

The relative cueing power of F0 and duration in German prominence perception

Oliver Niebuhr¹, Jana Winkler²

¹Mads Clausen Institute, IRCA, University of Southern Denmark, Sønderborg, Denmark

²Dept. of General Linguistics, Kiel University, Kiel, Germany

oniebuhr@mci.sdu.dk, admin@isfas.uni-kiel.de

Abstract

Previous studies showed for German and other (West) Germanic language, including English, that perceived syllable prominence is primarily controlled by changes in duration and F0, with the latter cue being more powerful than the former. Our study is an initial approach to develop this prominence hierarchy further by putting numbers on the interplay of duration and F0. German listeners indirectly judged through lexical identification the relative prominence levels of two neighboring syllables. Results show that an increase in F0 of between 0.49 and 0.76 st is required to outweigh the prominence effect of a 30% increase in duration of a neighboring syllable. These numbers are fairly stable across a large range of absolute F0 and duration levels and hence useful in speech technology.

Index Terms: prominence, perception, F0, duration, German.

1. Introduction

Prominence in the sense of the degree to which a syllable stands out in perception is the pivot of many important functions in speech communication. Prominences indicate that the corresponding piece of information is new or shall be discussed from a new angle. Prominent syllables link lexical meanings with speaker attitudes, set boundary signals in syntagmatic structure, provide the basis for emphatic expressions, and they are of course the building blocks of speech rhythm whose look-ahead function guides the listener along the speaker's line of argument and "glues" (e.g., interrupted) phrases together. Not least because of this pivotal role, syllable prominence is probably one of the longest and best investigated features of prosody.

In many languages, speakers use prosody to make syllables stand out against neighboring syllables. Duration and F0 are typically considered the two major cues to prominence (but, see [1]). Both are positively correlated with prominence. That is, longer durations (e.g., of vowels) increase perceived prominence; and the same applies to larger pitch movements, especially when these movements are rises, cf. [2,3,4,5,6].

Although the nature of prominence is already pretty well understood and was investigated from very different angles [7], a few questions still remain open. In view of the growing attention paid to prominence matters in the rapidly growing fields of speech technology and human-machine-interaction [8,9], a very important question concerns the *quantification* of prominence-cue hierarchies. For example, for German, which is the subject of the present study, [10] concluded that F0 is the primary cue to prominence, as it varied more between prominent and non-prominence syllables than duration. The data of a further production study by [11] lent support to the conclusion of [10] for fully vs. partially prominent syllables. Moreover, the production evidence is consistent with percep-

tion evidence on German ([6,12,13]). But, the fact that F0 is obviously higher in the German prominence-cue hierarchy than duration does not tell us *how much higher* it actually is. That is, how powerful is F0 compared to duration? How much of a change in duration is necessary to counterbalance a given change in F0? Is this amount of change the same for pitch accents in upstep and downstep contexts? Our study addresses these questions and provides initial cue power estimates for German that can, for example, be used in speech-synthesis applications or quantitative models of prominence [14].

Previous perception experiments that were focused on establishing the prominence hierarchies cannot answer quantitative questions on relative cueing power or allow deriving cue power estimates from their results. The reason for this is that F0 and duration were varied and judged in different stimulus series, see the left panel of Figure 1. Due to this setup, the two prominence cues could never directly compete with each other. Accordingly, the conclusion of, for example, [6] that F0 is a more powerful cue to syllable prominence in German than duration does not rely on a quantitative comparison. Rather, it relies on the qualitative fact that a successively increasing syllable duration was able to outweigh a weaker F0 prominence cue in stimulus series 1, but not a stronger F0 prominence cue in stimulus series 2. In order to put a number on the extent to which F0 outweighs duration in terms of perceived prominence, an experimental setup is needed in which F0 and duration continua are orthogonally combined within the *same* stimulus series. Moreover, the two parameters have to be varied in ranges that allow multiple alternations of the stronger prominence parameter and prevent one parameter to completely override the other. Such a setup is illustrated in the right panel of Figure 1. It was used in the present study.

