



# An analysis of prosodic boundary detection in German and Austrian German read speech

Barbara Schuppler<sup>1</sup>, Bogdan Ludusan<sup>2,3</sup>

<sup>1</sup> Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

<sup>2</sup> Phonetics Workgroup, Faculty of Linguistics and Literary Studies, Bielefeld University, Germany

<sup>3</sup> CITEC, Bielefeld University, Germany

b.schuppler@tugraz.at, bogdan.ludusan@uni-bielefeld.de

## Abstract

With speech annotation being one of the most time-consuming and costly aspects of speech corpora development, there is a significant interest in the development of automatic annotation tools. The present study focuses on variant-independent prosodic boundary annotations for German. We test a previously proposed unsupervised approach, which posits prosodic boundaries based only on acoustic cues. The experiments were conducted on read speech from two corpora, one of Standard German, the Kiel Corpus of Spoken German, and the other of Austrian German, the Graz Corpus of Read and Spontaneous Speech. Averaging across all speakers in the dataset, the tool attained an area under the precision-recall curve of 0.308 and 0.215, for the Kiel corpus and the GRASS corpus, respectively. The significant differences obtained in detection across the two varieties were accompanied by large differences between speakers, as well. This was confirmed by a subsequent analysis of the acoustic cues employed in the process, which showed important differences in the way speakers make use of those cues for marking prosodic structure. We discuss these findings with respect to the current literature and their implication for variant-independent automatic annotation.

**Index Terms:** prosodic phrase boundaries, automatic detection, Austrian German, acoustic features

## 1. Introduction

Investigations in speech science and technology require the development of corpora for the languages studied, as well as for their varieties. One of the most time-consuming and costly aspect of developing speech resources is their annotation. For the creation of phonetic segmentations, automatic tools have been based on *forced alignment* (e.g., [1, 2]). The advantage of such tools is that errors that occur are mostly systematic and can therefore be taken into account in the analysis. Given the higher complexity of prosodic phenomena, their manual annotation requires even more time and yields lower inter-labeler agreements than segmental transcriptions. This paper contributes to the body of work on prosodic annotation tools with a special focus on developing variant-independent prosodic boundary annotations for German. The aims of this paper are two-fold: The first aim is to develop a tool that facilitates the currently ongoing prosodic transcription process of GRASS, the first large scale database for Austrian German, containing both read and conversational speech [3]. So far, only a small portion of the recorded speech materials have been manually annotated for prosodic boundaries. With the help of the tool presented here, we intend to annotate the rest of the corpus by means of a semi-automatic approach (similar to e.g., [4] for French). The second

aim is to analyse how strongly the acoustic features used by our system contribute to the perception of boundaries by German vs. Austrian listeners (as given by the perception-based annotations). We hope that this analysis will allow us to improve the proposed system, while shedding further light on how prosodic boundaries are marked across German varieties.

### 1.1. Prosodic boundary detection tools

Several prosodic annotation systems have been previously proposed. Some of them combine acoustic, lexical and syntactic features (e.g., [5] for American English, [6] for French), others use lexical and syntactic information alone (e.g., [7] for Dutch). Since our aim is to build a tool that can be incorporated into the annotation process of a not-yet annotated database and also into an automatic speech recognition (ASR) system, the requirement for the system is to exclusively use acoustic information.

For German, only a few prosodic annotation tools have been built that use acoustic features alone. For instance, N. Braunschweiler [8] proposed ProsAlign, a system that automatically produces GToBI labels, does not require a pre-segmentation of the speech stream and should, in principle, be language independent. The proposed approach discovers 56% of the manually established labels and could, thus, be integrated in a semi-automatic annotation procedure. Since the development of ProsAlign, other prosodic annotation systems, not based on GToBI, have been developed for German (e.g., KIM, DIMA) showing certain advantages compared to existing approaches [9]. For Austrian German, no tool is available at this point.

