Interspeech 2018
2-6 September 2018, Hyderabad

Truncation and compression in Southern German and Australian English

Jenny Yu[1,2] *and Katharina Zahner*[3]

[1]The MARCS Institute, Western Sydney University, Australia
[2]ARC Centre of Excellence for the Dynamics of Language, Australia
[3]University of Konstanz, Germany

jenny.yu@westernsydney.edu.au, katharina.zahner@uni-konstanz.de

Abstract

Nuclear pitch accents are realized differently when there is little sonorant material (as in monosyllabic compared to disyllabic words): Southern British English speakers compress rises and falls, while Northern German speakers truncate falls and compress rises [1] (Grabe 1998). This leads to different phonetic surface patterns for final falls. Within these languages, dialectal variation affects alignment and the frequency of occurrence of nuclear tunes. We test whether the differences in compression and truncation use are a stable cross-linguistic phenomenon (and occur in other varieties of English and German) or whether they are limited to the varieties tested in [1]. Here, we investigated productions of rises and falls in Australian English and Southern German in words with different proportions of sonorant material. Australian English speakers compressed rises and falls, while Southern German speakers only compressed rises but truncated falls, consistent with Grabe's findings for Southern British English and Northern German. This indicates consistent use of strategies within a language, even though the varieties under investigation display other phonetic differences from previous varieties tested. We discuss implications of these findings for automatic labelling.

Index Terms: speech production, pitch accent, compression, truncation, regional variation, German, Australian English

1. Introduction

Transmitting pitch information requires time and voicing, and when there is limited sonorant material (for instance, when a word is short or contains short vowels) it can be difficult to realize an intended intonational contour. Two strategies languages adopt to cope with this problem are *compression* and *truncation* (e.g., [1],[2]). Compression involves accelerating the realization to accommodate the full contour within a shorter period. In truncation, the pitch slope remains unchanged and simply ends earlier when voiced material runs out, so that the original contour is never completed (but see [3] for other strategies).

Grabe [1] found that English and German speakers adopt different strategies in this situation. Southern British English and Northern German speakers produced surnames with progressively less voiced material (*Sheafer*, *Sheaf*, and *Shift* in English; *Schiefer*, *Schief*, and *Schiff* in German) in questions versus statements, i.e., in contexts requiring rising and falling contours, respectively. Results showed that for English, rate of fundamental frequency (f0) change increased significantly as the duration of voiced materials decreased in three steps in both falling and rising contexts. This indicates that English speakers were compressing pitch movements when segmental material decreased for both rises and falls. German speakers also showed this pattern of compression for rises. For falls however, their rate of f0 change decreased from the longest word to the shortest, reflecting the use of truncation.

Different phonetic surface patterns for phonological final falls arise from this cross-linguistic difference in the use of strategies. While compressed nuclear falls usually display all underlying tonal targets ((L)ow-toned leading tone, (H)igh accentual tone, and an L phrasal accent), truncated falls may not; even though the phonological structure is the same, see Figure 1, tonal labelling according to GToBI [4].

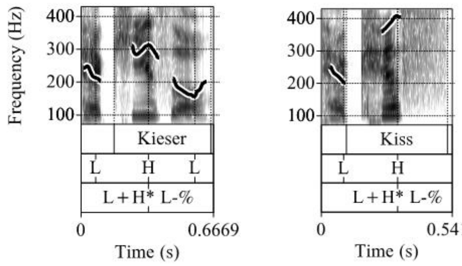

Figure 1: *Realizations of Herr (Mr) 'Kieser' and Herr 'Kiss' produced with a fall by a German speaker (105K, female).*

This difference in phonetic manifestation may not only cause difficulty for human annotators, it also raises challenges for automatic labelling systems, such as AuToBI [5], as these systems derive the phonological structure of contours from phonetic features like tonal targets of the surface contour.

This paper addresses the question of whether the differences in compression and truncation are a stable cross-linguistic phenomenon (and occur in other varieties of English and German) or whether they are limited to the varieties tested in [1]. Consistency in the use of strategies within a language would increase accuracy in automatic classification. We chose two varieties that have been reported to differ to the ones tested in [1] regarding phonetic alignment and distribution of pitch accent types: Southern German and Australian English.

