Microprosodic variability in plosives in German and Austrian German

Margaret Zellers[1], *Barbara Schuppler*[2]

[1]Institute for Scandinavian Studies, Frisian Studies and General Linguistics, Kiel University, Germany
[2]Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria

mzellers@isfas.uni-kiel.de, b.schuppler@tugraz.at

Abstract

Fundamental frequency (F0) contours may show slight, microprosodic variations in the vicinity of plosive segments, which may have distinctive patterns relative to the place of articulation and voicing. Similarly, plosive bursts have distinctive characteristics associated with these articulatory features. The current study investigates the degree to which such microprosodic variations arise in two varieties of German, and how the two varieties differ. We find that microprosodic effects indeed arise in F0 as well as burst intensity and Center of Gravity, but that the extent of the variability is different in the two varieties under investigation, with northern German tending towards more variability in the microprosody of plosives than Austrian German. Coarticulatory effects on the burst with the following segment also arise, but also have different features in the two varieties. This evidence is consistent with the possibility that the fortis-lenis contrast is not equally stable in Austrian German and northern German.

Index Terms: phrase boundaries, plosives, phonetic detail, bursts, F0, microprosody, German, Austrian German

1. Introduction

This study investigates microprosodic features of German plosives, analyzing data from two varieties of German (Northern German and Austrian German) of different standard pronunciations of the fortis and lenis plosives [1] and different phonological coarticulation processes [2].

German fortis plosives have been reported to have longer closures, stronger bursts and longer voice onset time (VOT) than lenis plosives [3]. In Austrian German, however, the fortis/lenis contrast does not arise in all positions. [1] found that only velar lenis and fortis plosives differ in aspiration in initial position, while bilabial and alveolar plosives do not differ significantly in aspiration. Additionally, they found that lenis plosives are never aspirated, regardless of position, and that intervocalic fortis plosives have longer closures than intervocalic lenis plosives.

Microprosodic effects—that is, small variations in prosodic features which have not normally considered to contribute to meaning—have nonetheless been shown to arise systematically and to contribute to listeners' ability to perceive particular sounds. One such phenomenon that has often been investigated is variation in fundamental frequency (F0) in the vicinity of plosives. Studies as early as [4, 5] demonstrated that vowels following fortis plosives tend to have higher F0 than vowels following lenis plosives in similar contexts. [6] additionally demonstrated, for northern German, that "macro" features of prosody (in this case, the larger F0 movements associated with intonation) can modulate this effect; if the F0 contour through the plosive is rising or level, F0 is relatively higher following a fortis plosive as compared to a lenis plosive, but if the F0 contour is falling, higher F0 occurs in the vowel *preceding*, rather than following, the fortis plosive. [7] further reported that modifying F0 in a vowel preceding a plosive could influence the perception of the plosive as fortis or lenis. [8] also confirms these findings for the German variety of German. Similar findings have been reported for a variety of languages; within the Germanic family, [9] uses this F0 effect to argue that the fortis-lenis contrast in American English rests on the feature +/-stiff rather than on voicing. [10] and [11], on the other hand, find that fortis plosives with a relatively longer VOT have relatively lower F0 in the following vowel than fortis plosives with a relatively shorter VOT. This result suggests that these two cues to the fortis/lenis contrast could be in a trading relationship in American English. In Afrikaans, there appears to be a shift in terms of which cues to the fortis/lenis contrast are most important: older listeners depend more on VOT to make this distinction, while younger listeners depend more on F0 cues.

F0 perturbations are not the only microprosodic variations which are meaningful for plosives. Burst frequency has been shown to have different shapes and center points for different places of articulation of plosives [12, 13]. These studies found that bursts from labial plosives were harder to classify than those from alveolar or velar plosives, but that nonetheless each place of articulation had a characteristic energy distribution in the burst. [14] report that burst intensity also varies in northern German plosives as a function of strength of an adjacent prosodic boundary; fortis plosives showed both a shorter VOT and a lower burst intensity when they were adjacent to stronger prosodic boundaries, while lenis burst intensities and VOT were not affected.

