10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Perceptual Prominence of Accent Types and the Role of Expectations

Christine T. Röhr[1], *Stefan Baumann*[1], *Petra. B. Schumacher*[2], *Martine Grice*[1]

[1]I*f*L Phonetik, University of Cologne, Germany
[2]IdSL I, University of Cologne, Germany
{christine.roehr; stefan.baumann; petra.schumacher; martine.grice}@uni-koeln.de

Abstract

This paper is concerned with how accent types can contribute to the perceived prominence of the accented words, both when they are presented in a sentence devoid of context and when a context is used to build expectations of more or less prosodic prominence. In a prominence rating experiment in German, target words were presented with four different levels of prosodic prominence (L+H*, H*, H+L*, deaccented) in isolated sentences. In a second prominence rating experiment, words with two of these levels of prominence (L+H* and H+L*) were presented with a prior context evoking the expectation that the upcoming information was either exciting or neutral. In a third experiment, this reduced set of stimuli was also assessed for appropriateness, i.e. as to whether the target stimulus matched expectations evoked by the context provided. Results indicate that, both with and without prior context, pronounced accentual rises (L+H*) involve a particularly high level of perceptual prominence. Contextually induced expectations had no effect on the perception of prominence. However, the extra prominence for L+H* was only appropriate in the exciting context. Falling accents (H+L*) were lower in prominence, and were appropriate in both contexts, indicating that they did not need to be contextually licensed.

Index Terms: prominence, prosody, accent types, speech perception, appropriateness, expectations, German

1. Introduction

There is ample evidence for West-Germanic languages that information status and focus domains of different sizes and types are marked by the level of prominence expressed by different pitch accent types. That is, new or less accessible information and words in narrow and contrastive focus have been found to involve greater prosodic prominence (see e.g. for information status: [1, 2, 3, 4]; for focus: [5, 6, 7]). Accordingly, listeners have been found to be able to interpret an element's level of prominence as indicative of the information structure that the speaker is intending to convey (e.g. [7, 8, 9]).

For German, a direct relation has been attested between the impression of an element standing out (i.e. perceptual prominence) and different accent types, which can be attributed to three tonal dimensions in the acoustic signal [10] (see also [11]). Most important for German listeners is the direction of pitch movement in the vicinity of the accented syllable, with rises being more prominent than falls (see also [12]). The extent of the pitch excursion/onglide (steep - shallow) and the height of the f0 target corresponding to the starred tone (high - mid - low) are also relevant, but are perceptually less decisive.

However, there is also evidence that the perception of prominence is not only determined by the auditory input but may also be affected by intrinsic properties of words, language-specific expectations, and interpretations of pragmatic and phonological context (e.g. [12, 13]). For instance, [13] showed that expectations about the focus structure may even overwrite the prominence information provided by the acoustic signal.

The current paper investigates German accent types with the aim of providing further evidence for their perceptual prominence (on more naturally produced data) and to further investigate the effect of contextually induced expectations that are not related to information structural aspects. Two perception experiments assessed untrained listeners' intuitions about the degree of prominence of words in isolated and in contextualized sentences (prominence rating task), so as to investigate both the prosodic and contextual influences on prominence perception. In a third study, we additionally evaluated the appropriateness of the sentences with respect to the context (appropriateness rating task). Three pitch accent types (L+H*, H*, H+L*) plus deaccentuation were tested in the absence of context, while two pitch accent types (L+H* and H+L*) were tested in relation to prior contexts that trigger the expectation that the following test sentence either contains information about an exciting and unusual event (EXCITING) or information about an ordinary, predictable event (NEUTRAL). Two previous EEG studies on the same stimuli [14] revealed that a high level of prosodic prominence (L+H*) – regardless of context – consumes attentional resources and engenders processing costs. Further, inappropriate accents engendered prediction error responses. A production study on the contextualized sentences [15] confirmed that exciting information was made more prominent by a predominant use of rising accents (L+H*, H*) with large tonal onglides, compared to neutral information which was made less prominent by using fewer and smaller rising onglides as well as more and larger falling onglides (H+!H*, H+L*). Based on these previous findings, we formulated the following hypotheses for the current perception experiments.

Hypothesis 1a (prominence - isolated sentences):
If no context is provided and only the acoustic signal plays a role, we expect (according to [10]) an increase in perceptual prominence of the target items from deaccentuation (Ø) through falling (H+L*) and shallow rising (H*) to steep rising (L+H*) nuclear pitch accents.