For instance, [6] found that Southern German speakers align prenuclear accentual rises significantly later than Northern speakers. Australian English shows use of rising terminal contours even in statements ("uptalk"; [6]), while such patterns have not been reported for Standard Southern British English. For German, late-peak accents (L*+H) are reported to be a favored pitch accent type in Southern Germany, particularly in so-called hat patterns [7], possibly rendering this accent type more frequent than in Northern Germany. Given these phonetic and phonological differences within the languages, dialectal variation may be the source of the observed differences in [1].

10.21437/Interspeech.2018-2513

2. Method

2.1. Participants

Twelve native speakers each of German (M_{age} = 25.45 years, SD = 3.98 years, 9 females) and of Australian English (M_{age} = 33.86 years, SD = 11.65 years, 5 females) participated for a small payment. German speakers were recruited in Konstanz (n = 8) and Tübingen (n = 4) and grew up in Baden-Wuerttemberg. English speakers were recruited in Sydney.

2.2. Stimuli

Table 1 shows the test items: Twelve equivalent surnames for each of English and German. They formed four sets of three names, with each name in a set representing one step on a continuum that varied in scope for voicing /i/ vowels (e.g., *Sheafer, Sheaf, Shift*). One set was identical to Grabe's [1] (the *sh*-continuum). Three additional sets comprised names varying in structure and sounds (fricatives and stops).

Table 1: *Test items in English and German (italicized)*

	Step 3: Long	**Step 2: Mid**	**Step 1: Short**
sh-Continuum	Sheafer *Schiefer*	Sheaf *Schief*	Shift *Schiff*
s-Continuum	Seefer *Siefer*	Seef *Sief*	Siff *Siff*
g-Continuum	Geesser *Gieser*	Geese *Gies*	Giss *Giss*
k-Continuum	Keesser *Kieser*	Keese *Kies*	Kiss *Kiss*

As in [1], test items occurred in syntactically similar carrier phrases in two lists, one a question, designed to elicit a nuclear rising tone on the test word, and the other a declarative statement for a nuclear falling tone (see (2) and (3) for examples). Each carrier also had a short preceding introductory paragraph to set the scene (1). The test item, in phrase-final position, was followed by a polysyllabic appositional phrase which served as a control (to indicate the underlying phonological specification of a test word in case of truncation). In addition, to mask the systematic durational differences along the continuum, 28 filler stimuli were created for each language: 14 declaratives and 14 questions with English and German surnames in various positions.

(1) Introductory paragraphs
English: Anna and Peter are watching TV. A photograph of this week's National Lottery winner appears. Anna says: ''Look, Peter!''
German: Anna und Peter sehen fern. Ein Lottogewinner wird vorgestellt. Anna sagt: ''Na sowas!''

(2) Carrier phrases for falls (test items are bolded):
English: It's Mr. **Sheafer**! Our new neighbour!
German: Das ist doch Herr **Schiefer**! Unser neuer Nachbar!

(3) Carrier phrases for rises:
English: Isn't that Mr. **Sheafer**? Our new neighbour?
German: Ist das nicht Herr **Schiefer**? Unser neuer Nachbar?

2.3. Procedure

The existing equipment used in the three testing locations (Konstanz, Tübingen, and Sydney) differed slightly; however, the experimental procedure was identical across sites. Subjects were recorded individually in a sound-attenuated booth in one session lasting approximately an hour. The session also included a perception task [8] and L2 production, though these are not reported here. The experiment was controlled using *Presentation* software. Productions were recorded using a microphone and digitized directly onto a PC (44.1 kHz, 16Bit, stereo). Following [1], items were blocked by contour (questions vs. statements) for each language, such that participants completed all 26 statement or question items before completing the other type. Contour order was also counterbalanced and randomly assigned to each subject.

Each trial began with the presentation of an introductory paragraph on screen which subjects were instructed to read silently for context. Participants then read the corresponding carrier phrase out loud. To prevent participants from ignoring contextual information, the carrier phrase only appeared underneath the introductory paragraph after three seconds. Subjects were instructed to read the sentences consistently, and in the same way. In addition, participants were asked to speak naturally, and not exaggerate their speech. After producing the sentence, participants pressed a button to progress onto the next item. Each block began with three practice trials to familiarize the participant with producing statements and questions. The experimenter provided feedback if subjects exaggerated or spoke unclearly during practice trials and monitored the quality of recordings during experimental trials (from outside the booth). Fifty-two test trials (24 critical trials, 28 filler trials) followed, which were presented in an individually randomized order (randomization was done by *Presentation*).

