

# Investigating Prosodic Variation in British English Varieties using *ProPer*

Hae-Sung Jeon<sup>1</sup>, Stephen Nichols<sup>2,3</sup>

<sup>1</sup>School of Humanities, Language and Global Studies, University of Central Lancashire, UK

<sup>2</sup>Faculty of Linguistics, Philology and Phonetics, University of Oxford, UK

<sup>3</sup>Department of Language and Linguistics, University of Essex, UK

HJeon1@uclan.ac.uk, sn21014@essex.ac.uk

## Abstract

This study used *ProPer* (Albert et al., 2022, *ProPer: PROsodic analysis with PERiodic energy* [computer software], <https://doi.org/10.17605/OSF.IO/28EA5>), a tool for automatic prosodic analysis, to investigate prosodic variation in four varieties of British English. The central question was how to efficiently describe systematic prosodic variation without extensive manual annotations. We analysed *ProPer* parameters, the magnitude of F0 change between two successive syllables, synchrony (i.e. the general trend of the F0 movement within a syllable), and periodic energy mass. We then carried out audiovisual analysis of the F0 movement. Our overall assessment is that *ProPer* allows holistic prosodic analysis, significantly reducing researchers' workload in prosodic annotations. The *ProPer* analysis is optimised for syllable-sized intervals. It requires the examination of a vast amount of acoustic information and possibly the careful design of experimental materials for collecting laboratory speech data.

**Index Terms:** intonation, prosody, speech production, English

## 1. Introduction

The quantitative investigation of prosodic variation poses non-trivial challenges to researchers. Conventionally, researchers undergo intensive training in phonological and phonetic prosodic annotation, but various analytical methods have their limitations. For instance, the ToBI [see 1] is dialect-specific and therefore unsuitable for cross-varietal investigation [cf. 2]. Some phonetic analysis algorithms, such as the INTSINT [e.g. 3], specifically model F0 contours, and researchers need to collect information about other acoustic parameters such as timing and intensity separately. The speech data annotation is extremely labour-intensive, and further challenges remain in quantitative analysis of the annotated data.

We used a new, freely available tool, *ProPer* (*PROsodic analysis with PERiodic energy*) [4], which makes a holistic analysis possible. *ProPer* produces visual representations of acoustic parameters including the F0 contour over an utterance and its strength [5] (e.g. Figure 1). It calibrates a range of acoustic parameters which can be submitted to statistical analysis. *ProPer* allows analyses of any language with stable vocalic intervals (i.e. periodic energy). The present study examined how the *ProPer* analysis and minimal manual analyses could be used in describing prosodic differences across four varieties of British English. Our intention was to test the new tool on existing data to explore how it can be used for novel data.

## 2. Target Varieties and Speech Data

This study used speech data from Cambridge, Belfast, Liverpool and Newcastle upon Tyne varieties of English in the IViE corpus [6]. These were chosen to include the standard variety (Cambridge, Standard Southern British English, SSBE) and three varieties which have prosodic properties that are distinct from SSBE. Unlike SSBE, which typically marks the end of a declarative with a pitch fall, Liverpool, Belfast, and Newcastle English show a final rise in pitch [7, p. 133]. In particular, Belfast English seems to use rising tones more frequently compared to other varieties [8].

In yes-no questions (which were analysed in the present study), [8] reported that Cambridge speakers used both final falls (L%, 44%) and rises (H%, 56%), Belfast always used a final rise, and Newcastle speakers showed a similar trend to Cambridge speakers. For the rise, both Cambridge and Newcastle speakers produced H\* LH% at similar frequencies. The difference was that Cambridge speakers often produced L\* HH% (with a steep rise from the stressed syllable) whereas Newcastle speakers produced L\*H%. For yes-no questions, Liverpool speakers dominantly used a final rise [9].

## 3. Methodology

### 3.1. Data

Three sentences, 'are you growing limes or lemons?', 'is his name Miller or Mailer?', and 'did you say mellow or yellow?' were chosen from the Cambridge, Belfast, Liverpool and Newcastle varieties respectively [6]. In total, 144 utterances were selected (3 sentences × 12 speakers × 4 areas).