Hypothesis 1b (prominence - contextualized sentences):
If a context is present, we expect that a context expressing excitement leads to an increase in perceptual prominence for both rising (L+H*) and falling (H+L*) accents.

Hypothesis 2 (appropriateness - contextualized sentences):
Test sentences with a prominent accent (L+H*), typically found on new information, should match the EXCITING but not the NEUTRAL context. Conversely, test sentences with a less prominent accent (H+L*), typically found on contextually derivable information, should match the NEUTRAL context and be permissible after the EXCITING context.

10.21437/SpeechProsody.2020-75

2. Methods

2.1. Stimuli

The following stimuli also served as experimental stimuli in two EEG studies [14] and have been used as reading material in a production study [15] (see above). All stimuli were read by a trained 48-year-old male phonetician (native German speaker) and were recorded in a sound-attenuated booth with a sampling rate of 44100 Hz and 16 bit resolution (mono).

In the first set of stimuli, four intonation contours were tested in isolated (i.e. context-free) sentences as exemplified in (1) with a sentence-initial pronominal subject *wir* "we", an auxiliary verb *haben* "have", a direct object (e.g. *Milena*) and a sentence-final past participle (e.g. *getroffen* "met"). (Capital letters in (1) indicate accented syllables.) The last argument of the test sentences, i.e. the direct object, served as target word involving different levels of prosodic prominence. The target word was realized with three different nuclear accent types (1a), categorized according to GToBI [16], or without an accent (1b). In particular, we tested a steep rising, late-peak accent (L+H*), a shallow rising accent (H*) and a falling, early-peak accent (H+L*) as well as deaccentuation of the target word (Ø) accompanied by a steep rising (L+H*) nuclear accent on the sentence-final participle. All test sentences ended on a final low boundary tone.

(1) (a) *Wir* *haben* ***MiLEna*** *getroffen.*
L+H* / H* / H+L*

(b) *Wir* *haben* ***Milena*** *geTROFfen.*
Ø L+H*

pronoun auxiliary noun participle

"We met Milena." (lit. "We have Milena met.")

Sixty test sentences with different lexical material (direct object and participle) were constructed and realized with each of the four intonation contours (240 test sentences). As target words we used 37 proper nouns (person names like *Milena*) and 23 common nouns in their plural form without a definite article (such as *Bananen* "bananas"). All target words are trisyllabic with primary stress on the second syllable, have a simple segmental structure and are predominantly voiced to enable a continuous f0 trajectory. Furthermore, all sentence-final participles start with the prefix *ge-* and are at least three syllables long so as to avoid truncation or compression of the f0 contour on the preceding nuclear accented target word. Further, 80 filler sentences were created with an identical syntactic structure but varied in the position of the nuclear accent, so that each of the four positions of the sentence (pronoun, auxiliary, noun and participle) equally often carried a shallow rising (H*) nuclear accent.

We controlled both the perceptual and acoustic equivalence of the respective accent types in the test sentences. Figure 1 shows the mean f0 contours as well as the individual f0 traces of all test sentences per condition superimposed. These f0 traces indicate that the intonation was produced in a very stable manner in each of the four conditions. For some stimuli, the f0 values of the accentual peaks or valleys were adjusted (with *Praat*: [17]) if they exceeded a designated range of 10-15 Hz around the mean value for a specific condition (number of resynthesized stimuli per condition: L+H* = 23, H* = 19, H+L* = 25 and deaccentuation = 19).

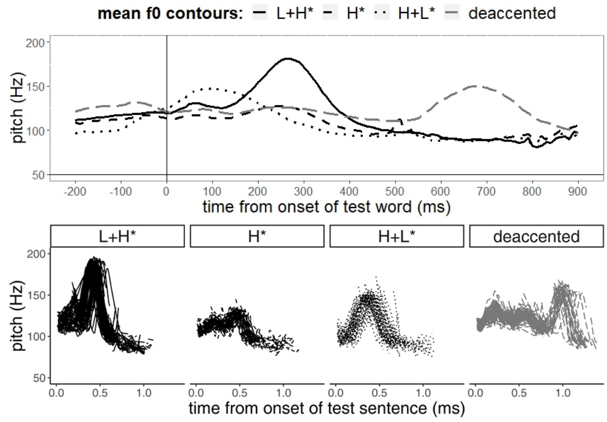

Figure 1: *Top panel: Mean f0 contours of all test sentences per condition temporally aligned with the onset of the target word (at vertical bar); bottom panel: Individual f0 traces of all test sentences per condition superimposed, from the onset until the end of the test sentence.*

Table 1 shows the means and standard deviations of the accentual peaks (for deaccentuation: the f0 value in the middle of the stressed vowel), the tonal onglide (i.e. the pitch excursion towards the starred tone of the accented syllable in semitones [18]; see also [7]) and the duration of the target words for each condition. The values show stepwise differences in the height of the accentual H tone with the highest value for L+H* accents, a mid value for H+L* accents and the lowest value for H* accents (which is similar to the f0 value of deaccented target syllables). As intended, target words with a steep rising (L+H*) accent and a falling (H+L*) accent also exhibit larger tonal onglides than targets with a shallow rising (H*) accent.