For Austrian German, there is less knowledge than for northern German about microprosodic effects and segmental variation. For vowels, prosodically weak positions are prone to monophthongization [15] and vowel duration and quality are affected by whether they occur in stressed or unstressed syllables [16, 17]. [18] found differences between Austrian German and northern German in the prosodic features differentiating phrase-boundary-adjacent plosives from non-phrase-boundary-adjacent plosives, with plosive duration distinguishing between boundary-adjacent and non-boundary- adjacent plosives in Austrian German but not in northern German; this contrasts with [14]'s report for northern German, but may result from the fact that [18] used only word-initial plosives. [18] also found that stress marking was driven primarily by burst duration in the northern German data but by closure duration in the Austrian German data. [14] propose that features related to maintaining the fortis/lenis contrast should be less affected by prosodic boundaries, while those which are not of use to the contrast may be more variable between boundary-adjacent and non-boundary-adjacent positions.

Given that in northern German and Austrian German lenis and fortis plosives are distinguished by a different set of acoustic cues, and that they have different effects on the phonetic detail of the surrounding segments, we hypothesize that these varieties have different strategies for prosodic strengthening, both in terms of stress marking and prosodic boundary marking. However, it is important to clarify the distinctions between the two varieties in terms of which cues vary in systematic ways in relation to both the fortis/lenis contrast and in prosodically strong or weak positions. Thus, the current study investigates the relationship of microprosodic variation, namely F0 perturbations, burst intensity, and burst Center of Gravity (CoG), to plosive place of articulation, the fortis/lenis contrast, and prosodic structure, using data from both northern and Austrian German.

2. Materials and methods

2.1. Speech material

This study draws from the same speech material as [18], comprising read speech from two corpora, the Kiel Corpus of Spoken German [19] and the Graz Corpus of Read and Spontaneous Speech (*GRASS*) [20]. The speakers in these corpora represent two different varieties of German (i.e., (northern) German and Austrian German), and the read material in both corpora overlaps substantially, enabling us to make a more direct comparison. The subset of the data involved in this study is slightly different to that selected by [18], since additional hand-corrections of annotations were completed since the time of that study, as well as semi-automated corrections of the F0 contours (cf. Section 2.2 below). The re-selection of the data subset led to a better balance of plosive tokens in the current study (407 tokens spoken by 10 speakers from the Kiel corpus and 329 tokens spoken by 38 speakers from GRASS corpus) than that available at the time of the previous study.

Both corpora contain high quality audio recordings but were annotated with different methods. The Kiel Corpus was fully manually annotated phonetically and prosodically (cf. [19]), while the *GRASS* corpus was automatically segmented using MAUS [21, 22]. Thus, word-initial plosives in *GRASS* were manually separated into closure and burst, and misalignments hand-corrected. Plosive bursts were considered to begin at the first visible perturbation of the signal following the plosive closure, and ended either when periodicity became visually apparent in the waveform, or in the case of plosives followed by a fricative rhotic, when a change in the visual signal corresponded with a change in the auditory impression.

2.2. Acoustic measurements

Acoustic analysis was carried out using Praat [23]. Since automatic F0 contour extraction in the vicinity of plosives is particularly prone to error, all F0 contours were hand-checked, and a smoothed version of the whole contour was generated, using the *mausmooth* tool [24]. During this process, octave errors were corrected, and spurious F0 points arising from, for instance, high frequency noise in plosive bursts were removed. Otherwise, no changes were made to the contours. The generation of the smoothed contour proceeded according to the standard parameters in *mausmooth*.

Macro- and micro-F0 features were then extracted. The macro-F0 feature is the overall trajectory of the F0 contour across the plosive, and is operationalized as the difference between the F0 in the middle of the vowel following the plosive minus the F0 in the middle of the vowel preceding the plosive, as calculated using the smoothed contour, to avoid redundancy with the microprosody extraction. The numerical value is then classified as either rising, falling, or level (*Pitch_Lab*), with rises or falls of less than 1 semitone classified as level.

