Speech Prosody 2016
31 May - 3 June 2016, Boston, USA

Effects of information structure, syllable structure, and voicing on nuclear falling pitch accents in German

Frank Kügler, Susanne Genzel

Department of Linguistics, Potsdam University, Germany
{kuegler,sgenzel}@uni-potsdam.de

Abstract

Previous work showed that tonal alignment distinguishes information status (given/new), that syllable structure affects alignment and that the scaling and alignment of tones correlate with focus. We investigate the interaction of these effects on the tonal realization of nuclear falling accents in German. In a production experiment, syllable structure [± open syllable] and voicing of the following consonant [± voice] was systematically varied in four different information structure contexts: broad focus, narrow focus, contrastive focus and givenness. The results show that [± voice] of the following consonant had a significant effect: The F0 peak of the nuclear falling accents is systematically realized in the post-accented syllable if it was followed by a [+voice] consonant while it is realized towards the end of the accented syllable if it was followed by [-voice] consonant. The results further replicate well-established effects of information structure on the tonal alignment and scaling. The reported results will add to a growing body of studies on fine phonetic detail. In addition, results will also have consequences for annotation of German intonation.

Index Terms: German intonation, tonal alignment, tonal scaling, nuclear falling pitch accent, voicing, syllable structure

1. Introduction

Fine phonetic details distinguish tonal categories and/or express pragmatic meaning [1]. Peak alignment distinguishes information status (e.g. [2–6] for German), i.e. early vs. late peaks correlate with given vs. new (focus) information. Syllable structure has been shown to affect the alignment of tones in a number of languages: the presence of a [+voice] coda leads to a later alignment of the peak in H* accents [7–11]. Scaling and alignment of tones correlate with focus [12–14]. Higher scaling and later alignment corresponds to higher prominence. [15] proved this perceptually.

Although a lot is known about individual factors affecting F0 peak alignment of pitch accents, there is no study investigating the interaction of information structure effects and structural effects such as syllable structure or voicing of surrounding segments. For instance, analyzing a corpus of spoken German [8] show that peak alignment in nuclear falling accents is later if the accented syllable has a [+voice] coda (mean: 41.7% of syllable duration) than if it has a [-voice] coda (mean: 27.7% of syllable duration). They also found that there is virtually no difference between open and closed syllables, and that tenseness of vowels does not affect peak position. However, [8] particularly looked at F0 peaks within the accented syllable. Thus, any peak delay beyond the accented syllable was not covered.

In general, the "segmental anchoring hypothesis" put forward by [16, 17], and supported by [18] for tone languages, claims that both the start and the end of an F0 movement are precisely aligned with the segmental string, i.e., they define landmarks of tonal targets. Given that an F0 peak of a nuclear falling accent represents a landmark, we will analyze its phonetic properties and will establish its variation looking at an analysis window that comprises both the accented syllable and the surrounding syllables, i.e. the preceding and following syllable.

1.1. Research question and Hypotheses

The research question for the present study is to investigate whether there is systematic variation of F0 alignment and/or F0 scaling as a function of information structure, syllable structure and segmental context, i.e. voicing of consonants that follow the accented vowel.

The following hypotheses concerning nuclear falling accents in German are pursued in this study.

- H1: The effect of information status: given vs. new (focus) correlates with early and late peaks [2–6].
- H2: The effect of syllable structure: F0 peaks in closed syllables with a [+voice] coda occur later than in closed syllables with a [-voice] coda [8].
- H3: The effect of focus: Peaks are scaled higher and are aligned later in focus [12–14].

As for the interaction of information structure and structural effects it is largely unknown whether alignment patterns are affected to such an extent that, for instance, earlier peaks due to an obstruent coda overlap with earlier peaks due to given information. In order to investigate these effects we run a production study where speakers were asked to read target sentences as answers to pre-recorded questions.

2. Production study

2.1. Method

2.1.1. Speech materials

Four target items functioning as first names were used that meet the segmental criteria for syllable structure [±open] and segmental context of the following consonant [±voice] (cf. Table 1). The accented syllable of the target words is always centric. Two items contain an open accented syllable (/ri/ and /li/), and the following consonant is the onset of the post-accented syllable. The other two items contain a closed accented syllable (/rin/ and /lit/), and the following consonant is the coda consonant of the accented syllable. Note that according to German phonology, the vowel of the open syllable is necessarily a tense vowel, while the vowel in the closed

10.21437/SpeechProsody.2016-139

syllables is a lax one [19]. We assume that vowel quality does not affect F0 peak alignment [20]. The consonant following the accented vowel is either voiced (an alveolar nasal) or unvoiced (an alveolar stop).

