistent paratone-final marker is the long pause, normally exceeding one second" [9].

5. Discussion

According to [35] 'Only a few studies have looked at global patterns in spontaneous speech', giving [24] and [36] as examples.

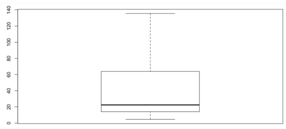

Figure 5: *Duration (s.) of paratones in talk #2*

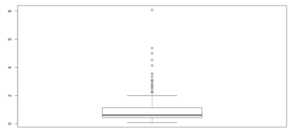

Figure 6: *Inter-paratone break duration (s.) in talk #2*

5.1. Major and Minor Paratones?

[14] entertains the division between major and minor paratones. If we assume this distinction, examples of those minor paratones can be found in rhetorical attempts to capture audience attention towards the end of the paratone (“relaunch signals”). The existence of two rhetorical devices has considerably hindered agreements between the two annotators.

Paratone headers? Some speakers gave a title to each subdevelopment, a bit like the header of the paratone, which has resulted in the existence of very short sections just before the initial Paratone. This is similar to the head or pre-head in the British type of analysis of intonational tone-units. Consensus between the two annotators was reached on the need to acknowledge this type of discourse structure, especially as some speakers were reading from prepared material. The time span of the pause following this kind of preamble was not systematic, ranging in one speaker from 0.42sec to more than one second.

Relaunch signals Periodically, in the middle of the paratone, one speaker seems to attempt to get the audience’s attention by specifically using the pitch reset signal, which does not correspond to a new topic (and therefore a new paratone). In those relaunch cases, the acoustic cues of the pitch reset do not match a new semantic component.

5.2. The Role of Declination

To try to capture the importance of absolute value of pitch (claims 1 to 4), we plotted the pitch targets on paratones, only to realise that the apparently random distribution of extreme targets (Top, Bottom) across the paratone (see Figure 1) could have another explanation. An alternative analysis consisted in taking into account declination, right from its physiological motivation, i.e. the slow decrease of air pressure observable in each breathing group. We therefore manually annotated inspirational breaks on a supplementary tier for talk #2. Due to the absence of equipment (e.g. for thoracic and abdominal signals), the analysis of respiratory parameters is necessarily insufficient (cf relative lung volume as analysed in [37]). However, we can report inhalation duration, as it can be measured on the spectrogram for some of the recordings. Figure 7 plots the probability density functions of the different INTSINT pitch targets along the normalized duration of all the breathing groups. The y-axis measures the probability density, while the x-axis continuously measures how far into the breathing group the patterns can be found. Declination seems to take its toll as most of the B (bottom) pitch targets occur at the end of the breathing groups, whereas T (Top) are concentrated at the beginning.

Declination seems to explain why detecting paratones on the basis of absolute of pitch would fail. The quality of the sound files prevents us from verifying this with all the recordings, but it may well be the case that pitch levels are partly dependent on breathing. Coordination and planning of the paratone structure solely with pitch levels may not be maintained across breathing groups, whereas sequences of INTSINT pitch targets and speech acceleration might be.

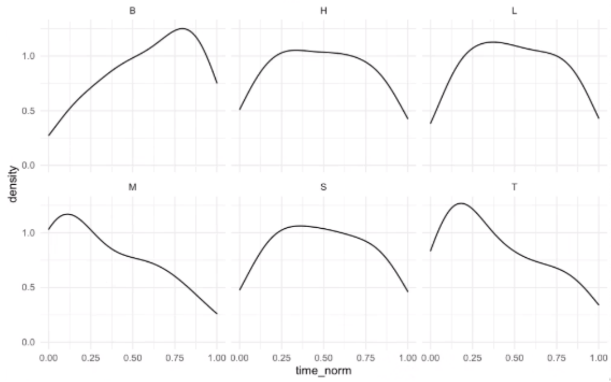

Figure 7: *Density of 6 INTSINT pitch target types over breathing groups (normalised duration) for talk #2*

5.3. The Role of Phonostyles

The inter-paratone duration may vary according to the variability of speech rates across speakers. The two talks that we manually analysed had similar speech rates (4.26 and 4.24) but these constitute the fastest speakers of the dataset. It may well be the case that the inter-paratone break is longer for slower speakers.