The micro-F0 features are the values of the mean of the last two pitch points before the plosive closure (in the remainder of the text referred to *F0-INTO*), and the mean of the first two pitch points following the release of the plosive (*F0-OUTOF*; whether or not a burst is present); that is, the last pitch points in the preceding segment and the first pitch points in the next segment. These were further limited in that they were required to fall within the final half of a segment preceding the plosive, or the first half of a segment following a plosive, so that we could be relatively confident that any microprosodic effects observed could reasonably have arisen based on the plosive articulation.[1]

Center of gravity (*burst-COG*) and mean intensity (*burst-DB*) were calculated for the plosive bursts. Each burst was extracted from the signal using Praat's Extract Part function with a rectangular window. Praat converted the sound into a spectrum, and center of gravity was extracted with weighting by the power spectrum. The intensity was calculated from the extracted burst using Praat's default settings and is reported in decibels (dB) relative to the auditory threshold (2x10-5 Pascal per Praat's settings).

2.3. Statistical analysis

We built linear mixed effects regression models with the acoustic measurements as dependent variables: burst duration (*burst-DUR*), *burst-COG, burst-DB, F0-INTO* and *F0-OUTOF*. All models included the independent variables *Speechrate*, calculated as the average syllable duration of the utterance, *Plosive* (values: /p/, /b/, /d/, /t/, /g/, /k/) or *Fortis_Lenis* (values: fortis and lenis), and variables describing the preceding (*Prev_*) and following (*Foll_*) segments of the token analyzed: their manner of articulation (*_Manner* with the values pause, nasal, vowel, fricative, plosive, approximant), whether they were voiced or not (*_Voicing*) and their duration in seconds (*_Duration*). *Boundary* (yes, no) described whether the word-initial plosive occurred at a prosodic phrase boundary or not. As durational features may vary between content and function words [25], and as stressed syllables tend to have longer realizations than unstressed ones [26], we added the variables *Word_Class* (values: content vs. function word) and *Word_Stress*, which described whether the initial syllable of the tokens bore canonical stress. Although speakers do not obligatorily produce lexical stress on any given token in connected speech, this automatic classification enabled an improvement of our study without substantial annotator time investment. We additionally included the variables *Sex* (male, female), *Variety* (German, Austrian), as well as the random variables *Speaker*, *Word*, and *Sentence*.

We used the `lmer()` function of the `lme4` package in *R* [27] to build the linear mixed effects regression models. After including all independent variables and their interactions (two and three-way) into the models, predictors and interactions were reduced by stepwise backward selection. Correlating variables were either added separately or orthogonalized. Non-significant factors and interactions were removed as models would still significantly improve as given by their AIC value and their degree

[1] Note, however, that since some of our plosives were preceded or followed by voiceless segments, these values could in some cases not be extracted.

of freedom. Random variables were only kept in the model if they improved the model as given by model comparison using the `anova()` function [28, 29]. The threshold significance value is set at α = .05 for all tests.

3. Results and Discussion

3.1. Microprosodic effects in F0

After reducing all non-significant factors and interactions, the final model for *F0_INTO* had the following syntax: $lmer(F0\_INTO \sim (Boundary + Variety)^2 + (Pitch\_Lab + Variety)^2 + (Variety + Word\_Stress)^2 + (Speechrate + Variety)^2 + (Foll\_Manner + Variety)^2 + (1|Word) + (1|Sentence) + (1|Speaker)$. Overall, the models showed that F0 going into a plosive is highly significantly lower for tokens within phrase boundaries than for boundary adjacent tokens (β = -1.60, t = -3.39, $p < .001$), and this effect is significantly higher in the German than in the Austrian data (β = 2.13, t = 3.43, $p < .001$). Furthermore, F0 upon entering a plosive during a rising contour is relatively higher than F0 upon entering a plosive during a level or falling contour (β = 0.64, t = 2.54, $p < .05$), and this effect is stronger in German than in Austrian read speech (β = -0.78, t = -2.75, $p < .01$). Finally, plosives at the onset of stressed syllables tend to be realised with significantly lower *F0_INTO* values than plosives at the onset of not-stressed syllables (β = -1.12, t = -2.28, $p < .05$), and also this effect is stronger in the German than in the Austrian data (β = 1.17, t = 2.19, $p < .05$).