Table 1: *Means and standard deviations (in brackets) of the accentual peak [Hz], the accentual onglide [semitones] and the duration [ms] of the target words.*

Condition:	Pitch peak [Hz]	Tonal onglide [st]	Duration [ms]
L+H*	188.1 (4.0)	6.6 (1.6)	415 (39)
H*	130.8 (4.2)	2.1 (0.8)	406 (41)
H+L*	151.0 (4.6)	-7.9 (1.0)	430 (43)
Ø	126.2 (6.6)	--	425 (44)

Altogether, the first set of stimuli contains 320 items, involving 240 test/critical items (60 lexically different sentences x 4 intonation contours) plus 80 filler items (20 lexically different filler sentences x 4 accent positions).

In the second set of stimuli, two intonation contours of the stimuli introduced above were tested in relation to two pre-contexts (two-by-two design): The test sentences with a steep rising (L+H*) accent and those with a falling (H+L*) accent on the target word were crossed with two types of pre-contexts that generate different expectations about appropriate prosody in the test sentences, as exemplified in (2): Context (a) builds up an expectation for "new", exciting information, guiding the attention of the listener to an event that is expected to be highlighted by prosodic prominence. By contrast, context (b) establishes that nothing "new" or unusual will follow and thus builds up an expectation for a more neutral prosodic realization, i.e. less prosodic prominence.

(2) (a) *Rate mal, was uns heute passiert ist!* EXCITING
*Wir haben **MiLEna** getroffen.* L+H*/H+L*
"Guess what happened to us today! We met Milena."

(b) *Heute ist nichts Besonderes passiert.* NEUTRAL
*Wir haben **MiLEna** getroffen.* L+H*/H+L*
"Today, nothing special happened. We met Milena."

Sixty EXCITING and NEUTRAL context sentences were created along the lines of (2a) and (2b) with different lexical material (120 context sentences) and were produced with as neutral prosody as possible. We endeavoured to keep the contexts semantically unspecific and to control them with respect to their effect on the information structure of the following sentences: In all test sentences the target word and the sentence-final participle are part of a broad focus domain and are not derivable from the context. Furthermore, the 80 filler sentences (see above) were also provided with EXCITING and NEUTRAL contexts so that both context types were equally distributed over the four different filler conditions (H* accent on each word of the sentence).

No adjustments were made to the context stimuli. For the experimental session, context and target sentences were merged into one stimulus separated by a pause of 400 ms. Altogether, the second set of stimuli also contains 320 items, involving 240 test/critical items (60 lexically different context sentences x 2 context types crossed with 60 lexically different test sentences x 2 intonation contours) plus 80 filler items (40 lexically different context sentences x 2 context types evenly combined with 20 lexically different filler sentences x 4 accent positions).

2.2. Procedures

The three studies were conducted through web-based questionnaires implemented with the *SoSci Survey* software available at www.soscisurvey.de [19]. In all studies critical and filler items were evenly distributed and pseudo-randomized across four different lists: In order to prevent repetition effects, test sentences with the same lexical material were assigned to different lists. This means that each experimental list contained 60 lexically different test sentences (15 sentences per experimental condition) and 20 filler sentences (5 sentences per condition). Furthermore, in order to avoid systematic order effects in the exposure to the stimuli, experimental stimuli were presented in different condition sequences across the lists. Each participant saw only one of the four experimental lists. Accordingly, each participant rated 80 stimuli (60 critical and 20 filler items) in the main part of an experiment. At the beginning of a study, participants were familiarized with the experimental procedure by means of a short practice part including six stimuli.