Table 1. *Target items split by syllable structure of accented syllable* [±open] *and segmental context of following consonant* [±voice].

Syllable structure	[+voice]	[–voice]
[+open]	kaˈriːna	loˈliːta
[-open] (closed)	koˈrɪna	meˈlɪta

The target words were embedded in carrier sentences (1) such that only one sentence accent is produced, which is realized on the target word. The most neutral realization of a single accented sentence in German is with a nuclear falling pitch accent [21], indicated by H*L in (1). Two different carrier sentences, (1a) and (1b) were used. Each of the carrier sentences was put in four different contexts, see (2), to elicit the intended information structure on the target word in (1a.). The actual contexts contained more content than presented in (2). They were presented prior to the questions introducing the discourse situation and the discourse referents. Information structural terminology (focus and givenness) is based on [22].

(1) a. H*L
Es ist Karina dran. 'It is Karina's turn.'
b. H*L
Es hat Karina Schuld. 'It was Karina's fault.'

(2) a. Broad focus: Anything new?
b. Narrow focus: It's whose turn?
c. Contrastive focus: Is it Susanne's turn?
d. Givenness: I propose that it is Karina's turn.

2.1.2. Speakers

Eight female undergraduates in their twenties were recorded. All were native speakers of Standard German spoken in and around Berlin and reported no speech or hearing impairment. They either received course credit or were paid a small fee.

2.1.3. Recording procedure and data processing

The contexts spoken by a male voice had been previously recorded and were presented together with a target sentence both visually on screen and auditorily over headphones. The pre-recorded context sentences ensured that no uncontrolled variation of an experimenter speaking the context questions would affect the data elicitation. Speakers were asked to read and listen to the context and then to speak out the answer displayed on the screen as a response to the question. Subjects were familiarized with the task through written and verbal instructions. In case of hesitations or false starts, participants were asked to repeat the sentence. Recordings took place in a sound-proof chamber equipped with an AT4033a audiotechnica studio microphone, using a C-Media Wave sound card at a sampling rate of 44.1 kHz with 16 bit resolution. Presentation flow was controlled by the experimenter, and participants were allowed to take a break at any point. A total of 256 test sentences were recorded (2 σ-structure x 2 voicing contexts x 2 carrier sentences x 4 information structures x 8 speakers).

The 256 target sentences were hand-annotated using standard segmentation cues [23] and were subjected to phonetic analysis using Praat software [24]. The annotation comprised the levels of the syllable and of the segments. Each realization was rated as containing a falling nuclear pitch accent. The F0 peak was identified in two ways. In case of [+voice] target words (i.e., kaˈriːna, koˈrɪna), the F0 peak was located at the F0 maximum in or around the accented syllable. In case of a [-voice] target word (i.e., loˈliːta, meˈlɪta), the higher F0 peak was chosen comparing between the F0 maxima in and after the accented syllable. In case of a voiceless stop, the F0 maximum was measured excluding the first 30 ms of the vowel (cf. [11, 25]) which reflects the effect of microprosodic perturbation caused by the voiceless stop. This concerns [+voice] items when the F0 peak is before the accented syllable, and [-voice] items when the F0 peak is after the accented syllable. Note that 31.6 % of the nuclear falling accents were realized as plateau contours instead of single peak accents (cf. [26, 27] for British English). We follow [11] and analyze the final turning point of the flat plateau-like stretch as the F0 peak.

As measuring points, both the time of the F0 peak (t_{peak}) relative to the accented syllable's beginning (t_{beg}) and end (t_{end}) as well as the F0 maximum, were collected to tables. The relative alignment in per cent was calculated by equation (3). The peak position was set in relation to the beginning of the accented syllable.

$$Align_{rel}(\%) = \frac{(t_{peak} - t_{beg})}{(t_{end} - t_{beg})} * 100 \quad (3)$$

2.2. Results

2.2.1. Averaged pitch contours

To assess the overall F0 pattern on the target words, we plotted time-normalized mean F0 contours based on seven measuring points during the sonorant part of each of the three syllables of the target words. Figure 1–3 show the averaged contours realized in broad focus, narrow focus and for given information, respectively; the left panel shows a comparison between open and closed syllables of the [+voice] target words, the right panel shows the equivalent comparison of the [-voice] target words. 95 % confidence intervals accompany the mean contour.