The final model for *F0_OUTOF* had the following syntax: $lmer(F0\_OUTOF \sim (Pitch\_Lab + Variety)^2 + (Speechrate + Variety)^2 + Prev\_Manner + Prev\_Voicing + (1|Word) + (1|Sentence) + (1|Speaker)$. As expected, F0 upon leaving a plosive during a rising contour is overall relatively higher than F0 upon leaving a plosive during a falling contour (β = 1.14, t = 5.42, $p < .001$). For a level contour the values tend to be smaller than for a falling contour, but this effect was only significant in the German data (β = 0.85, t = 2.23, $p < .05$). In both varieties, *F0_OUTOF* is significantly lower at higher speech rates (β = -32.43, t =-2.84, $p < .01$), and this is the factor of this model with the highest effect size. In addition, the manner of articulation of the previous context showed to play a significant role in both varieties: *F0_OUTOF* tends to be lower after fricatives (β = -1.99, t =-1.91, $p < .1$), nasals (β = -1.33, t =-2.38, $p < .05$), stops (β = -1.88, t =-1.77, $p < .1$) and vowels than after approximants. In both varieties lenis plosives showed to have significantly higher values for *F0_OUTOF* (β = 2.85, t =2.19, $p < .05$) than fortis plosives, and this effect was smaller in the Kiel than in the GRASS corpus (β = -2.87, t = -2.17, $p < .05$) and smaller at higher speech rates (β = -44.42, t = -2.21, $p < .05$).

3.2. Microprosodic effects in burst features

After reducing all non-significant factors and interactions, the final model for *Burst_DUR* had the following syntax: $lmer(Burst\_DUR \sim (Variety + Stressed)^2 + Speechrate + (Prev_Manner + Plosive)^2 + (Prev_Voicing + Variety)^2 + (Foll_Manner + Variety)^2 + (1|Word) + (1|Sentence) + (1|Speaker)$. Overall, the bursts of /k/ (β = 4.47e-02, t = 10.26, $p < .001$), /g/ (β = 1.40e-02 , t = 3.34, $p < .001$) and /t/ (β =2.74e-02 , t = 7.39, $p < .001$) are significantly longer than /d/ in both corpora, with /k/ and /t/ being especially long; see Figure 1. Furthermore, we found effects of previous and following context. When vowels follow word-

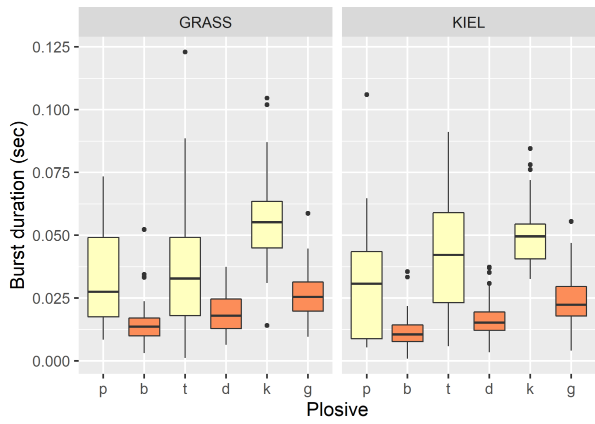

Figure 1: *Duration of plosive bursts across the six plosive types, in GRASS and the Kiel Corpus.*