This paper investigates a tool based on a system previously proposed by Ludusan and Dupoux [10]. It posits prosodic boundaries based on four acoustic cues: duration of the following pause, duration of the syllable nucleus, the nucleus-onset-to-nucleus-onset duration, and f0 reset. These cues have been shown to be employed for boundary identification in a wide variety of languages, being exploited even by infants in the process of early language acquisition.

### 1.2. Perception of prosodic boundaries

Among the acoustic cues have been found to be relevant for the perception of prosodic boundaries, the existence of a pause in the speech signal is considered to be one of the most salient. Besides numerous linguistic studies that investigated the relationship between pause and boundary perception (e.g., [11, 12]), the electrophysiological study by Männel and Friederici [13] revealed that 3-year-olds perceive the boundary also when the existence of a pause is the only acoustic cue marking the boundary. This is in line with the findings of Petrone et al. [14], showing that adult German listeners give categorical responses

for prosodic boundaries in the case of pauses, while more gradual transitions were observed with f0 and final lengthening cues. Their experimental design, however, did not allow to show the weights of these three cues. Final lengthening is another well studied cue to prosodic boundaries (e.g., [15, 16, 17]). In addition to longer segment durations before a phrase boundary, domain initial speech sounds may be produced with a stronger contact between articulators and with a decreased coarticulation to the previous context [18].

For Austrian German, only a few studies have focused on prosody (e.g., [19, 20]), and even less is known about the production and perception of prosodic boundaries. Ulbrich [21] compared prosodic phrasing in the three German standard varieties (Germany, Austria, Switzerland). She based her analysis on the following measures: number of intra-sentential boundaries, number and duration of intra-sentential pauses, f0 reset, and phrase initial and phrase final syllable duration. She found that, whereas f0 reset and pausing was used equally among the different varieties, German speakers followed more closely the punctuation marks (syntax-prosody relationship) and had the lowest change in speech rate across boundaries. Recently, Schuppler and Zellers [22] found that plosives at prosodic boundaries are produced differently in German and Austrian German read speech. While Austrian speakers tend to lengthen the duration of the closure of plosives, German speakers lengthen the duration of the burst.

## 2. Materials and Methods

### 2.1. Read speech material

We tested the system on read speech from two corpora, the Kiel Corpus of Spoken German [23], which contains speech from, mainly, Northern Germany, and the Graz Corpus of Read and Spontaneous Speech (GRASS; [3, 24]), which contains speech from eastern-Austrian speakers. As these corpora were annotated with different methods (manual phonetic segmentations for the Kiel corpus vs. semi-automatic segmentations for GRASS) and wanting to have comparable input data for a more accurate comparison, we created automatic segmentations for both corpora, using MAUS [1]. These automatic annotations were subsequently manually corrected, which was especially necessary for the Austrian data, as MAUS has no acoustic models for the Austrian German variant. At the prosodic level, a subset of GRASS was manually annotated for prosodic boundaries, using the same criteria as for the Kiel Corpus. Each utterance was annotated by one phonetically trained transcriber and checked by two other transcribers. This procedure was chosen to guarantee a high transcription quality. Based on a small validation set of 269 words from 47 utterances, which were randomly presented to the transcribers during the transcription process, we calculated Cohen's kappa for the decision whether a boundary should be placed after a word or not. The obtained inter-annotator agreement was good to high.

For the present study, we chose the recordings corresponding to the components *Nordwind* and *Buttergeschichte*, which exist in both corpora. For the GRASS corpus, we used a number of additional recordings containing materials which can be found also in the Kiel dataset. This resulted in 391 and 368 sentences (from 38 and 30 speakers) for the GRASS dataset and Kiel corpus, respectively.