Comparing the [+voice] and [-voice] realizations of the target words in broad focus (Figure 1) it can be seen that the F0 peak tends to be earlier in case of [-voice] target words (right panel) and later in case of [+voice] target words (left panel). In addition, in closed syllables the F0 peak tends to be earlier than in open syllables (dashed lines vs. solid lines).

Similar to broad focus, in narrow focus the F0 peak is generally realized later in [+voice] target words than in [-voice] target words (left vs. right panel in Figure 2). To be more precise, the F0 peak tends to be realized in the post-accented syllable in case of [+voice] target words while it is realized in the accented syllable in the [-voice] target words.

In case of the [-voice] target word with open syllables (solid line in right panel of Figures 1 and 2) a large microprosodic F0 boost is visible at the beginning of the post-accented (third) syllable. Note that the F0 maximum in that syllable is detected only *after* the first 30 ms (see section 2.1.3).

From Figure 3 it becomes obvious that given information tends to be realized with early F0 peaks, usually located in the syllable preceding the accented one. This is in line with previous findings (cf. [2–6]).

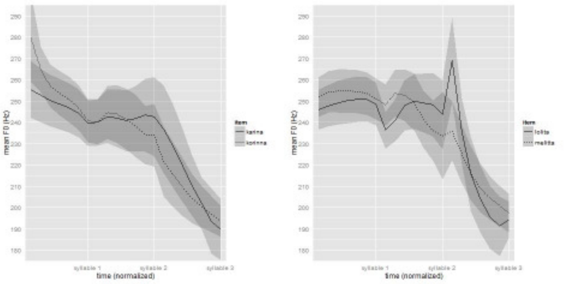

Figure 1: [+voice] *target words (left) and* [-voice] *target words (right) in broad focus.*

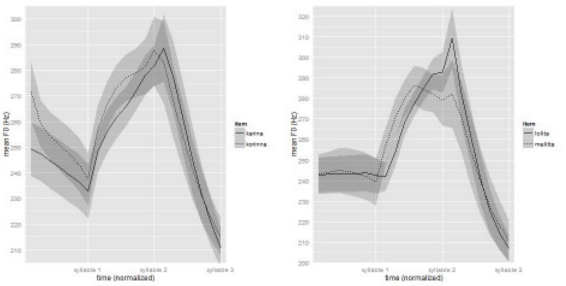

Figure 2: [+voice] *target words (left) and* [-voice] *target words (right) in narrow focus (narrow and contrastive focus collapsed).*

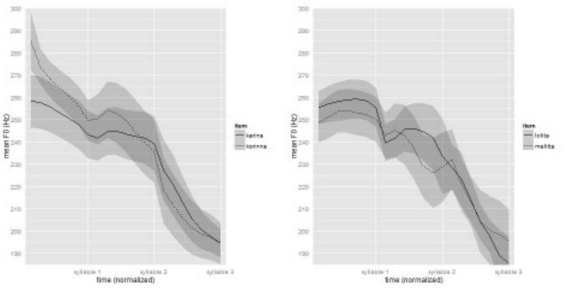

Figure 3: [+voice] *target words (left) and* [-voice] *target words (right) as given information.*

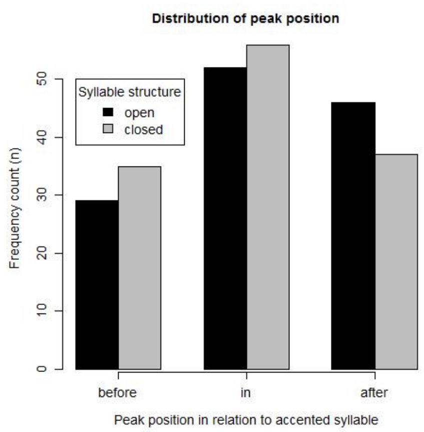

Figure 4: *Distribution of F0 peaks before, in or after the accented syllable split by* SYLLABLE STRUCTURE*. Black bars refer to open syllables, grey bars refer to closed syllables.*

Table 2. *Report of the linear mixed effects model with* SYLLABLE STRUCTURE *as fixed factor and position of F0 peak as dependent variable.*

Variable	Coefficients	SE	*t*-value
(Intercept)	0.81771	0.11614	7.041
σ-structure = closed	0.01517	0.14770	0.103