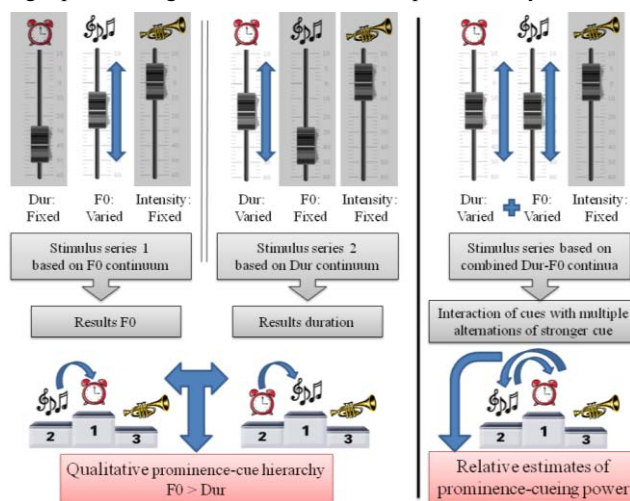


Figure 1: The present quantification-oriented experimental setup (right) in comparison to hierarchy-oriented setups (left).

There is one major exception in which a direct competition of prominence cues was implemented like in the right panel of Figure 1. This exception is the seminal experiment series by Fry [2,15]. However, even though Fry explicitly addressed the issue of the "relative strength" of prominence cues (2:135), he was only able to draw conclusions on the competition between duration and intensity, but not on the key competition between F0 and duration. For the latter competition, Fry's F0 variation only created all-or-none effects on syllable prominence instead of showing a gradual, quantifiable interaction with the variation in duration.

One reason for this outcome could be that Fry varied F0 by changing the relative frequency scaling between two flat F0 stretches (i.e. F0 "plateaux" with a roughly steady pitch) on adjacent syllables. Such F0 plateaux are fairly rare in natural speech and thus could have stimulated a psychoacoustic listening mode. A psychoacoustic listening mode would foster categorical judgments based on, for example, musical interval steps or the change from downward to upward pitch intervals for which listeners are particularly sensitive [16]. In addition, we think that the all-or-none effects in Fry's F0 continua could have been further fostered by the fact that he used low-quality pattern-playback syntheses of isolated disyllables. In contrast, our study is based on PSOLA resyntheses of entire stimulus sentences, and our F0 manipulation was conducted in the form peak-range extension rather than relative plateau scaling.

So, overall, our stimuli are a lot more natural than those of Fry; and, indeed, our stimuli yielded more insightful results with respect to the relative cueing power of F0 and duration, based on a judgment task that was inspired by that of Fry.

2. Method

2.1. Participants

The participants of the perception experiment were 61 native speakers of Standard German, 42 females and 19 males. They were all between 21 and 50 years old (average age 22.8 years) and undergraduate students of Empirical Linguistics at Kiel University. None of the students had taken part in a prosody lecture or prosody experiment by the time of our study. They received course credits for their participation.

2.2. Target items

The stimuli's key elements were three target items. Table 1 shows these items together with their respective phonological-phonetic representation and English translation. Each item has a one-word interpretation (1-3a) and a two-word interpretation (1-3b). The one-word interpretations are compound nouns, the two-word interpretations are noun sequences or a sequence of adverb and noun ("mehr Salz", 3b). All target-item interpretations are common expressions in Standard German.

Table 1: 3x2 target items of the perception experiment and their one-word (left, 1-3a) and two-word (right, 1-3b) interpretations. Stressed syllables are underlined.