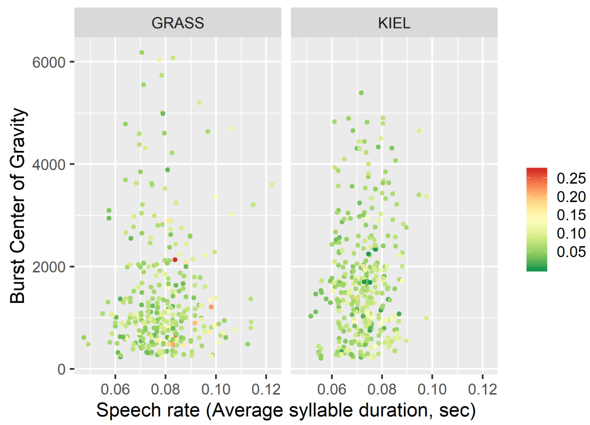

Figure 2: *Burst center of gravity as conditioned by speech rate (average syllable duration, ASD) and previous segment duration, both measured in seconds.*

initial plosives, bursts tend to be longer than for approximants (β = 1.43e-02, t = 5.45, $p < .001$), with an even stronger effect in the Kiel corpus (β = -8.24e-03, t = -2.84, $p < .01$). Burst duration is marginally significantly smaller in plosives following voiced rather than voiceless segments (β = -1.06e-02, t = -1.66, $p < .1$) and this effect is stronger for the German than for the Austrian data (β = 4.66e-03, t =1.81 , $p < .1$).

The final model for burst center of gravity had the following syntax: $lmer(Burst\_COG \sim Speechrate + (Foll_Manner + Plosive)^2 + (Foll_Manner + Variety)^2 + (1|Word) + (1|Sentence) + (1|Speaker)$. In general, in both varieties, CoG values were significantly higher for /d/ (β = 1169.98, t = 2.26, $p < .05$), /g/(β = 1190.30, t = 2.52, p <).05, /k/(β = 2544.77, t = 3.62, $p < .001$) and /t/(β = 2448.50, t = 9.48, $p < .001$) than for /b/ and /p/, with /k/ having the highest CoG values. These differences among the different plosives are smaller for /k/ (β = -1652.58, t =-2.27 , $p < .05$) and /t/(β = -684.77, t = -2.39, $p < .05$) when preceding a vowel. The effect of segmental context (i.e., that the segment following the plosive is a vowel) is marginally significantly stronger in the Kiel corpus than in the GRASS corpus (β = 329.29, t = 1.89, $p < .1$), as shown in Figure 2.

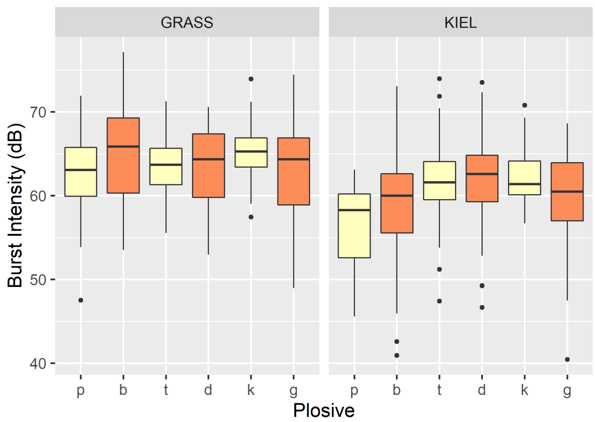

Figure 3: *Burst intensity in decibels (dB). Fortis plosives are shown in light blue, lenis in dark blue. Plosive place of articulation is more strongly distinguished by burst intensity in northern German than in Austrian German.*

The final model for burst intensity had the following syntax: $lmer(Burst_DB\ Boundary + Prev\_DUR + Follow\_Dur + (Prev\_Manner + Fortis\_Lenis + Variety)^2 + (Fortis\_Lenis + Variety)^2 + (1|Word) + (1|Sentence) + (1|Speaker))$. In the Kiel corpus, lenis plosives showed to have significantly lower values for burst intensity (β = -4.37, t = 2.35, p $<$.05) and higher values when plosives are already preceded by plosives (β = 4.66, t = -3.39, p $<$.001). Furthermore, in the Kiel data, the alveolar bursts /d/ (β = 4.49, t = 3.09, p $<$.01) and /t/ (β = 3.32, t = 2.90, p $<$.01) were produced with higher intensity than bilabial plosives. In both varieties, we observed the lowest intensity values for /g/ (β = -11.18, t = -2.31, p $<$.05). Burst intensity for fortis and lenis plosives is shown in Figure 3. In both varieties, context also played a role. After fricatives, the alveolar plosives (/d/:β = 3.68, t = 1.78, p $<$.1; /t/: β = 4.44, t = 2.03, p $<$.05) and the velar plosives (/g/: β = 3.90, t = 1.81, p $<$.1, /k/: β = 7.05, t = 3.12, p $<$.01) are produced with significantly higher intensity than when preceded by vowels or approximants. In addition, /t/ plosives have a significantly lower intensity when preceded by nasals compared to when preceded by approximants (β = -3.60, t = -2.39, p $<$.05). Finally, we also observed effects of following context in both varieties. When followed by stops (β = 14.34, t = 2.17, p $<$.05), /g/ plosives have higher intensity values than when followed by approximants. Among all manners of articulation, however, /g/ has an overall higher intensity when followed by voiced than when followed by voiceless segments (β = 11.93, t = 2.15, p $<$.05).