In the two prominence rating studies the participants' task was to evaluate how highlighted/prominent each word in a target sentence sounded. The participants were told to give their judgments for each word by placing a roll bar on a continuous horizontal line without scaling (visual analogue scale) with the left pole labelled "not at all highlighted" and the right pole labelled "strongly highlighted" (see Fig. 2). The responses were encoded as interval data ranging from 1 (left pole) to 100 (right pole). Hence, higher ratings reflect a higher degree of perceptual prominence. The evaluation was carried out for each stimulus on a separate page. Both the context sentence (if applicable) and the target sentence were presented acoustically and orthographically. Participants were able to control when and how often to play a stimulus.

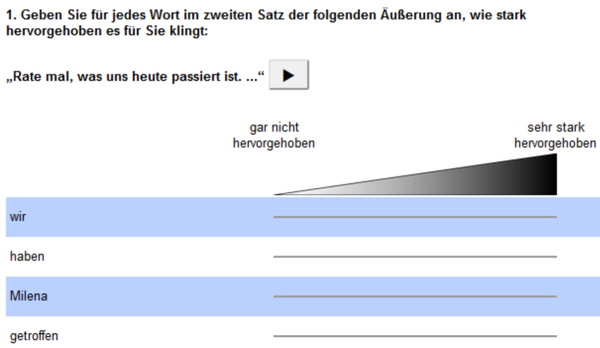

Figure 2: *Example from the prominence rating task with contextualized test sentences.*

In the appropriateness rating study, the participants' task was to evaluate how well the melody of a target sentence matched the preceding context. Judgments were given by placing a roll bar on a visual analogue scale with the left pole labelled "does not match at all" (= 1) and the right pole labelled "matches very well" (= 100) (see Fig. 3). Higher ratings reflect a higher degree of appropriateness. The evaluation was carried out for each stimulus on a separate page. While both the context and the target sentence were presented acoustically, only the context sentence was presented orthographically. Participants were able to control when and how often to play a stimulus.

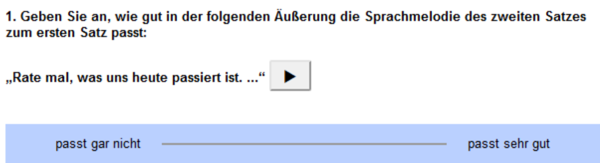

Figure 3: *Example from the appropriateness rating task with contextualized test sentences.*

2.3. Participants and analysis

The three perception experiments were conducted separately with different groups of native German speakers. None of them reported any auditory or visual impairment.

- <u>Prominence ratings of isolated sentences:</u> 22 participants (7 female, 15 male), aged between 19 and 49 years (mean = 26.2, SD = 8.4)
- <u>Prominence ratings of contextualized sentences:</u> 23 participants (12 female, 11 male), aged between 18 and 40 years (mean = 25.2, SD = 5.7)
- <u>Appropriateness ratings of contextualized sentences:</u> 23 participants (16 female, 7 male), aged between 19 and 43 years (mean = 24.2, SD = 5.7)

To test for statistical significance of the ratings on the visual analogue scales, we calculated linear mixed-effect models using the *lmer()* function from the *lme4* package for *R* [20]. The models for the isolated sentences (prominence ratings) included *PROSODY* (L+H*, H*, H+L*, Ø) and *WORD* (pronoun, auxiliary, noun, participle) as fixed factors as well as by-subject intercepts and slopes for *PROSODY* as random factors. The models for the contextualized sentences (prominence and appropriateness ratings) included *CONTEXT* (EXCITING, NEUTRAL), *PROSODY* (L+H*, H+L*) and *WORD* (if applicable) as fixed factors as well as by-subject intercepts and slopes for *PROSODY* and *CONTEXT* as random factors.

3. Results and discussion

3.1. Prominence ratings

Likeliood ratio tests on the prominence ratings of the isolated sentences revealed a significant effect of PROSODY ($\chi^2 = 37.84$, $p < .001$) and WORD ($\chi^2 = 3125.1$, $p < .001$) as well as an interaction between both factors ($\chi^2 = 1992.8$, $p < .001$). Figure 4 shows that accented words were generally perceived as more prominent than unaccented words.