### 3.2. *ProPer* analysis

Although *ProPer* does not require manual annotations, we annotated syllable boundaries in each utterance using Praat [10]. The marking of syllable boundaries allowed us to compare the same intervals across the same sentences spoken by different speakers. For words with an ambisyllabic consonant (e.g. mellow), the boundary was placed between the preceding vowel and the ambisyllabic consonant. The syllable-sized interval served as an analysis window for each acoustic parameter. Then *ProPer* produced periograms (e.g. Figure 1). *ProPer* produced output files listing a wide range of acoustic parameters, such as F0, ΔF0, interval duration, Tonal Centre of Gravity [11], synchrony [12], periodic energy mass, etc. We report the results of the following (see [5] for details):

- ΔF0: The difference in F0 between the target syllable and its preceding syllable.

- Synchrony: The distance between the Tonal Centre of Gravity and the Centre of Mass of the Periodic Energy Curve (TCoG - CoM). Synchrony represents the overall F0 trend within the target syllable ( $> 0$ : rise,  $= 0$ : level, or symmetrical fall-rise or rise-fall,  $< 0$ : fall).
- Periodic energy mass: The integral of duration and power, i.e. prosodic prominence or strength.

We used relative measures normalised for each token (i.e. F0 change relative to, for instance, the pitch range of a given utterance).<sup>1</sup>

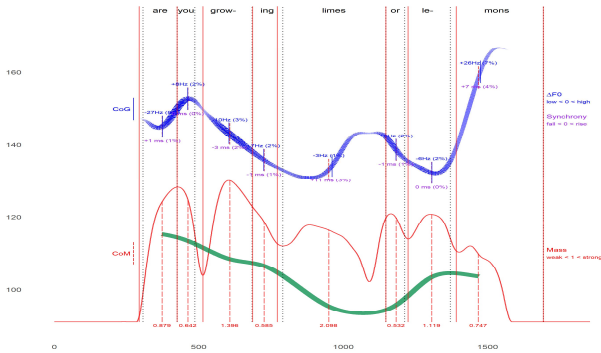


Figure 1: Periogram sample (Belfast). X-axis represents time (ms) and y-axis frequency (Hz). The blue line represents the F0 contour with its thickness showing the prosodic strength. The green line shows speaking rates. The red curve shows periodic energy. Dotted vertical lines show the syllable boundaries in Praat TextGrid. Red vertical lines show the interval boundaries determined by ProPer.

### 3.3. Audiovisual analysis

We examined two words in each utterance bearing either the prenuclear (e.g. ‘mellow’ in ‘did you say mellow or yellow?’) or nuclear accent (e.g. ‘yellow’). As *ProPer* operated in short domains (e.g. within a syllable or between two syllables), we intended to examine the broad F0 contour shape associated with an accent. We listened to each utterance and visually examined the F0 contour, referring to the periogram. We identified whether the F0 associated with each accent formed a rise-fall with a peak (i.e. the maximum F0) or a fall-rise with a valley (i.e. the minimum F0). The time point of the F0 maximum or the minimum was then semi-automatically detected using Praat, and annotated in a point tier. We marked the onset and the offset of the F0 plateau (the flat stretch of F0) associated with each accent. For defining a plateau, we used the 4% criterion following [13]. For instance, if the F0 contour was valley-shaped, falling to the minimum at 120 Hz ( $4\% = 4.8$  Hz), then the plateau onset and offset were annotated at the time points where the F0 was 115 Hz in its contour before and after the minimum. When the plateau could not be identified in a continuously rising or falling contour, the accent was excluded from the analysis. The author SN carried out the initial analysis and the annotations were cross-checked by the author HSJ.

<sup>1</sup> We chose the ‘token’ normalisation method for F0-related measures to account for the acoustic variation related to both speakers and utterances, as one speaker may vary their phonetic properties differently across utterances. Alternatively, normalisation can be done relative to the whole ‘data’ or for ‘speaker’. Normalisation relative to ‘speaker’ is

## 4. Analysis and Results

### 4.1. ProPer outputs

We examined the Area effect for the stressed syllables and following unstressed syllables in the words carrying either the prenuclear or nuclear accent (e.g. ‘mellow’ and ‘yellow’ in ‘did you say mellow or yellow?’). R [14] was used for all statistical analyses. Initially, we constructed a mixed model with a fixed factor, Area, and the random factors, syllable label and speaker, with a random intercept [15]. The initial model was compared to a lower-order model constructed with the two random factors (*anova* function,  $df = 3$ , Table 1).