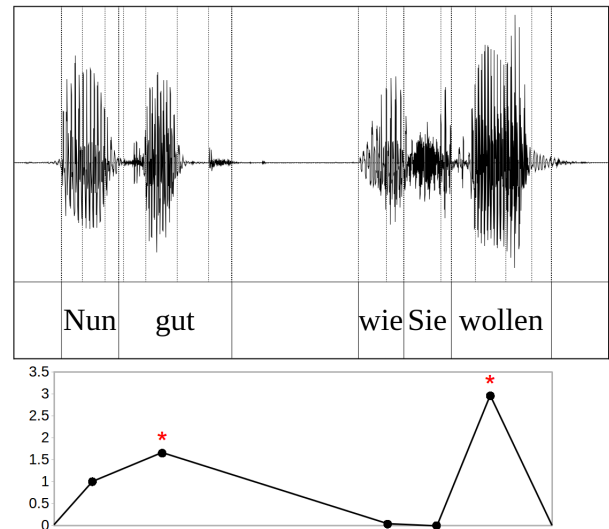


Figure 1: Waveform, transcription and the obtained prominence detector function for the utterance: "Nun gut, wie Sie wollen". The syllables corresponding to the local maxima of the function are represented by a red asterisk. Prosodic boundaries will be placed after these syllables.

### 2.2. Prosodic boundary detection

For the detection of prosodic boundaries, we employed a previously proposed system [10], which posits boundaries based on four acoustic cues:

- duration of the following pause
- duration of the syllable nucleus
- the nucleus-onset-to-nucleus-onset duration
- f0 reset

They represent the acoustic correlates of a number of phenomena associated with prosodic structure (pausing, final lengthening, initial strengthening and pitch reset). Furthermore, these cues have been shown to mark prosodic boundaries in a variety of languages, being also employed by infants in early language segmentation. The values of these cues is computed for each syllable in the utterance and then, for each cue only the local maxima are considered. Based on these values, the system computes a syllable-based detector function, by summing the values of the individual cues. Before the summation, the features were normalized between [0, 1] and were given the same weight in the calculation of the detector function. Prosodic boundaries are placed in correspondence to the local maxima of the obtained detector function. An example, based on a sentence contained in the employed corpus, is given in Figure 1.

We ran the algorithm separately on the sentences belonging to each speaker from the two corpora. This being a preliminary study, we extracted the duration features from the automatic segmentation, while f0 was extracted using Praat [25]. The autocorrelation method was used for this, with a pitch range set between 75 and 400 Hz. For the syllabic nuclei for which no f0 value was found using Praat (for example, in the case of creaky voice), we employed the SWIPE algorithm [26] to obtain f0 values for the nucleus. All utterance-internal prosodic boundaries found by the system were evaluated against the manual boundaries, by means of precision, recall and F-score (the harmonic

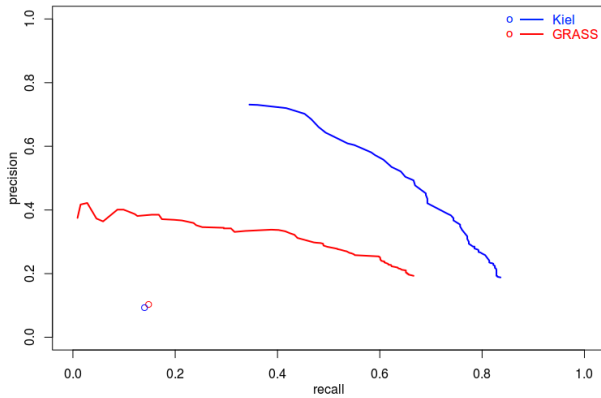


Figure 2: Precision-recall curves obtained on the two datasets with the automatic detection system (solid lines) and with a random baseline (circles).

mean of the former two measures). In order to observe the behaviour of the system on the input data, we varied the threshold used for placing boundaries, deriving a precision-recall curve. For an easier comparison between the two language varieties, we employed the area under the precision-recall curve (AUC) as evaluation metric. The proposed approach was then compared to a baseline, which posits randomly, for each utterance, a number of boundaries equal to the number of manually annotated prosodic boundaries for that utterance.