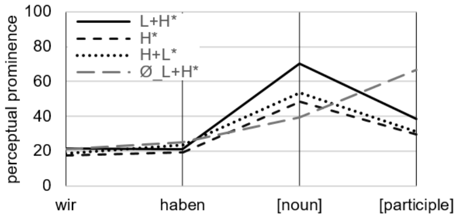

Figure 4: *Mean prominence ratings of each word in isolated sentences with different intonation.*

As to different accent types, L+H* accents were clearly rated as most prominent, with similar values for L+H* accents on the noun (M = 70, SD = 19) and on the participle in the Ø condition (M = 67, SD = 22). H* (M = 48, SD = 22) and H+L* (M = 53, SD = 21) accents received medial prominence values. Hence, Hypothesis 1a is mostly confirmed. As to the unaccented words, pronouns (M = 20, SD = 17) and auxiliaries (M = 22, SD = 17) were rated as least prominent. However, deaccented nouns (Ø: M = 39, SD = 20) as well as unaccented participles (L+H*: M = 39, SD = 20; H*: M = 30, SD = 17; H+L*: M = 31, SD = 20) reveal somewhat higher prominence scores, which might be related to a higher probability for being accented in the given sentence structure. Nevertheless, these relatively high values are surprising since we would rather expect a clearer perceptual contrast between adjacent accented and unaccented words. Instead, it may be that the prominence of one word bleeds into adjacent ones, possibly owing to the fact that cues are rarely confined to the accented word, especially in late peak accents, where they can extend onto following syllables.

As to the contextualized sentences, a likelihood ratio test showed that CONTEXT had no effect on the prominence ratings ($\chi^2 = .07$, $p = .79$). However, further model comparisons revealed a significant effect of PROSODY ($\chi^2 = 22.68$, $p < .001$) and WORD ($\chi^2 = 4954.9$, $p < .001$) as well as an interaction between these two factors ($\chi^2 = 400.03$, $p < .001$) and between CONTEXT, PROSODY and WORD ($\chi^2 = 408.65$, $p < .001$). Figure 5 shows that the prominence ratings of the sentences with an L+H* (M = 71, SD = 19) and H+L* (M = 51, SD = 24) nuclear accent on the target word were replicated when presented in context. However, Hypothesis 1b was not confirmed, since contextually induced expectations turned out to have no effect on the perception of prominence.

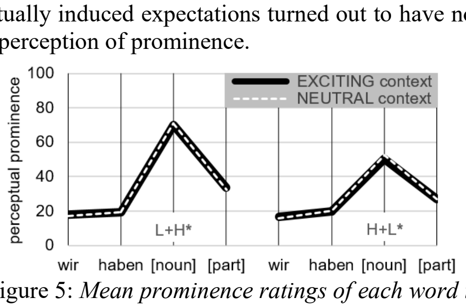

Figure 5: *Mean prominence ratings of each word in contextualized sentences with an L+H* or H+L* nuclear accent.*

In general, both prominence rating studies suggest that in German a pronounced rising pitch on a stressed syllable (as in L+H* accents) is a decisive tonal cue to prominence that leads to enhanced perceptual prominence (in line with [10, 12, 14]).

3.2. Appropriateness ratings

Likeliood ratio tests on the appropriateness ratings of contextualized sentences revealed no effect of CONTEXT ($\chi^2 = 2.39$, $p = .12$), but a significant effect of PROSODY ($\chi^2 = 6.72$, $p < .01$) as well as an interaction between CONTEXT and PROSODY ($\chi2 = 33.36$, $p < .001$). Figure 6 shows that Hypothesis 2 is confirmed:

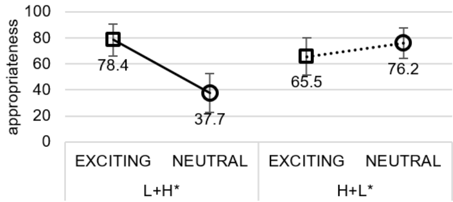

Figure 6: *Mean appropriateness ratings and standard deviations of contextualized sentences with an L+H* or H+L* nuclear accent.*

Prominent L+H* accents were perceived as appropriate in the EXCITING context but not in the NEUTRAL one. Less prominent H+L* accents obtained the highest appropriateness scores in the NEUTRAL context, but were also perceived as appropriate in the EXCITING context. This suggests that falling (H+L*) accents - similar to H* accents and due to their lower perceptual prominence - may serve as prosodic default realization which is in line with the observation that H+L* accents commonly occur in reading style, especially in isolated sentences (all new, broad focus) or paragraph-finally [21].

4. Summary and conclusions

The present paper assessed the perceptual prominence of different prosodic realizations of stimuli without and in context, and their appropriateness with respect to expectations as to how exciting the content of a following utterance will be.

Results provide further evidence for the relevance of the tonal onglide in prominence perception. Accents with a pronounced rising f0 (as in L+H*) involve an extra high level of perceptual prominence, whereas accents with a falling f0 (as in H+L*) were less conspicuous, involving a medium level of prominence. The different levels of perceptual prominence appear to license the use of a particular accent type in different contexts: While L+H* accents were only appropriate as a marker of EXCITING information, H+L* accents were equally appropriate as a marker of NEUTRAL and EXCITING information.