3. Acoustic analysis and data treatment

From 576 productions (2 languages x 12 subjects each x 2 contours x 12 surnames), 82 items were excluded due to pitch tracking errors (glottalization), pronunciation accuracy (e.g., the production of the correct vowel in target), missing data. We further only included items in which a fall was intended in a statement context and a rise in a question context (based on the realization of the polysyllabic appositional phrase in which the underlying phonological form became apparent, second part in Figure 2). The final data set had 494 productions (230 English (113 falls, 117 rises), 264 German (129 falls, 135 rises)). Figure 2 shows an example Praat [7] annotation of a target item, produced by a German speaker (107K, female).

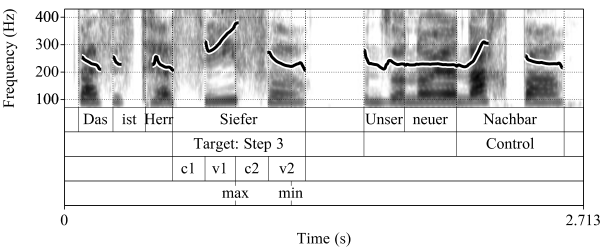

Figure 2: *Spectrogram and f0 contour with annotation (German; "Siefer" (long: step 3), fall). The tiers show the words in German (1), the target word (and its control, 2), its segments (3), and the f0 minimum and f0 maximum of the accentual movement (4) in the target).*

Segment boundaries were set using standard segmentation criteria [9]. For each f0 movement, we annotated the minimum and maximum values. In monosyllabic words, both values usually occurred in the vowel of the test word. In disyllabic words the first tonal target (minimum in rises, maximum in falls) occurred in the first vowel and the second tonal target in the second vowel (maximum in rises, minimum in falls). A Praat script automatically extracted duration and f0 values.

To analyze effects of voicing duration on f0 movement, "rate of f0 change" (RoCh) was calculated by dividing f0-excursion (f0 maximum minus f0 minimum) of a fall or rise by movement duration. Comparing across the three continuum steps (from longest to shortest), a RoCh increase is termed compression, and a stable RoCh truncation (see [1]). Note that f0 excursion was calculated in semitones (st) to account for differences in gender (in Grabe [1] participants were female only).

Duration, RoCh, and f0 excursion were analysed using linear mixed effects regression models (lmer) in R with *language* (English, German), *contour* (fall, rise), and *step* (1, 2, 3) as fixed factors; *word type* (sh-continuum, s-continuum, g-continuum, k-continuum) was added as a control predictor. *Participants* and *items* were modelled as crossed random factors [10, 11]. Random slopes for the fixed factors were added, and retained if the model fit improved [12, 13]. To account for the fact that multiple comparisons were made, we adjusted the p-values using the Holm correction [14] (p.adjust() in R). For completeness, both raw p values (p) and the adjusted p values (p_{adj}) are reported.

4. Results

Duration. First, we checked whether the duration manipulation had the desired effect of providing successively less sonorant material (from longest to shortest sonorant portion, e.g., *Schiefer* to *Schiff*). Figure 3 shows the f0 duration (absolute time difference between f0 min and f0 max) for rises (left panel) and falls (right panel).

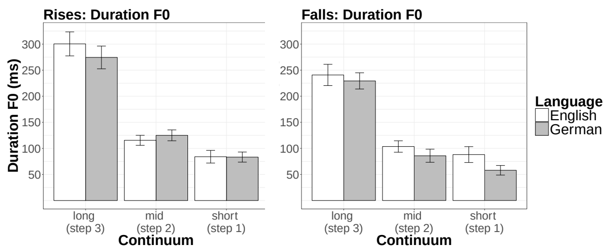

Figure 3: *Duration of f0 movement (ms) for rises and falls.*

Statistical results showed that for both **German and English**, f0 duration **decreased from the longest to the shortest** sonorant portion: For both contours, the difference between step 3 vs. 2, and step 3 vs. 1, i.e., disyllabic vs. monosyllabic items, was highly significant (individual comparisons all $p < 0.0001$, $p_{adj} < 0.0001$); the difference of f0 duration in the two monosyllabic words (step 1 vs. 2) was smaller, but also significant for German falls and English rises (both $p < 0.01$, $p_{adj} < 0.04$), but for German rises and English falls the difference was only numerical post p correction ($p = 0.02$, $p_{adj} = 0.12$ and $p = 0.06$, $p_{adj} = 0.3$, respectively).