Table 1: The Area effect in model comparisons ( $\chi^2$ ,  $p < .05^*$ ,  $< .001^{**}$ ,  $< .0001^{***}$ )

	Stressed	Unstressed
<b>Prenuclear</b>		
$\Delta F0$	9.07*	22.79***
Synchrony	2.71	17.03***
Mass	11.76**	1.34
<b>Nuclear</b>		
$\Delta F0$	4.11	15.44**
Synchrony	9.61	18.18***
Mass	4.11	8.19*

#### 4.1.1. $\Delta F0$

For prenuclear position, Area had an effect on  $\Delta F0$  for both the stressed and unstressed syllables, e.g. the F0 movement between both *say* and *me-*, and between *me-* and *-llow* in ‘did you say mellow or yellow?’. Figure 2 shows the general trends that, over *say mellow*, F0 rose gradually or fell for Cambridge, it rose for Belfast, it rose gradually for Liverpool, and it was relatively flat for Newcastle. To elaborate, for Cambridge, the F0 differences clustered at zero between *say* and *me-* (the ‘stressed’ condition), while F0 either rose or fell at similar frequencies between *me-* and *-llow*. The wide distribution of  $\Delta F0$  for the unstressed syllable was notable. For Belfast, F0 rose slightly or fell between *say* and *me-* with a rise between *me-* and *-llow*. Liverpool showed a similar trend to Belfast. For Newcastle,  $\Delta F0$  values clustered near zero.

For nuclear position, Figure 3 shows that F0 changed little for the ‘stressed’ condition, i.e. between the unstressed syllable and the following stressed syllable for all Areas (e.g. *or* and *ye-* in ‘did you say mellow or yellow?’). Area had an effect only for the unstressed syllable (e.g. between *ye-* and *-llow*). Figure 3 shows that the trend for Cambridge was a nuclear final fall, Belfast showed a slight rise, Liverpool showed more rises than falls, and Newcastle showed a slight fall.

preferred when the data include multiple utterances spoken by the same speakers (A. Albert, *pc*).

#### 4.1.2. Synchrony

The Area effect was significant only for the unstressed syllables (Table 1) for both prenuclear and nuclear positions. Figures 4 and 5 show that the synchrony showed a similar trend between the prenuclear and nuclear positions for all areas. For Cambridge, F0 was more likely to fall in both positions (e.g. *me-llow* and *ye-llow* in ‘did you say mellow or yellow?’). For Belfast, F0 tended to rise. For Liverpool, synchrony clustered near zero for the prenuclear position, while F0 was more likely to rise in the nuclear position. For Newcastle, F0 was more likely to fall. Figures 4 and 5 show that synchrony showed a wider distribution for the prenuclear than the nuclear position.

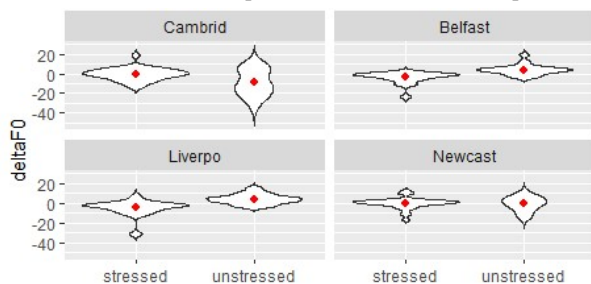


Figure 2:  $\Delta F0$  for prenuclear position.

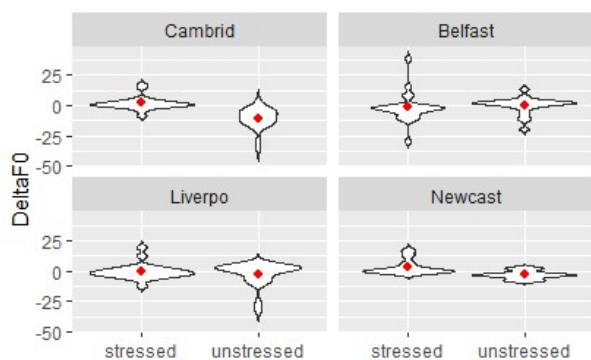


Figure 3:  $\Delta F0$  for nuclear position.