Table 3. *Report of the linear mixed effects model with* [±voice] *as fixed factor and position of F0 peak as dependent variable.*

Variable	Coefficients	SE	*t*-value
(Intercept)	1.0540	0.1138	9.259
voice = [-voice]	-0.4579	0.1021	-4.485

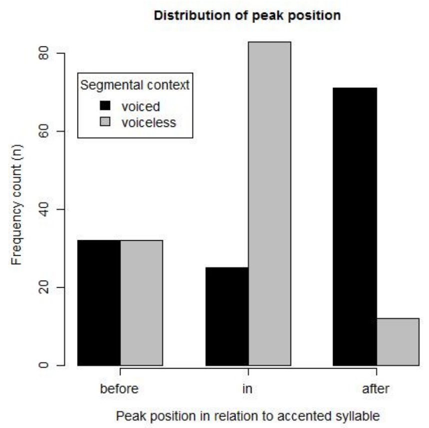

Figure 5: *Distribution of F0 peaks before, in or after the accented syllable split by* SEGMENTAL CONTEXT*. Black bars refer to* [+voice] *target words, grey bars refer to* [-voice] *target words.*

2.2.2. Syllable structure and F0 peak position

To assess the effect of SYLLABLE STRUCTURE on the alignment of the F0 peak we fit a multilevel model with SYLLABLE STRUCTURE (with levels open/closed) as fixed factor applying crossed random factors ‘speaker’ and ‘item’ [28], and alignment in terms of ‘position of F0 peak’ as dependent variable. Random slopes [29] for speakers were integrated into the models assuming that differences exist for each speaker’s individual realizations. Treatment-coding was applied using level ‘open’ as baseline. Table 2 reports the results of the model showing that the factor SYLLABLE STRUCTURE is not significant. This means that the distribution of the F0 peak does not depend on syllable structure in contrast to [7–11]. The distribution of the F0 peak illustrated in Figure 4 clearly shows that the majority of F0 peaks are realized in the accented syllable, but that an almost equal high number of F0 peaks are realized both before and after the accented syllable.

2.2.3. Segmental context and F0 peak position

The averaged time-normalized F0 contours indicated that the location of the F0 peaks tends to depend on the segmental context. To assess the effect of SEGMENTAL CONTEXT we fit an identical linear mixed effects model as for SYLLABLE STRUCTURE. The model used level [+voice] as baseline. Table 3 reports the model outcome sowing that the peak position is significantly different between [+voice] and [-voice] target words. As is shown in Figure 5, the F0 peak is mainly realized in the accented syllable in case of [-voice] target words while it is mainly realized in the post-accented syllable in case of [+voice] target words.

2.2.4. Information structure

To assess the effect of INFORMATION STRUCTURE on the alignment of F0 peaks we fit an identical linear mixed effects model as for SYLLABLE STRUCTURE. The model used level 'broad focus' as baseline. Table 4 gives the outcome of the model showing that F0 peaks in narrow and contrastive focus are realized significantly later than in broad focus while givenness has no effect compared to broad focus. Figure 6 displays the distribution of pitch peaks between the three syllables. In broad focus as in given constituents, F0 peaks are realized before or on the accented syllable, while in narrow / contrastive focus pitch peaks are realized either in or after the accented syllable in the majority of cases.

2.2.5. F0 scaling and alignment

We computed a Pearson product-moment correlation coefficient to assess the relationship between the F0 scaling and F0 alignment. Figure 7 displays the correlation between relative F0 peak alignment and F0 scaling. A negative alignment refers to an F0 peak in the syllable preceding the accented one, while an alignment greater than 100% refers to an F0 peak in the post-accented syllable. The correlation coefficients indicate that given referents show systematically more early peak realizations (cf. [2, 3]), which are scaled lower by trend, compared to broad focus where a positive correlation between scaling and relative alignment could be established (cf. Figure 7). A similar significant positive correlation could be established for narrow/contrastive focus.