3.3. Discussion

For the features which are associated with plosive identity and the fortis/lenis distinction, systematic effects in these features tended to be stronger in northern German than in Austrian German. Previous work by [1] already suggested that, in at least some contexts, the fortis/lenis contrast is not as strongly marked in Austrian German as in other varieties, and our findings are also consistent with this result. It is possible that the fortis/lenis distinction in Austrian German is in the process of collapsing, or of shifting to rest on a different parameter; Zürich German, for example, bases its fortis/lenis contrast on closure duration, with fortis closures up to 4 times longer than lenis closures [30]. However, more evidence comparing varieties, as well as longitudinal studies, are necessary to determine whether this is actually the case.

[14] suggest that cues that are crucial to the fortis/lenis distinction should be less influenced by prosodic factors than cues that do not contribute to the segmental distinction. For burst duration, which was shown by [18] to be the driver of plosive duration differences related to lexical stress in northern German but not in Austrian German, there appear to be more structured coarticulatory effects in the northern German data than in the Austrian German data; this is marginally also the case for center of gravity of the burst. On the other hand, burst intensity, which distinguishes fortis and lenis plosives in both varieties, shows similar coarticulatory effects in both varieties. If prosodic factors as defined by [14] are taken to mean those signalling phrasing, then these results are not inconsistent. Rather, they contribute to the case for a different fortis/lenis contrast strength between Austrian and northern German; a more stable contrast in northern German allows cue trading patterns to arise, whereas in Austrian German, cue trading may be lost into less structured variation if the fortis/lenis contrast is in the process of weakening or shifting its acoustic basis.

We found a contrast in micro-F0 effects in fortis and lenis plosives in both varieties in our study, although without the interaction between macro-F0 and micro-F0 reported by [6]. F0 is higher following lenis plosives than following fortis plosives; in contrast to the burst features, this effect is stronger in the Austrian German data than in the northern German data, supporting the proposal that the fortis/lenis contrast in Austrian German, rather than disappearing, may be transitioning to a different set of acoustic cues. Plosives at the onset of stressed syllables arise with lower F0 in the preceding syllable in both varieties, with a stronger effect in northern German.

4. Conclusions

Based on 736 word-initial plosive tokens from German and Austrian German read speech, we found that micro-F0 effects and features of plosive bursts contribute to the fortis/lenis distinction in Austrian German and northern German, but that variations in these acoustic features arise to different degrees in the two varieties, both in terms of the contrast itself and in terms of what kinds of variability shows in prosodically prominent locations. It remains to be seen whether listeners are able to use this variation in perception to identify prosodically strong locations. Future research will give a fuller picture of the functionality of these prosodic variations for speakers and listeners in both of these varieties.

5. Acknowledgements

The work by Barbara Schuppler was supported by the Elise Richter grant (V638 N33) from the Austrian Science Fund. We would like to thank the transcribers David Ertl, Katerina Petrevska (Graz) and Jens Jäger (Kiel) for their efforts.

6. References

[1] S. Moosmüller and C. Ringen, "Voice and aspiration in Austrian German plosives," *Folia Linguistica*, vol. 38, pp. 43–62, 2004.