Furthermore, listeners' prominence ratings appear to rely primarily on acoustic cues: Contrary to previous findings for English (e.g. [13]), listeners in our study did not take contextual information into account when rating perceptual prominence.

5. Acknowledgements

This research has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 281511265 – SFB 1252 "Prominence in Language" (in project A01) at the University of Cologne. We would like to thank Christian Weitz and Matthias Gauser for assistance with the experimental setup.

6. References

[1] J. B. Pierrehumbert and J. Hirschberg, "The Meaning of Intonational Contours in the Interpretation of Discourse," in P. R. Cohen, J. Morgan and M. E. Pollak (eds.), *Intentions in Communication*, 1990, pp. 271–311. Cambridge: MIT Press.

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

[3] A. Chen, E. Den Os and J. P. De Ruiter, "Pitch Accent Type Matters for Online Processing of Information Status: Evidence from Natural and Synthetic Speech," *The Linguistic Review*, vol. 24, no. 2–3, pp. 317–344, 2007.

[4] C. T. Röhr and S. Baumann, "Prosodic Marking of Information Status in German," *Proceedings of the 5th International Conference on Speech Prosody*, vol. 100019, pp. 1–4, 2010.

[5] K. Ito, S. Speer, and M. Beckman, "Informational Status and Pitch Accent Distribution in Spontaneous Dialogues in English," *Proceedings of the 2nd International Conference on Speech Prosody*, pp. 279-282, 2004.

[6] S. Baumann, J. Becker, M. Grice and D. Mücke, "Tonal and Articulatory Marking of Focus in German," in *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS)*, pp. 1029–1032, 2007.

[7] M. Grice, S. Ritter, H. Niemann and T. B. Roettger, "Integrating the Discreteness and Continuity of Intonational Categories," *Journal of Phonetics*, vol. 64, pp. 90–107, 2017.

[8] M. Breen, E. Fedorenko, M. Wagner and E. Gibson, "Acoustic Correlates of Information Structure," *Language and Cognitive Processes*, vol. 25, pp. 1044–1098, 2010.

[9] C. T. Röhr and S. Baumann, "Decoding Information Status by Type and Position of Accent in German," *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)*, pp. 1706–1709, 2011.

[10] S. Baumann and C. T. Röhr, "The Perceptual Prominence of Pitch Accent Types in German," *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, vol. 298, pp. 1–5, 2015.

[11] M. Grice and S. Baumann, "An Introduction to Intonation - Functions and Models," in J. Trouvain and U. Gut (eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice*, 2007, pp. 25–51. Berlin & New York: De Gruyter.

[12] S. Baumann and B. Winter, "What Makes a Word Prominent? Predicting Untrained German Listeners' Perceptual Judgments," *Journal of Phonetics*, vol. 70, pp. 20–38, 2018.

[13] J. Bishop, "Information Structural Expectations in the Perception of Prosodic Prominence," in G. Elordieta and P. Prieto (eds.), *Prosody and Meaning*, 2012, pp. 239–270. Berlin & New York: Mouton De Gruyter.

[14] C. T. Röhr, I. Brilmayer, S. Baumann, M. Grice and P. B. Schumacher, "Signal-Driven and Expectation-Driven Processing of Accent Types," Language, Cognition & Neuroscience, submitted.

[15] C. T. Röhr, H. Niemann, S. Baumann and M. Grice, "The Influence of Expectations on Tonal Cues to Prominence," *Journal of Phonetics*, submitted.

[16] M. Grice, S. Baumann and R. Benzmüller, "German Intonation in Autosegmental-Metrical Phonology," in S.-A. Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 2005, pp. 55–83. Oxford: Oxford University Press.

[17] P. Boersma and D. Weenink, Praat: Doing Phonetics by Computer (Version 5.3.80) [Computer Program], 2013, retrieved from http://www.praat.org/

[18] 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.

[19] D. J. Leiner, SoSci Survey (Version 3.1.06) [Computer software], 2019, available at https://www.soscisurvey.de

[20] D. Bates, M. Mächler, B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, vol. 67, no. 1-48, 2015.

[21] M. Grice and S. Baumann, "Intonation in der Lautsprache: Tonale Analyse," in U. Domahs and B. Primus (eds.), *Handbuch Laut, Gebärde, Buchstabe*, 2016, pp. 84–105. De Gruyter.