F0 analyses. Figure 4 shows RoCh (Hz/st) for falls and rises and Figure 5 illustrates the average f0 excursion (st). A visual inspection shows similarities to Grabe [1]'s findings: While rate of change increases as a function of limited sonorant material for rises in both languages and for falls in English, German falls do not show a step-wise increase (grey bars, right panel Figure 4). Furthermore, the f0 excursion dramatically decreases from "long" to "short" (grey bars, right panel Figure 5), while this decrease is reduced for English falls and English rises (white bars, both panels), and absent altogether for German rises (grey bars, left panel).

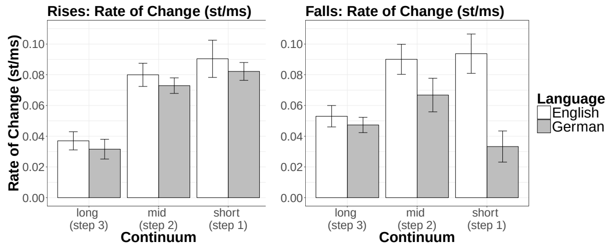

Figure 4: *Rate of change (st/ms) for rises and falls.*

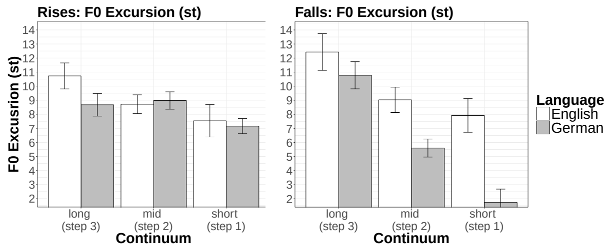

Figure 5: *Average f0 excursion (f0 maximum minus f0 minimum) for rises and falls (in st).*

For **RoCh**, German and English were considered separately, since the interaction between language and contour was significant ($p = 0.003$, $p_{adj} = 0.05$). **In the German dataset,** rate of change was affected differently by step in rises from in falls ($p < 0.0001$, $p_{adj} < 0.0001$), which is why we investigated German rises and falls separately. **For German rises**, there is a significant main effect of step, reflecting the increase in rate of change from the longest to the shortest word, i.e., **compression** ($p < 0.0001$, $p_{adj} = 0.001$): The differences between step 1 vs. step 3 and step 2 vs. step 3 were significant (both comparisons $p < 0.0001$, $p_{adj} < 0.002$), while the difference between step 1 vs. step 2 was not ($p = 0.07$, $p_{adj} = 0.85$). **For German falls**, rate of change was not affected by step ($p = 0.07$, $p_{adj} = 0.85$), indicating **truncation** (difference between step 1 vs. 2: $p = 0.03$, $p_{adj} = 0.44$; 1 vs. 3: $p = 0.18$, $p_{adj} = 1$; 2 vs. 3: $p = 0.02$, $p_{adj} = 0.35$). Contrary to the German dataset, there was **no significant interaction** between contour and step for **RoCh in English** ($p > 0.35$, $p_{adj} = 1$), thus the RoCh was not affected differently by step for falls and rises. Importantly, there was an effect of step ($p < 0.0001$, $p_{adj} < 0.0001$), which reflects the increase in rate of change from the longest to the shortest word, i.e., **compression** for both falls and rises (step 1 vs. 3 and step 2 vs. 3, both comparisons $p < 0.0001$, $p_{adj} < 0.0001$; the difference between step 1 vs. step 2 was not significant post p correction ($p = 0.08$, $p_{adj} = 0.85$).

For **f0 excursion**, language also interacted with contour ($p = 0.004$, $p_{adj} = 0.06$): **For German**, step affected f0 excursion differently in rises from in falls ($p < 0.0001$, $p_{adj} < 0.0001$). **For German rises**, the f0 excursion did not progressively decrease: the differences between step 1 vs. step 3 and step 2 vs. step 3 was not significant (step 1 vs. 3: $p = 0.007$, $p_{adj} = 0.11$; 2 vs. 3: $p = 0.20$, $p_{adj} = 1$), but the f0 excursion was smaller in step 1 than in step 2 ($p < 0.0001$, $p_{adj} = 0.002$). Hence, compression slightly affects the frequency domain, (but more strongly the temporal domain). **For German falls**, f0 excursion significantly decreased for each step of the continuum (all p for individual comparisons < 0.002, $p_{adj} < 0.04$). This corroborates the truncation strategy evident from the analysis of RoCh. **In English**, contrary to the German dataset, there was **no significant interaction** between contour and step for f0 excursion ($p = 0.32$, $p_{adj} = 1$) and the interaction term was removed from the model. Importantly, there was a main effect for step: the difference between step 1 vs. step 2 was not significant ($p = 0.01$, $p_{adj} = 0.15$), but the difference between step 1 vs. step 3 and step 2 vs. step 3 was (p for individual comparisons < 0.0001, $p_{adj} < 0.0001$), i.e., the monosyllabic words have a smaller f0 excursion than the bisyllabic targets. The compression strategy thus also involves a narrowing in the frequency domain in English.