Table 4. *Report of the linear mixed effects model with INFORMATION STRUCTURE as fixed factor and position of F0 peak as dependent variable.*

Variable	Coefficients	SE	*t*-value
(Intercept)	1.00912	0.20853	4.839
IS = narrow focus	-0.38756	0.16446	-2.357
IS = contr. focus	-0.39975	0.18162	-2.201
IS = given	0.04696	0.13973	0.336

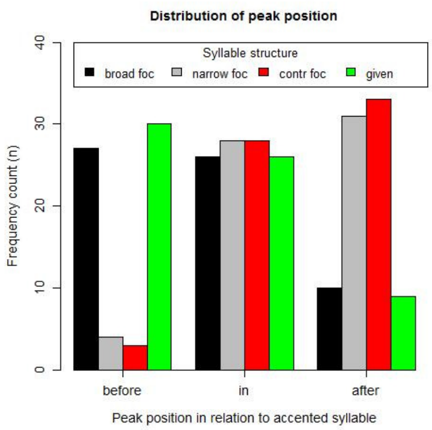

Figure 6: *Distribution of F0 peaks before, in or after the accented syllable split by INFORMATION STRUCTURE. Black bars refer to broad, grey bars to narrow, red bars to contrastive focus, and green bars to givenness.*

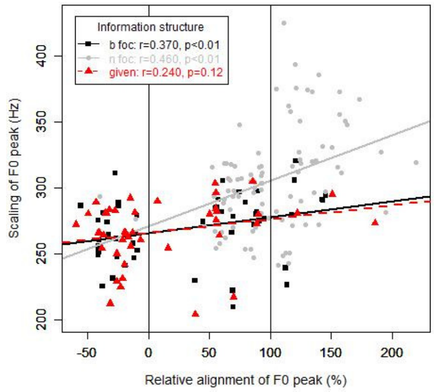

Figure 7: *Correlation between F0 peak alignment and F0 scaling split by INFORMATION STRUCTURE.*

2.3. Discussion

This study replicated some of the well-known effects of information structure on F0 peak alignment (cf. Figure 6; [12–14]). Narrow focus leads to a later alignment. It also revealed that given information is realized with earlier F0 peaks, usually before the accented syllable, than narrow focus but not earlier than broad focus. Peaks on given material are scaled lower. Presumably, the non-significant correlation between scaling and alignment (cf. Figure 7) reflects their status as less prominent than broad focus accents [30, 31].

Concerning structural effects, this study revealed that [±voice] of a following consonant has a significant effect on F0 peak alignment in German nuclear falling accents. At first sight and in line with [8], [+voice] codas show later F0 peaks alignment. However, our results show that considering [±voice] of the following consonant matters. In line with [7–11], F0 peaks were aligned significantly later on [+voice] than on [-voice] words. Unlike [8] however, results showed that, the F0 peak is realized in the post-accented syllable independent of syllable structure. Whether or not a following [+voice] consonant belongs to the accented syllable's coda does not matter for peak alignment. Similarly, independent of syllable structure, a following [-voice] consonant caused F0 peaks to be realized on the accented syllable (cf. Figure 5). Thus, studies on fine phonetic details need to consider a larger analysis window than only the accented syllable as in [8].

3. Conclusions

This study revealed a significant effect of segmental context on the alignment of F0 peaks in nuclear falling accents in German. It emphasizes that, contrary to previous research [8], an analysis window that comprises a pre- and a post-accented syllable is necessary to capture the relevant variation. At the same time, the results need to be considered in the annotation of German intonation: a following [+voice] consonant causes later alignment of F0 peaks, however the segmental context does not change the tonal category of a falling H*L accent.

4. Acknowledgements

This work was funded by the German Research Association [DFG grant number SFB 632 "Information structure", project D5], http://www.sfb632.uni-potsdam.de/en/cprojects/d5.html.

5. References

[1] B. Post, M. D'Imperio, and C. Gussenhoven, "Fine phonetic detail and intonational meaning," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 4-10 August 2007, 2007, pp. 191–196.

[2] K. J. Kohler, "Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics," in *Arbeitsberichte (AIPUK)*, vol. 25, *Studies in German Intonation*, K. J. Kohler, Ed, Kiel: Institut für Phonetik und digitale Sprachverarbeitung, 1991, pp. 115–185.

[3] S. Baumann, *The Intonation of Givenness. Evidence from German*. Tübingen: Niemeyer, 2006.

[4] O. Niebuhr and G. Ambrazaitis, "Alignment of Medial and Late Peaks in German Spontaneous Speech," in *Proceedings of Speech Prosody 2006, Dresden, Germany*, Dresden: TUDpress, 2006, pp. 161–164.

[5] O. Niebuhr, *Perzeption und kognitive Verarbeitung der Sprechmelodie: Theoretische Grundlagen und empirische Untersuchungen*. Berlin: Walter de Gruyter, 2007.