One-word items: noun compounds	English	Two-word items: {N/Adv}+N	English
(1a) <u>D</u> osengemü <u>s</u> e	canned vegetables	(1b) <u>D</u> osen <u>G</u> emü <u>s</u> e	cans of vegetables
(2a) <u>M</u> utterer <u>u</u> nde	topsoil	(2b) <u>M</u> utter <u>E</u> rde	mother earth
(3a) <u>M</u> eersal <u>z</u>	sea salt	(3b) <u>m</u> ehr <u>S</u> al <u>z</u>	more salt

The target items include two stressed syllables. The decisive point is: Whether listeners arrive at the one-word or the two-word interpretation of each target item is solely based on prosody, more specifically, on the relative prominence difference between the two stressed syllables. Stronger prominence on the first stressed syllable triggers one-word interpretations. Stronger prominence on the second syllable triggers two-word interpretations of the target items.

This semantics-driven approach, which was inspired by the experiments of Fry [2,15], allowed us to obtain relative prominence judgments from naïve listeners without giving them complex metalinguistic instructions like 'please judge whether syllable X or Y sounds more salient to you'. Listeners are undoubtedly able to produce valid results based on such instructions, even on very fine-grained prominence scales [17, 18,19,20,21]. Yet, in view of the large number of stimuli that our research question required, we preferred a judgment task that is closer to everyday speech perception and as easy as possible to perform for listeners. They simply had to name the words they perceived in the stimuli, which potentially also yields more generalizable results.

The three target items differed in the number of unstressed syllables in between the two stressed ones. Target item (1a-b) had two unstressed syllables in between the stressed ones. In target item (2a-b), the stressed syllables framed just one unstressed syllable; and in item (3a-b) there was no intermediate unstressed syllable. The stressed syllables were immediately adjacent. Nota that is also meant that the time interval between the two stressed syllables was successively reduced from long (215 ms) through short (105 ms) to zero.

All three target items were embedded in simple sentence frames that were, according to extensive pilot testing, semantically neutral and thus equally compatible with both the one-word and the two-word interpretation of each item. The sentence frames ended on a noun that was associated with the nuclear accent. Thus, the two stressed target syllables of a target item were always in prenuclear position in the stimuli.

2.3. F0 and duration manipulation

The sentence frames were realized with target items embedded by a trained female phonetician (JW, 2nd author), a native speaker of Northern Standard German. She produced prosodically neutral base stimuli that served as starting points for the F0 and duration manipulation. The term 'neutral' means two things here: First, a flattened F0 contour; and, second, target items with two equivalently (weakly) stressed syllables that made the target items ambiguous with respect to one-word or two-word interpretations (checked by pilot tests). The recordings were made digitally (48 kHz/24-bit) at Kiel University.

F0 and duration manipulations were done in Praat and resynthesized using PSOLA [22]. The duration manipulation was based on a triangular shape that spanned the stressed syllables and peaked in the stressed vowel. It consisted of equal-sized increases of 30 %, implemented by successively lifting the peak of the manipulation triangle. Thus, the lengthening mainly concerned the syllable nucleus, which corresponds to the way duration changes are realized in natural speech production [23,24]. The duration continuum had 6 steps: 40 %, 70 %, 100 %, 130 %, 160 %, 190 %. Percentages were used in order to follow the accepted method of [6].

The F0 manipulation was done by first creating rising-falling F0 peaks on the two stressed syllables (H*, [32]) and then raising the peak maximum of one of these peaks in 10 equal-

sized steps of 0.5 semitones (st). In order to reduce the overall number of stimuli, we refrained from implementing all 10 F0 steps at each duration step. Instead, we implemented only 5 F0 steps at each duration step, and started one step higher in the F0 continuum for each step higher in the duration continuum. This strategy is schematically illustrated in Figure 2. The manipulation strategy and its two parameter ranges were submitted to feasibility checks in series of formal pre-tests.