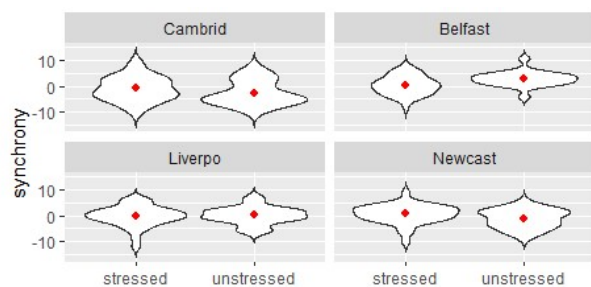


Figure 4: Synchrony for prenuclear position.

#### 4.1.3. Mass

The Area effect was significant for the stressed syllable in prenuclear position (e.g. *me-* in ‘did you say mellow or yellow?’) and for the unstressed syllable following the nuclear syllable (e.g. *ye-llow*) (Table 1). In the mixed model (see Section 4.1), for the prenuclear stressed syllable, compared to Cambridge (intercept = 1.82, SE = 0.26), the relative mass was lower for Belfast (est. = -0.28, SE = 0.97), higher for Liverpool (est. = 0.03, SE = 0.1) and lower Newcastle (est. = -0.16, SE =

0.97). That is, the prenuclear stressed syllables were relatively acoustically strong within the utterance for Cambridge and Liverpool.

For the unstressed syllable following the nuclear syllable, compared to Cambridge (intercept = 0.85, SE = 0.15), Mass was lower for Belfast (est. = -0.1, SE = 0.12), and higher for Liverpool (est. = 0.1, SE = 0.12) and Newcastle (est. = 0.24, SE = 0.12). That is, the unstressed syllable was acoustically strong for Cambridge, Liverpool and Newcastle compared to Belfast. The reason for this is not clear, but it is possible that the steep F0 rise for Belfast occurred with a rapid decrease in periodic energy.

Overall, compared to Cambridge, the relative Mass was consistently lower for Belfast and higher for Liverpool. The trend for Newcastle was not consistent. Examination of the Mass distributions (not reported here due to the space limitation) revealed that, overall, Mass was reduced in the unstressed syllables compared to the stressed syllables for all Areas, but less so for Newcastle.

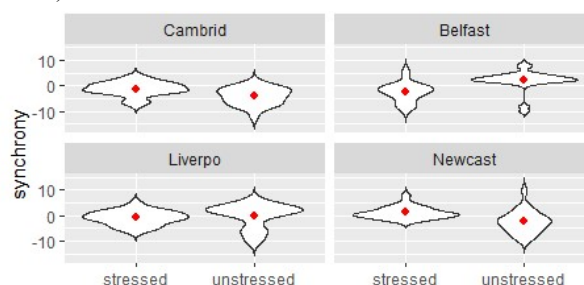


Figure 5: Synchrony for nuclear position.

#### 4.2. Manual analysis

For Cambridge (CB), the majority of the prenuclear accents were peak-shaped, whereas they were valley-shaped for Belfast (BF) and Liverpool (LP). Newcastle (NT) showed the similar frequency between the peak- and valley-shaped accents (Table 2). For the nuclear position, Cambridge, Belfast and Liverpool showed the same pattern as in their respective prenuclear position; mainly peak-shaped for Cambridge, and valley-shaped for Belfast and Liverpool. For Newcastle, in nuclear position, peaks were more common than valleys.

Table 2: The F0 extremum (peak with max. vs valley with min.) frequency and mean plateau duration (ms).

	N	Mean	SE	N	Mean	SE
Prenuclear, max.			Prenuclear, min.			
CB	20	93.97	11.23	4	253.03	48.51
BF				34	166.15	12.43
LP	4	107.45	32.55	27	154.52	15.16
NT	16	162.18	16.25	13	161.34	17.39
Nuclear, max.			Nuclear, min.			
CB	23	109.99	8.32			
BF	3	112.84	29.91	29	139.82	10.77
LP	6	158.59	27.32	26	155.87	12.32
NT	18	141.94	8.99	9	119.59	17.78

We constructed mixed-effect models (as in Section 4.1) to examine the effect of Area, Extremum Type (peak with max. vs valley with min.) and their interaction on plateau duration. For prenuclear position, there was a significant effect of Extremum ( $df = 1, \chi^2 = 6.32, p < .05$ ) and the Area  $\times$  Extremum interaction ( $df = 2, \chi^2 = 12.24, p < .01$ ). Figure 6 shows that for Cambridge, while F0 tended to form a clear peak (i.e. a short plateau for max.) in prenuclear position, the F0 formed a long plateau when it was valley-shaped with a final rise (but there were only a few tokens for the latter, Table 2). For Belfast which predominantly had accent valleys, the F0 plateau tended to be longer compared to Cambridge. For Liverpool, the plateaux associated with the frequent accent valleys (i.e. minimum) were longer compared to peaks. For Newcastle, plateau duration did not seem markedly differ between accent peaks and valleys.