[6] M. Grice, S. Baumann, and N. Jagdfeld, "Tonal association and derived nuclear accents-The case of downstepping contours in German," *Lingua*, vol. 119, pp. 881–905, 2009.

[7] van Santen, Jan P. H. and J. Hirschberg, "Segmental effects on timing and height of pitch contours," in *Proceedings ICSLP-1994*, 1994, pp. 719–722.

[8] B. Möbius and M. Jilka, "Effects of Syllable Structure and Nuclear Pitch Accents on Peak Alignment: A Corpus-based Analysis," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 4-10 August 2007, 2007, pp. 1173–1176.

[9] M. D'Imperio, "The role of perception in defining tonal targets and their alignment," PhD Thesis, The Ohio State University, Ohio, 2000.

[10] P. Welby and H. Loevenbruck, "Anchored down in Anchorage: Syllable structure, rate, andsegmental anchoring in French," *Italian Journal of Linguistics*, vol. 18, no. 1, pp. 73–124, 2006.

[11] P. Prieto, "Tonal alignment patterns in Catalan nuclear falls," *Lingua*, vol. 119, no. 6, pp. 865–880, 2009.

[12] S. Baumann, M. Grice, and S. Steindamm, "Prosodic Marking of Focus Domains - Categorical or Gradient?," in *Proceedings of Speech Prosody 2006, Dresden, Germany*, Dresden: TUDpress, 2006, pp. 301–304.

[13] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics*, vol. 36, no. 4, pp. 680–703, 2008.

[14] C. Gussenhoven, *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press, 2004.

[15] D. R. Ladd and R. Morton, "The perception of intonational emphasis: continuous or categorical?," *Journal of Phonetics*, vol. 25, no. 3, pp. 313–342, 1997.

[16] D. R. Ladd, D. Faulkner, H. Faulkner, and A. Schepman, "Constant segmental anchoring of F0 movements under changes in speech rate," *Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1543–1554, 1999.

[17] D. R. Ladd, I. Mennen, and A. Schepman, "Phonological conditioning of peak alignment in rising pitch accents in Dutch," *Journal of the Acoustical Society of America*, vol. 107, no. 5, pp. 2685–2696, 2000.

[18] Y. Xu, "Consistency of Tone-Syllable Alignment across Different Syllable Structures and Speaking Rates," *Phonetica*, vol. 55, no. 4, pp. 179–203, 1998.

[19] R. Wiese, *The phonology of German*. Oxford: Oxford University Press, 1996.

[20] M. Jilka and B. Möbius, "The Influence of Vowel Quality Features on Peak Alignment," in *8th International Conference on Speech Communication and Technology*, Antwerp, Belgium: International Speech Communication Association, 2007, pp. 2621–2624.

[21] C. Féry, *German Intonational Patterns*. Tübingen: Niemeyer, 1993.

[22] M. Krifka, "Basic notions of information structure," *Acta Linguistica Hungarica*, vol. 55, no. 3, pp. 243–276, 2008.

[23] A. Turk, S. Nakai, and M. Sugahara, "Acoustic segment durations in prosodic research: A practical guide," in *Methods in empirical prosody research*, S. Sudhoff, D. Lenertová, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, and J. Schließer, Eds, Berlin: De Gruyter, 2006, pp. 1–27.

[24] P. Boersma and D. Weenink, *Praat: doing phonetics by computer [Computer program]*, 2014.

[25] S. Ritter and M. Grice, "The Role of Tonal Onglides in German Nuclear Pitch Accents," *Language and Speech*, vol. 58, no. 1, pp. 114–128, 2015.

[26] R.-A. Knight, "The realisation of intonational plateaux: effects of foot structure," in *Cambridge Occasional Papers in Linguistics 1*, Cambridge, 2004, pp. 157–164.

[27] 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, no. 3, pp. 223–244, 2008.

[28] D. Bates, M. Maechler, B. Bolker, and S. Walker, *lme4: Linear mixed-effects models using Eigen and S4*, 2013.

[29] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," *Journal of Memory and Language*, vol. 68, no. 3, pp. 255–278, 2013.

[30] S. Baumann and C. T. Röhr, "The perceptual prominence of pitch accent types in German," in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK: University of Glasgow, 2015, p. 0384: 1-5.

[31] S. Baumann, C. T. Röhr, and M. Grice, "Prosodische (De-) Kodierung des Informationsstatus im Deutschen" *Zeitschrift für Sprachwissenschaft*, vol. 34, no. 1, pp. 1–42, 2015.