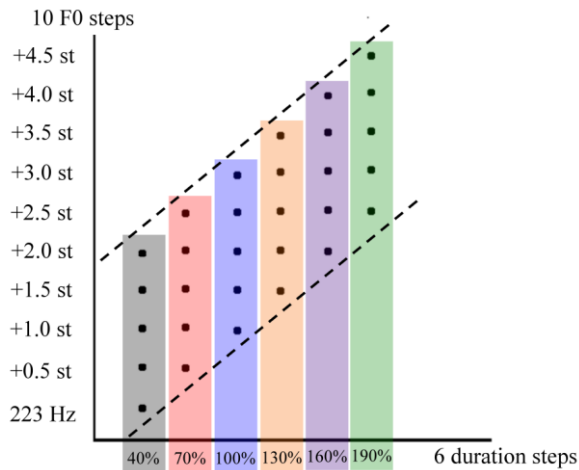


Figure 2: F0 and duration changes in the stimuli; all 6 duration steps (+30 %) from 40 % to 190 % were combined with 5 equal-sized F0 steps (+0.5 st). Each dot represents one combined F0 and duration condition.

The orthogonally combined F0 manipulations were conducted in two prominence-cue orders. In the order D-F0, duration was changed in the first and F0 on the second stressed syllable of the target item. In order F0-D, it was the first stressed syllable whose F0 was changed, whereas the second stressed syllable was changed in duration. Note that, with respect to the H* pitch accents on the two stressed syllables, F0-D is a constant *downstep* and D-F0 a constant *upstep* condition, cf. [30].

The F0 and duration manipulations yielded 60 stimuli for each target item, 30 with the cue order D-F0, and another 30 with the cue order F0-D; and as we included three target items, a total of 180 stimuli were resynthesized.

2.4. Experimental procedure

The PRAAT-MFC platform was used to conduct the perception experiment on individual desktop PCs in a silent, sound-treated room at Kiel University. Fifteen separate sessions were conducted, each with 3-5 participants.

At the beginning of a session, the participants were familiarized with the target items and their one-word and two-word interpretations. Then, they were informed that their task would simply be to listen to 180 stimulus utterances consisting of the target items embedded in constant sentence frames and decide spontaneously after each stimulus whether they had perceived the respective one-word or two-word interpretation.

The instruction was followed by a training set of 12 stimulus utterances that included all three target items but were otherwise randomly selected for each participant.

The participants heard the stimuli over headphones at a constant, pre-adjusted loudness level.

When listening to the stimuli, the participants saw the corresponding orthographic representations of the one-word

and two-word interpretations on the screen and had to click on either of them in order to make their judgments. The 180 stimuli were presented without repetition in individually randomized orders. A two-minute half-time break was made after the 90th stimulus in order to reduce fatigue and auditory habituation artifacts like the speech-to-song illusion [25].

3. Results

Differences in the frequencies of one-word interpretations of the target items across all 61 participants were analyzed in a four-way repeated-measures ANOVA based on the independent variables F0, duration, cue order (D-F0/downstep vs. F0-D/upstep), and the number of unstressed syllables.

The number of unstressed syllables in between the two stressed ones had no significant main effect on the one-word and two-word interpretations of the target items, neither was any of the corresponding interactions significant, which is why the three unstressed-syllable conditions were pooled in the descriptive results summary of Figures 3(a)-(b).