[2] B. Schuppler, M. Adda-Decker, and J. A. Morales-Cordovilla, "Pronunciation variation in read and conversational Austrian German," in *Proc. INTERSPEECH*, 2014, pp. 1453–1457.

[3] K. J. Kohler, "Phonetic explanation in phonology: the feature fortis/lenis." *Phonetica*, vol. 41 3, pp. 150–74, 1984.

[4] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105–113, 1953.

[5] M. Haggard, S. Ambler, and M. Callow, "Pitch as a voicing cue," *The Journal of the Acoustical Society of America*, vol. 47, no. 2B, pp. 613–617, 1970.

[6] K. J. Kohler, "F0 in the production of lenis and fortis plosives," *Phonetica*, vol. 39, no. 4-5, pp. 199–218, 1982.

[7] ——, "F0 in the perception of lenis and fortis plosives," *The Journal of the Acoustical Society of America*, vol. 78, no. 1, pp. 21–32, 1985.

[8] M. Jessen, *Phonetics and phonology of tense and lax obstruents in German*. John Benjamins Publishing, 1999, vol. 44.

[9] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, pp. 425–441, 2009.

[10] A. A. Shultz, A. L. Francis, and F. Llanos, "Differential cue weighting in perception and production of consonant voicing," *The Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. EL95–EL101, 2012.

[11] O. Dmitrieva, F. Llanos, A. A. Shultz, and A. L. Francis, "Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English," *Journal of Phonetics*, vol. 49, pp. 77–95, 2015.

[12] K. N. Stevens and S. E. Blumstein, "Invariant cues for place of articulation in stop consonants," *The Journal of the Acoustical Society of America*, vol. 64, no. 5, pp. 1358–1368, 1978.

[13] S. E. Blumstein and K. N. Stevens, "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," *The Journal of the Acoustical Society of America*, vol. 66, no. 4, pp. 1001–1017, 1979.

[14] C. Kuzla and M. Ernestus, "Prosodic conditioning of phonetic detail in German plosives," *Journal of Phonetics*, vol. 39, pp. 143–155, 2011.

[15] S. Moosmüller, "The process of monophthongization in Austria (reading material and spontaneous speech)," in *Papers and Studies in Contrastive Linguistics*, 1998, pp. 9–25.

[16] ——, *Vowels in Standard Austrian German. An Acoustic-Phonetic and Phonological Analysis.* Habilitation, University of Vienna, 2007.

[17] D. El Zarka, B. Schuppler, C. Lozo, W. Eibler, and P. Wurzwallner, "Acoustic correlates of stress and accent in Standard Austrian German," in *Phonetik in und über Österreich, Veröffentlichungen zur Linguistk und Kommunikationsforschung: 31*, S. Moosmüller, C. Schmid, and M. Sellner, Eds. Vienna: ÖAW Austrian Academy of Sciences Press, 2017.

[18] B. Schuppler and M. Zellers, "Prosodic effects on plosive duration in German and Austrian German," in *Proceedings of 20th Interspeech, Graz, Austria*, 2019, pp. 1736–1740.

[19] 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.

[20] B. Schuppler, M. Hagmüller, and A. Zahrer, "A corpus of read and conversational Austrian German," *Speech Communication*, vol. 94C, pp. 62–74, 2017.

[21] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in *Proceedings of ICPhS*, 1999, pp. 607 – 610.

[22] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326 – 347, 2017.

[23] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 314–345, 2001. [Online]. Available: http://www.praat.org, last viewed 18-08-2018

[24] F. Cangemi. (2015) Mausmooth. [Online]. Available: http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

[25] A. Bell, J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *Journal of Memory and Language*, vol. 60, pp. 92–111, 2009.

[26] P. Ladefoged, *AUC Course in Phonetics, 2nd edition*. New York: Hartcourt, Brace, Javanovich, 1982.

[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.

[30] S. A. Fulop, "Acoustic correlates of the fortis/lenis contrast in Swiss German plosives," *The Journal of the Acoustical Society of America*, vol. 94, no. 3, pp. 1866–1866, 1993.