For nuclear position, only the Area effect was significant ( $df = 3, \chi^2 = 1.53, p < .05$ ). Figure 7 shows that Belfast and Liverpool with frequent accent valleys had a wider distribution of the plateau durations compared to Cambridge and Newcastle.

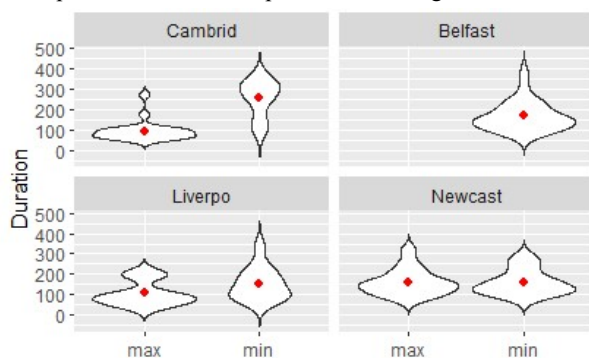


Figure 6: Plateau duration (ms), prenuclear position.

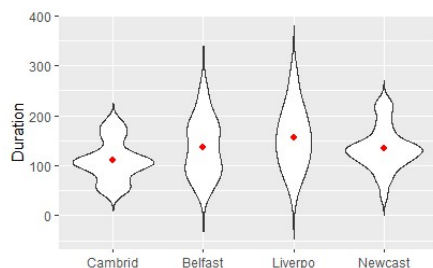


Figure 7: Plateau duration (ms) for nuclear position, data collapsed across max. and min.

## 5. Discussion

Overall, the results of the *ProPer* F0 parameter ( $\Delta F0$  and synchrony) analysis are in line with the previous findings. The intonational events at the post-stressed syllable in the nuclear position were crucial for characterising the varieties as discussed in literature [cf. 7]. In addition to pitch events, *ProPer* allowed us to examine detailed distributional properties of various acoustic dimensions. The *ProPer* analysis showed that, for Cambridge, F0 either rose or fell between the prenuclear stressed syllable and following unstressed syllable (e.g. ‘mellow’ in ‘did you say mellow or yellow?’), and in nuclear position (e.g. ‘yellow’), it tended to fall. For Belfast, F0 rose for both prenuclear and nuclear positions, forming a F0 valley accent. For Liverpool, F0 tended to rise in both positions as for Belfast, but its movement seemed more gradual compared to

that for Belfast. For Newcastle, F0 was relatively flat between the prenuclear stressed syllable and the unstressed syllable, and it slightly fell in prenuclear position. Newcastle showed less variation in F0 compared to other varieties. These results mirror the reported distribution of accent types in yes-no questions across the varieties (Section 2).

Moreover, compared to Cambridge, periodic energy mass was consistently lower for Belfast but higher for Liverpool. The trend for Newcastle was not consistent. For Newcastle, the contrast in the energy mass between stressed and unstressed syllables was relatively small compared to other varieties. The mass values can be influenced by the presence of creak and final lengthening; therefore, their interpretations will require some caution. Once these potential confounds are treated in the analysis process, the energy mass can be useful for investigating how different varieties contrast prominent and less prominent syllables (i.e. ‘prominence gradient’, [16]).

*ProPer* offered a significant advantage of timesaving, and the periograms were extremely useful for examining the F0 contours. The main challenge was interpreting the acoustic outputs measured in syllable-sized intervals to understand the prosodic events occurring over larger domains. For instance, in order to picture the F0 trajectory over a short word, the two parameters,  $\Delta F0$  and synchrony, were to be considered simultaneously. However, once researchers make informed choices in selecting the acoustic parameters to examine and conduct appropriate pre-processing of data, valid interpretations are possible. For instance, although this study did not examine different types of rising tones, e.g. ‘rise’, ‘rise-plateau’, ‘rise-plateau-slump’, and ‘rise-fall’ [7], such differences can be accounted for by examining TCoG values. However, more complex F0 movements, such as multitonal boundary tones (e.g. HLH%) associated with a syllable can pose a problem. The complex movement is represented as a single value of  $\Delta F0$  or synchrony; therefore, such tones may need to be excluded in the analysis.