Word type (sh-, s-, k-, g-continuum) did not interact with step for any of the response variables (f0 duration, RoCh, or f0 excursion, all $p > 0.21$, $p_{adj} > 0.54$). Thus, pitch patterns were consistent between all surname continuums.

Figure 6 summarizes our main finding, showing representative f0 contours of rises (compressed in both languages) and falls (compressed in Australian English, truncated in Southern German). Targets were excised from a single speaker in each language.

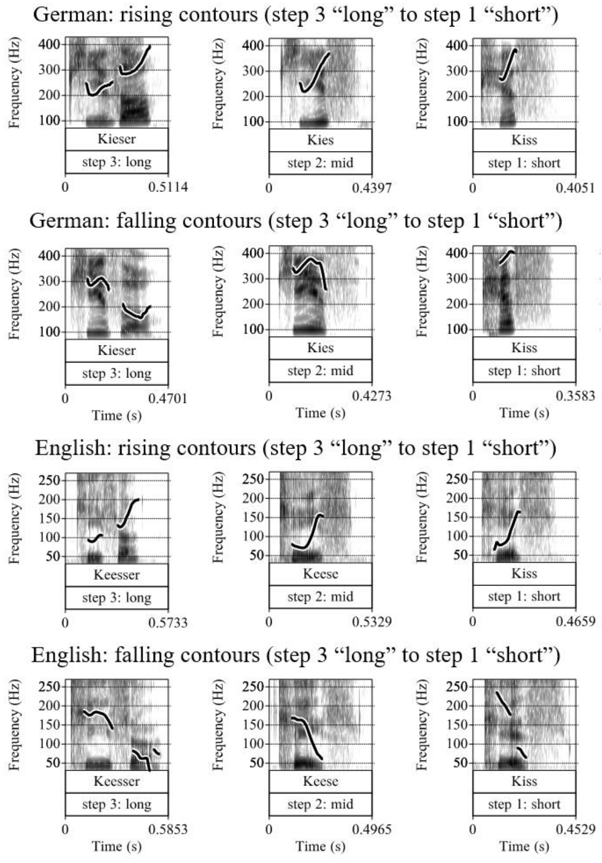

Figure 6: *Exemplar realizations of German and English rising and falling contours for long, mid, and short words in the k-continuum (for German: subject 105K, female; for English: subject 102S, male).*

5. Discussion

For English, rate of f0 change increased with decreasing syllable duration in fall and rise contexts. Thus, Australian English speakers compressed pitch movements for both falls and rises as segmental material decreased. Southern German speakers, however, showed this compression pattern for rises only. For falls, RoCh did not change from the longest to shortest word, reflecting truncation.

Despite alignment and distribution differences between the varieties tested here vs. by Grabe [1] our results are consistent with the findings for Southern British English and Northern German. This suggests that varieties within a language show more consistent use of compression and truncation strategies than previously thought. This interpretation is supported by [15] who found that although strategies could vary between different regional varieties of British English such as Belfast and Leeds, compression was generally adopted in most regions. Similarly, for German, Gilles [16] found that most varieties in Germany truncate for falls. Taken together, these findings depict English as typically a ‘compressing’ language and German as generally a ‘truncating’ language for falls.

Knowing this can help improve the accuracy of automatic intonation labelling systems like AuTobi [5]. Strategies like truncation potentially present challenges for automatic labelling systems that classify the pitch accents depending on phonetic features of the f0 movement, e.g., min and max, slope, mean etc. (see [17]). We argue that the inclusion of sonorant rime duration, syllable structure, and target language in current feature extraction routines may help reduce instances where truncated contours could be mistakenly classified as rises or plateaus (see Figures 1 and 3, German falling contours).