### 2.3. Statistical analysis

In order to investigate the contribution of the acoustic features to the perception of a prosodic boundary in German vs. Austrian German, we built linear mixed effects regression models with *boundary* (Y, N) as dependent variable and the previously mentioned acoustic measures *Pause* (Y, N), *Nucleus* (duration) and *f0\_Reset* as independent variables. As the nucleus-onset-to-nucleus-onset duration heavily correlates with *Pause* and *Nucleus* ( $r = 0.36$ ,  $p < 0.001$ ), the variable was orthogonalized and its residuals were added as a variable *Onset* to the models. We included also the factor *Variety* (German, Austrian), as well as the random intercepts *Speaker* (68 values), *Word* (606 values) and *Syllable* (1717 values). For a more detailed analysis of speaker dependency, separate models were built with *Speaker* random slopes, for each of the four acoustic cues.

The mixed effects logistic regression models were built using the `glmer()` function of the `lme4` package in R [27]. We started with full models, including all independent variables and their interactions (two-way or three-way). We, then, iteratively removed non-significant predictors and interactions as long as the model would still significantly improve given their AIC values, their degrees of freedom [28, 29] and a model-comparison using the `anova()` function. The threshold significance value was set at  $\alpha = .05$  for all tests.

## 3. Results and Discussion

### 3.1. Prosodic boundary detection performance

We present in Figure 2 the precision-recall curves obtained for boundary detection in the two varieties of German. Averaging across all speakers in a dataset, we attained an AUC of 0.308 for Kiel and an AUC of 0.215 for GRASS (difference significant at  $p < 0.01$ ). As seen in the figure, the system performed bet-

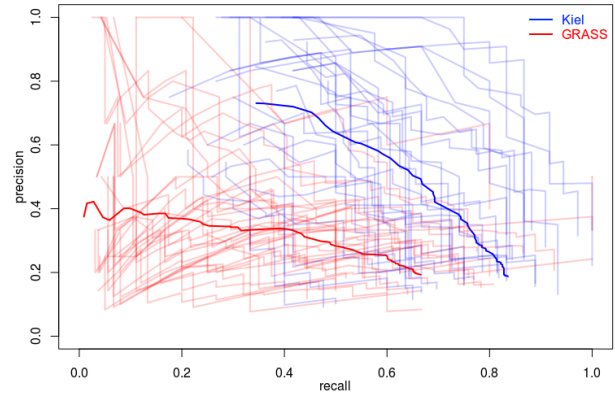


Figure 3: Precision-recall curves obtained on the two datasets. Solid lines are used for illustrating the results over the entire corpus, while each semi-transparent line represents the curve obtained for one speaker.

ter on both datasets compared to our random baseline (red and blue circles). Considering the parameter values giving the best performance on the two datasets, an F-score of 0.582 and 0.370 would be obtained for Kiel and GRASS corpus, respectively.

While significant differences have been obtained between the two datasets, we were also interested to analyse how the tool performed on a per-speaker basis. Figure 3 illustrates the per-speaker precision-recall curves for the two corpora, as well as the overall curve. One can observe important differences between speakers of the same variety, sometimes even larger than the differences in performance between the varieties themselves.

### 3.2. Contribution of features in the two varieties

To analyse how strongly the acoustic features used by our system contribute to the perception of boundaries by German vs. Austrian listeners (as given by the perception-based annotations), a mixed effects logistic regression model was built with the perceived boundary labels as dependent variable. Table 1 shows the final model. All acoustic cues contribute significantly to the perception of a boundary, with *Pause* having the highest effect size ( $\beta = 10.80$ )<sup>1</sup> and *f0\_Reset* having the lowest effect size ( $\beta = -1.23$ ). We also found a significant interaction between *Onset* and *Pause*: The effect of onset is significantly lower in the presence of a pause ( $\Delta = 0.37$ ) than in its absence ( $\Delta = 0.52$ ), indicating that even the orthogonalized onset cue contains information that is similar to the one included in the pause feature.

With respect to variety, the existence of a *Pause* showed to be an equally strong cue in both varieties. In the German data, however, the difference in *Onset* duration across words with no boundary separating them and those with boundaries between them tended to be significantly lower (0.43) than in the Austrian data (0.49). The contrast in *f0\_reset* between boundary and non-boundary tokens was significantly higher in the German (0.15) than in the Austrian data (0.13). Overall, significantly fewer boundaries were perceived in the German read speech sentences (0.12% of the tokens were followed by a boundary, compared to 0.15%, in the Austrian data).