The audiovisual classification of the F0 extremum type (i.e. accent peaks vs valleys) seemed to be redundant, as this could be coded in the data based on synchrony. Furthermore, reporting the frequency of peaks and valleys across varieties is not satisfactory as their phonetic properties differed across varieties, e.g. a peak for Newcastle was not as sharp as that for Cambridge. However, given that *ProPer* outputs were based on syllabic intervals, temporal properties of larger domains still required separate analysis.

## 6. Conclusions

For using *ProPer*, our recommendation is to carefully design speech materials to be suitable for analysis based on syllable-sized intervals. For investigating variation related to a pre-determined factor (‘English variety’ in the present study), the workflow would be most straightforward for local level analysis, e.g. focusing on prosodic events on a few target syllables. For quantitative analysis of laboratory speech, for instance, the scope can be limited to multisyllabic words rather than a whole utterance, with matching number of syllables across experimental conditions in the experimental materials.

## 7. Acknowledgements

We would like to thank Aviad Albert and Francesco Cangemi for their help in setting up the *ProPer* analysis and for their comments on the earlier version of this paper.

## 8. References

- [1] M. E. Beckman, J. Hirschberg, and S. Shattuck-Hufnagel, “The Original ToBI System and the Evolution of the ToBI Framework,” in *Prosodic Typology*, S.-A. Jun, Ed. Oxford University Press, 2005, pp. 9–54, 2005.
- [2] E. Grabe, F. Nolan, and K. Farrar, “IViE – A comparative transcription system for intonational variation in English,” *Proceedings of ICSLP 98*, Sydney, Australia, 1998.
- [3] D. Hirst, “A Praat plugin for Momel and INTSINT with improved algorithms for modelling and coding intonation,” in the *Proceedings of the ICPHS XVI*, Saabrücken, Germany, 6-10 August, 2007, pp. 1233–1236.
- [4] A. Albert, F. Cangemi, T. M. Ellison, and M. Grice, *ProPer: PROsodic analysis with PERiodic energy*, <https://doi.org/10.17605/OSF.IO/28EA5>, 2022.
- [5] A. Albert, F. Cangemi, and M. Grice, “Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration,” in *Speech Prosody 2018*, pp. 804–808, Poznań, Poland, Jun. 2018.
- [6] E. Grabe, B. Post, and F. Nolan, *The IViE Corpus*, Department of Linguistics, University of Cambridge, 2001, available at: <http://www.phon.ox.ac.uk/IViE>.
- [7] A. Cruttenden, *Intonation*, 2nd ed. Cambridge, UK: Cambridge University Press, 1997.
- [8] E. Grabe, “Intonational variation in urban dialects of English spoken in the British Isles,” in *Regional Variation in Intonation*, P. Gilles and J. Peters, Eds. Tübingen: Niemeyer, 2004, pp. 9–31.
- [9] C. Nance, S. Kirkham, K. Lightfoot, and L. Carroll, “Intonational variation in the North-west of England: The origins of a rising contour in Liverpool,” *Language and Speech*, 2020, 002383092096973
- [10] P. Boersma and D. Weenink, *Praat: Doing Phonetics by Computer*, ver. 6.1.16 [Computer program],” 2020.
- [11] J. Barnes, N. Veilleux, A. Brugos, and S. Shattuck-Hufnagel, “Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology,” *Laboratory Phonology*, vol. 3, article no. 2, Jan. 2012.
- [12] F. Cangemi, A. Albert, and M. Grice, “Modelling intonation: Beyond segments and tonal targets,” In the *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Australia, pp. 572-576, 2019.
- [13] R.-A. Knight, “The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux,” *Language and Speech*, vol. 51, 223–244, 2008.
- [14] R Core Team, *R: A Language and Environment for Statistical Computing, version 4.1.1*, 2021.
- [15] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol. 67, 1–48, 2015.
- [16] F. Nolan and E. L. Asu, “The Pairwise Variability Index and coexisting rhythms in language,” *Phonetica*, vol. 66, 64–77, 2009.