Given the implications of pitch realization strategies on labelling, future research should also investigate the strategies used by speakers of other dialects and languages, as well as of non-native speakers. Phonetic alignment differences based on dialectal variation, for instance, have been reported to be transferred in L2 speech [6]. Compression/truncation use may also potentially carry over in non-native productions. L2 productions collected in this study are currently being analyzed. Research on L2 speech may call for the inclusion of nativeness as a feature in future automatic labelling systems and may also help L2 speakers find additional ways to reduce foreign accent and increase intelligibility.

6. Conclusion

This study reveals consistent use of compensation strategies within a language, even when the varieties under investigation display other phonetic differences from previous varieties tested. It also sheds light on challenges for automatic labelling and offers suggestions on features that should be included for pitch accent classification to increase accuracy in labelling.

7. Acknowledgements

We thank Bettina Braun, Anne Cutler, Heather Kember, and Andrea Weber for feedback and guidance, and Ann-Kathrin Grohe for data collection and helpful comments. This work was funded by the ARC Centre of Excellence for the Dynamics of Language (project ID: CE140100041) and an Australia-Germany Joint Research Cooperation Scheme award to A. Weber (for SpeechNet BaWü) and A. Cutler.

8. References

[1] E. Grabe, "Pitch accent realization in English and German," *Journal of Phonetics,* vol. 26, pp. 129-143, 1998.

[2] Y. Erikson and M. Alstermark, "Fundamental Frequency correlates of the grave word accent in Swedish: the effect of vowel duration," *Speech Transmission Laboratory, Quarterly Progress and Status Report 2-3 KTH, Sweden,* 1972.

[3] J. Hanssen, "Regional variation in the realization of intonation contours in the Netherlands," Doctoral Thesis, LOT, Utrecht, 2017.

[4] M. Grice, S. Baumann, and R. Benzmüller, "German intonation in autosegmental-metrical phonology," in *Prosodic Typology. The Phonology of Intonation and Phrasing*, J. Sun-Ah, Ed., ed Oxford: Oxford University Press, 2005, pp. 55-83.

[5] A. Rosenberg, "AuToBI - a tool for automatic ToBI annotation," in *Interspeech 2010, 11th Annual Conference of the International Speech Communication Association*, Makuhari, Japan, 2010, pp-26-30.

[6] M. Atterer and D. R. Ladd, "On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German," *Journal of Phonetics,* vol. 32, pp. 177-197, 2004.

[7] H. Truckenbrodt, "Upstep of edge tones and of nuclear accents," *Tones and tunes: experimental studies in word and sentence prosody,* vol. 2, pp. 349-386, 2007.

[8] H. Kember, A. Grohe, K. Zahner, B. Braun, A. Weber, and A. Cutler, "Similar prosodic structure perceived differently in German and English," *Interspeech 2017, 18th Annual Conference of the International Speech Communication Association,* Stockholm, Sweden, 2017, pp. 1388-1392.

[9] A. Turk, N. Satsuki, and M. Sugahara, "Acoustic segment durations in prosodic reserach: A practical guide," in *Methods in Empirical Prosody Research*, S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek*, et al.*, Eds., Berlin, New York: De Gruyter, 2006, pp. 1-28.

[10] R. H. Baayen, *Analyzing Linguistic Data. A Practical Introduction to Statistics*. Cambridge: Cambridge University Press, 2008.

[11] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of Memory and Language,* vol. 59, pp. 390-412, 2008.

[12] D. M. Bates, R. Kliegl, S. Vasishth, and H. R. Baayen, "Parsimonious mixed models," *arXhiv preprint,* vol. arXiv:1506.04967, 2015.

[13] H. Matuschek, R. Kliegl, S. Vasishth, H. R. Baayen, and D. M. Bates, "Balancing type 1 error and power in linear mixed models," *Journal of Memory and Language,* vol. 94, pp. 305-315, 2017.

[14] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics,* vol. 6, pp. 65-70, 1979.

[15] E. Grabe, B. Post, F. Nolan, and K. Farrar, "Pitch accent realization in four varieties of British English," *Journal of Phonetics,* vol. 28, pp. 161-185, 2000.

[16] P. Gilles, "Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung," de Gruyter, Berlin [u.a.], 2005.

[17] G.-A. Levow, "Context in multi-lingual tone and pitch accent recognition," in *Interspeech, 9th Annual Conference of the International Speech Communication Association,* Lisbon, Portugal, 2005, pp. 1809-1812.