These results are, generally, in line with those obtained by

<sup>1</sup>Since all acoustic measures range between 0 and 1, the effect size equals the estimate of the model.

Table 1: *Model for human perception of prosodic boundaries as predicted by the four acoustic measures. The models include the significantly contributing random intercepts Speaker, Syllable, and Word.  $N = 15.046$ ,  $AIC = 3596$ , 7.*

|                     | estimate | z-value | p-value |
|---------------------|----------|---------|---------|
| Intercept           | -6.80    | -18.64  | < 0.001 |
| Variety(G)          | -0.85    | -3.62   | < 0.001 |
| Onset               | 8.98     | 9.70    | < 0.001 |
| Pause(Y)            | 10.80    | 17.65   | < 0.001 |
| f0_Reset            | -1.23    | -3.06   | < 0.01  |
| Nucleus             | 4.06     | 16.60   | < 0.001 |
| Variety(G):Onset    | -5.86    | -6.48   | < 0.001 |
| Variety(G):Pause(Y) | -0.45    | -0.37   | < 1.0   |
| Variety(G):f0_Reset | 2.87     | 5.89    | < 0.001 |
| Onset:Pause(Y)      | -3.82    | -2.51   | < 0.05  |

Ulbrich [21], based on recordings made with trained broadcast speakers. She found a significantly smaller number of sentence-medial phrase boundaries in the German data (as in our data) and that German speakers exhibited smaller changes in speech rate between boundary/non-boundary conditions (which corresponds to the smaller changes in onset duration for German than for Austrian speakers, in our data). Whereas she reported that f0 reset was equal among all varieties, we observed here that Austrians show less variation with respect to that feature than German speakers. Most importantly, our results show that the existence of a pause, independent of its absolute length, is a variety-independent cue to prosodic boundary perception in German.

Taking into account the findings of previous studies showing that boundaries in German tend to be generally marked not only by one, but multiple acoustic cues (e.g. [17]), we analysed also the distribution of the cues at boundary positions. It revealed that 81.4% and 77.3% of the prosodic boundaries (in the Austrian German and German data, respectively) are marked by more than one cue. Seeing how frequent is the use of multiple cues for marking boundaries in both varieties, one can envision the use of this information in the detection system.

Table 2: *Speaker-dependency of acoustic cues: ANOVA comparisons of models with and without random Speaker slopes, separately for the different acoustic cues, separately for the German and Austrian data.  $N_{GRASS} = 6955$ ;  $N_{KIEL} = 8091$ .*

| Features | Variety | $\Delta AIC$ | $\chi^2$ | p-value |
|----------|---------|--------------|----------|---------|
| Pause    | Germany | 31,1         | 37.07    | < 0.001 |
|          | Austria | 74.2         | 78.18    | < 0.001 |
| f0_Reset | Germany | 45,8         | 0        | < 1     |
|          | Austria | 24.3         | 26.31    | < 0.001 |
| Onset    | Germany | 2.8          | 1.18     | < 0.55  |
|          | Austria | 44.0         | 47.92    | < 0.001 |
| Nucleus  | Germany | 24.5         | 30.49    | < 0.001 |
|          | Austria | 33.4         | 37.44    | < 0.001 |

### 3.3. Speaker dependency of acoustic cues

Besides the observed language-variant significant difference, we also noted important inter-speaker variation (AUC standard deviation of 0.119 and 0.120 for the GRASS and the Kiel corpus, respectively) For that reason, we will investigate next whether these four acoustic cues are used in a language-variety versus speaker-specific manner.

In order to analyse any possible speaker dependency of the cues, we built models with *Speaker* random slopes, for each of the four acoustic cues and separately for the German and Austrian data. Models with and without random slopes were then compared using the `anova()` function to estimate whether their contribution to the models is significant. Table 2 shows the results of the ANOVA comparisons by means of the following statistical measures:  $\Delta AIC$ ,  $\chi^2$  and p-values.