6. Conclusion

In this paper, we revisited six characterising features of paratones posited in the pre-corpus era. The predictions based on absolute levels of pitch do not seem to hold, but relevant combinations of INTSINT pitch targets seem to correlate rather well with paratone boundaries. If rhythmic cues can be taken into account for automatic prediction of paratone boundaries, acceleration at the beginning of the paratone seems to be more robust than deceleration at the end. In both cases, signals seem to be clearer for initial boundaries.

The current study is a good illustration of how work within Digital Humanities can exploit technological advances to test assumptions from earlier decades. Our observations provide clear evidence to qualify Tench’s six characteristics of paratones, and as such, can be used as pointers for future research on automatic detection of paratone boundaries. Two approaches can be pursued. One could try to build the algorithm hierarchically, taking into account the most robust prosodic cues for paratones. The alternative approach would consist in using co-clustering techniques (see [32]) simultaneously dealing with relevant cues such as rhythmic cues (both as spacing of syllable peaks and phone intervals) and INTSINT pitch targets and F0 associated values.

7. Acknowledgements

We thank the reviewers for comments on an earlier version of this paper. The prosodic annotation of the corpus was funded by a grant to Alice Henderson’s lab from CORLI, the French Linguistic Consortium for Corpus, Language and Interaction, via their 2019 call for projects dedicated to corpus delivery (*Appel à Finalisation de corpus*).

8. References

[1] P. Thompson and H. Nesi, "Research in progress, the British Academic Spoken English (BASE) corpus project," *Language Teaching Research*, vol. 5, no. 3, pp. 263–264, 2001.

[2] BASE, "British Academic Spoken English Corpus," 2008. [Online]. Available: http://www.helsinki.fi/varieng/CoRD/corpora/BASE/

[3] JSCC, "John Swales Conference Corpus," 2006. [Online]. Available: https://varieng.helsinki.fi/CoRD/corpora/JSCC/

[4] MICASE, "Michigan Corpus of Academic Spoken English," 2007. [Online]. Available: https://lsa.umich.edu/eli/language-resources/micase-micusp.html

[5] B. Paltridge, *Genre and English for Specific Purposes*. Wiley Blackwell, 2012.

[6] W. Cheng, C. Greaves, M. Warren *et al.*, "The creation of a prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic)," *ICAME journal*, vol. 29, pp. 47–68, 2005.

[7] L. Pickering and P. Byrd, *Investigating connections between spoken and written academic English: Lexical bundles in the AWL and in MICASE*. University of Michigan Press, 2008, ch. 6, pp. 110–132.

[8] P. Tench, *The Intonation Systems of English*. Bloomsbury Publishing, 1996.

[9] G. Brown and G. Yule, *Discourse analysis*. Cambridge University Press, 1983.

[10] G. Brown, *Listening to spoken English*. Longman, 1977.

[11] M. Nespor and I. Vogel, *Prosodic phonology*. De Gruyter Mouton, 2012.

[12] H. Truckenbrodt, "The syntax-phonology interface," *The Cambridge Handbook of Phonology*, pp. 435–456, 2007.

[13] D. Büring, *Intonation and meaning*. Oxford University Press, 2016.

[14] G. Yule, "Speakers' topics and major paratones," *Lingua*, vol. 52, pp. 33–47, 1980.

[15] S.-p. Wang, "Multimodal research on tonal variations for pragmatic purposes in Mandarin," *Journal of Pragmatics*, vol. 113, pp. 53–70, 2017.

[16] K.-I. Kadooka, "Paratone as a discourse marker," *Estudos Linguísticos/Linguistic Studies*, vol. 2, pp. 27–40, 2008.

[17] G. O'Grady, *A Grammar of Spoken English Discourse: The intonation of increments*. Bloomsbury Publishing, 2010.