For Austrian German, the random speaker slopes significantly improve the models in combination with all acoustic cues, indicating that all of these features are highly speaker dependent. In the German data, only the existence of pauses and the nucleus duration are speaker dependent, f0\_reset and onset duration seem not to be speaker dependent. The latter result is in line with the findings of Holzgrefe-Lang and colleagues [17] which showed that both pitch reset and final lengthening need to occur simultaneously in order for native Germans to perceive an intonational phrase boundary. The necessity of these two cues for the perception of a prosodic boundary in German would explain why they were found to be used by the majority of speakers in our German data, although it is surprising to see that nucleus duration (another cue marking final lengthening) was employed in a speaker-dependent manner.

## 4. Conclusions

To summarize, we have compared here a signal-based prosodic boundary annotation tool on two varieties of German. We have seen a difference in performance between the two datasets, which seems to be congruent to the differences found in prosodic boundary marking between the two varieties. Furthermore, these differences are largely consistent to the findings of previous studies. In terms of the performance reached, we obtained encouraging results for using the system as a first pass annotation tool followed by a manual correction step.

The present study represents a feasibility study for variety-independent prosodic boundary detection. In the next step, we would like to expand the investigation to more spontaneous speech materials. Since important differences exist between read and spontaneous speech, it would be interesting to investigate how the proposed approach scales to conversational speech. Another direction this study opens is the integration of linguistic knowledge into the annotation tool, by exploring different ways of combining the individual cues.

## 5. Acknowledgements

The work done by Barbara Schuppler was supported by the Elise Richter grant (V638 N33) from the Austrian Science Fund. Bogdan Ludusan's work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 799022. We would like to thank the transcribers David Ertl and Katerina Petrevska for their efforts.

## 6. References

- [1] T. Kislser, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer, Speech & Language*, vol. 45, pp. 326–347, 2017.
- [2] B. Schuppler, S. Grill, A. Menrath, and J. A. Morales-Cordovilla, "Automatic phonetic transcription in two steps: forced alignment and burst detection," in *Statistical Language and Speech Processing. SLSP 2014. Lecture Notes in Artificial Intelligence*, L. Besacier, A. Dediu, and C. Martín-Vide, Eds., 2014, vol. 8791, pp. 132–143.
- [3] B. Schuppler, M. Hagmüller, J. A. Morales-Cordovilla, and H. Pessentheiner, "GRASS: the Graz corpus of Read And Spontaneous Speech," in *Proc. LREC*, 2014, pp. 1465–1470.
- [4] M. Avanzi, A. Lacheret-Dujour, and B. Victorri, "ANALOR. A tool for semi-automatic annotation of French prosodic structure," in *Proceedings of Speech Prosody*, 2008, pp. 119–122.
- [5] S. Ananthakrishnan and S. S. Narayanan, "Automatic prosodic event detection using acoustic, lexical, and syntactic evidence," *IEEE Trans. Audio Speech Lang Processing*, vol. 16, no. 1, pp. 216–228, 2008.
- [6] G. Christodoulides, M. Avanzi, and A. C. Simon, "Automatic labelling of prosodic prominence, phrasing and disfluencies in French speech by simulating the perception of naïve and expert listeners," in *Proceedings of Interspeech*, 2017, pp. 3936–3940.
- [7] E. Marsi, M. Reynaert, A. van den Bosch, W. Daelemans, and V. Hoste, "Learning to predict pitch accents and prosodic boundaries in Dutch," in *Proc. of the 41st Annual Meeting of the Association for Computational Linguistics*, 2003, pp. 489–496.
- [8] N. Braunschweiler, "ProsAlign - The Automatic Prosodic Aligner," in *Proc. ICPhS*, 2003, pp. 3093–3096.
- [9] F. Kügler, S. Baumann, B. Andreeva, B. Braun, M. Grice, J. Neitsch, O. Niebuhr, J. Peters, C. T. Röhr, A. Schweitzer, and P. Wagner, "Annotation of German intonation: DIMA compared with other annotation systems," in *Proceedings of ICPhS*, 2019, p. No. 181.
- [10] B. Ludusan and E. Dupoux, "Towards low-resource prosodic boundary detection," in *Proceedings of SLTU*, 2014, pp. 231–237.
- [11] J. Cole, H. Kim, H. Choi, and M. Hasegawa-Johnson, "Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from radio news speech," *Journal of Phonetics*, vol. 35, pp. 180–209, 2007.
- [12] Y. Mo and J. Cole, "Perception of prosodic boundaries in spontaneous speech with and without silent pauses," *The Journal of the Acoustical Society of America*, vol. 127, no. 3, p. 1956, 2010.
- [13] C. Männel and A. D. Friederici, "Neural correlates of prosodic boundary perception in German preschoolers: If pause is present, pitch can go," *Brain Research*, vol. 1632, pp. 27–33, 2016.
- [14] C. Petrone, H. Truckenbrodt, C. Wellmann, J. Holzgrefe-Lang, I. Wartenburger, and B. Höhle, "Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists," *Journal of Phonetics*, vol. 61, pp. 71–92, 2017.
- [15] M. Beckman and J. Edwards, "Lengthenings and shortenings and the nature of prosodic constituency," in *Between the grammar and physics of speech [Papers on Laboratory Phonology 1]*, J. Kingston and M. Beckman, Eds. Cambridge: Cambridge University Press, 1990, pp. 152–178.
- [16] S. Fuchs, J. Krivokapic, and S. Jannedy, "Prosodic boundaries in German: Final lengthening in spontaneous speech," *Journal of the Acoustical Society of America*, vol. 127, no. 3, p. 1851, 2010.
- [17] J. Holzgrefe-Lang, C. Wellmann, C. Petrone, R. Råling, H. Truckenbrodt, B. Höhle, and I. Wartenburger, "How pitch change and final lengthening cue boundary perception in German: converging evidence from ERPs and prosodic judgements," *Language, Cognition and Neuroscience*, vol. 31, no. 7, pp. 904–920, 2016.
- [18] T. Cho and P. Keating, "Effects of initial position versus prominence in English," *Journal of Phonetics*, vol. 37, pp. 466–485, 2009.
- [19] S. Moosmüller, "The interaction of prosody and phonotactics: Resyllabification in three varieties of German," *Italian Journal of Linguistics*, vol. 27, no. 1, pp. 111–132, 2015.
- [20] D. El Zarka, B. Schuppler, C. Lozo, W. Eibler, and P. Wurzwaller, "Acoustic correlates of stress and accent in Standard Austrian German," in *Phonetik in und über Österreich, Veröffentlichungen zur Linguistik und Kommunikationsforschung: 31*, S. Moosmüller, C. Schmid, and M. Sellner, Eds. Vienna: ÖAW Austrian Academy of Sciences Press, 2017.
- [21] C. Ulbrich, "Prosodic phrasing in three German standard varieties," in *Proc. 29th Annual Penn. Linguistics Colloquium*, 2006, pp. 361–373.
- [22] B. Schuppler and M. Zellers, "Prosodic effects on plosive duration in German and Austrian German," in *Proceedings of INTERSPEECH*, 2019, pp. 1736–1740.
- [23] K. J. Kohler, B. Peters, and M. Scheffers, "The Kiel Corpus of Spoken German—Read and Spontaneous Speech. New Edition, revised and enlarged," Available at <http://www.isfas.uni-kiel.de/de/linguistik/forschung/kiel-corpus/>, 2017.
- [24] B. Schuppler, M. Hagmüller, and A. Zahrer, "A corpus of read and conversational Austrian German," *Speech Communication*, vol. 94C, pp. 62–74, 2017.
- [25] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, vol. 5, no. 9/10, pp. 314–345, 2001. [Online]. Available: <http://www.praat.org>, last viewed 18-08-2018
- [26] A. Camacho, *SWIPE: A sawtooth waveform inspired pitch estimator for speech and music*. University of Florida Gainesville, 2007.
- [27] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.
- [28] R. H. Baayen, *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge University Press, 2008.
- [29] N. Levshina, *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2015.