[18] A. Wennerstrom, "Intonational meaning in english discourse: A study of non-native speakers," *Applied linguistics*, vol. 15, no. 4, pp. 399–420, 1994.

[19] S. Schuetze-Coburn, M. Shapley, and E. G. Weber, "Units of intonation in discourse: A comparison of acoustic and auditory analyses," *Language and Speech*, vol. 34, no. 3, pp. 207–234, 1991.

[20] L. Pickering, "The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse," *English for Specific Purposes*, vol. 23, no. 1, pp. 19–43, 2004.

[21] D. Brazil, *The communicative value of intonation in English*. Cambridge University Press, 1997.

[22] L. Pickering, "The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse," *English for Specific Purposes*, vol. 23, no. 1, pp. 19–43, 2004.

[23] J. Choi, D. Hindle, F. Pereira, A. Singhal, and S. Whittaker, "Spoken content-based audio navigation (scan)," in *Proceedings of the ICPhS*, vol. 99. Citeseer, 1999.

[24] E. Douglas-Cowie and R. Cowie, "Macrostructures in prosody: the case of phonecalls," in *Intonation: Theory, Models and Applications. Proceedings of an ESCA workshop*. Athens, Greece, 1997, pp. 103–106.

[25] O. Kang, D. Rubin, and L. Pickering, "Suprasegmental measures of accentedness and judgments of language learner proficiency in oral english," *The Modern Language Journal*, vol. 94, no. 4, pp. 554–566, 2010.

[26] C. De Looze and S. Rauzy, "Automatic detection and prediction of topic changes through automatic detection of register variations and pause duration," in *Interspeech 2009*, 2009, p. 2919–2922.

[27] C. L. Smith, "Topic transitions and durational prosody in reading aloud: Production and modeling," *Speech Communication*, vol. 42, no. 3-4, pp. 247–270, 2004.

[28] D. Hirst, "Phonetic and phonological annotation of speech prosody," in *(a cura di) Analisi prosodica. teorie, modelli e sistemi di annotazione, Atti del II Convegno Nazionale della Società Italiana di Scienze della Voce (AISV)*, R. Savy and C. Crocco, Eds. Torriana: EDK, 24-29: Università degli Studi di Salerno 30 novembre-2 dicembre 2005, 2006.

[29] B. Bigi, "SPPAS - Multi-lingual approaches to the automatic annotation of speech," *The Phonetician*, vol. 111-112, pp. 54–69, 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01417876

[30] N. H. De Jong and T. Wempe, "Praat script to detect syllable nuclei and measure speech rate automatically," *Behavior research methods*, vol. 41, no. 2, pp. 385–390, 2009.

[31] D. J. Hirst, "A praat plugin for momel and intsint with improved algorithms for modelling and coding intonation," in *Proceedings of the XVIth International Conference of Phonetic Sciences*, vol. 12331236, 2007, pp. 1223–1236.

[32] A. Méli, N. Ballier, A. Falaise, and A. Henderson, "An experiment in paratone detection in a prosodically annotated EAP spoken corpus," *Proc. Interspeech 2021*, pp. 2616–2620, 2021.

[33] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. [Online]. Available: https://ggplot2.tidyverse.org

[34] K. Peshkov, L. Prévot, and R. Bertrand, "Prosodic phrasing evaluation: Measures and tools," *Proceedings of TRASP*, 2013.

[35] A. Wichmann, *Intonation in text and discourse: Beginnings, middles and ends*. Taylor and Francis, 2000.

[36] S. Schuetze-Coburn, M. Shapley, and E. G. Weber, "Units of intonation in discourse: A comparison of acoustic and auditory analyses," *Language and Speech*, vol. 34, pp. 207 – 234, 1991.

[37] S. Fuchs, C. Petrone, J. Krivokapić, and P. Hoole, "Acoustic and respiratory evidence for utterance planning in German," *Journal of Phonetics*, vol. 41, no. 1, pp. 29–47, 2013.