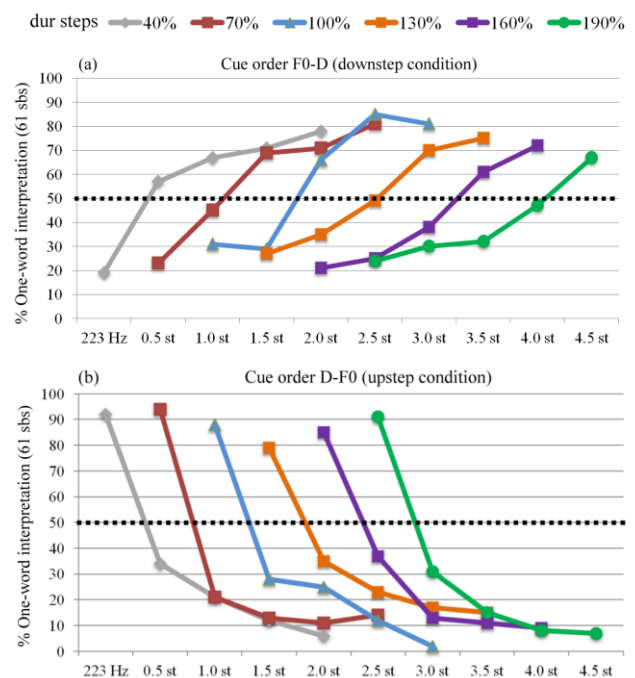


Figure 3: Percentages of one-word interpretations for each combination of duration and F0 manipulation in the F0-D (a) and D-F0 (b) cue-order conditions, pooled across the three target items and all 61 participants ($n=183$ per value). The dotted line marks the 50 % crossover point in target-item interpretation.

The main effect of cue order was significant ($F[1,59]=154.348, p<0.001, \eta^2_p=0.723$). One reason for this is that either target-word perceptions were clearer and F0 changes generally more effective in the D-F0 than in the F0-D condition (0-90 % one-word identifications in Fig. 3b vs 20-80 % in Fig. 3a). Another reason is that, overall, the F0-D condition caused more one-word interpretations than the D-F0 condition.

Furthermore, increases in F0 supported one-word interpretations in the F0-D condition, but suppressed one-word interpretations in the D-F0 condition. Thus, since the dependent variable was in all cases the frequency of one-word interpretations, we found significant interactions of cue order with

both F0 ($F[4,236]=83.990$, $p<0.001$, $\eta_p^2=0.587$) and duration ($F[5,295]=26.283$, $p<0.001$, $\eta_p^2=0.252$).

The two prominence variables F0 and duration also yielded significant main effects, with the effect of duration being stronger ($F[5,295]=52.432$, $p<0.001$, $\eta_p^2=0.471$) than that of F0 ($F[4,236]=11.157$, $p<0.001$, $\eta_p^2=0.159$). There was also a small but significant interaction between F0 and duration ($F[20,1180]=4.205$, $p<0.001$, $\eta_p^2=0.067$), reflecting that there were fewer one-word interpretations under strong syllable lengthening.

However, the most important result was that the size of the change in F0 or duration that was required to reverse target-item interpretation (in terms of the majority of judgments) was roughly constant *within* each cue-order condition, but of different magnitude *between* the two cue-order conditions. The 50 % crossover points in target-item interpretation that are marked by the dotted lines in Figures 3(a) and (b). Compare the distances between these cross-over points in Figure 3(a) with those in Figure 3(b). The exact 50 % crossover points were interpolated from logistic regression models (logit) fitted to the response curves in Figures 3(a) and (b). Table 2 shows that it required on average 1.51 steps in the F0 continuum, i.e. an increase of 0.76 st in pitch-accent peak maximum, to counterbalance a duration change of 30 % in the F0-D condition (Fig. 3a). In contrast, in the D-F0 condition, only 0.98 steps in the F0 continuum, or an increase of 0.49 st in pitch-accent peak maximum, were on average required to counterbalance a duration change of 30 % (Fig. 3b). A paired-sample t-test comparing these sets of st-values between the F0-D and D-F0 conditions across all 61 participants showed that they differed highly significantly ($t[60]=11.51$, $p<0.001$).

Table 2: 50 % crossover points in target-item interpretation in terms of the interpolated step in the F0 continuum (st) at each step in the duration continuum. Values were averaged across all 61 participants and shown separately for the two cue-order conditions; right column: mean F0 increase across the 6 duration steps (i.e. per 30 %).

	40 %	70 %	100 %	130 %	160 %	190 %	σ Increase in F0
F0-D	0.35	1.10	1.80	2.50	3.25	4.05	+0.76
D-F0	0.35	0.80	1.30	1.85	2.35	2.80	+0.49

4. Discussion

If the one-word and two-word interpretations in the stimuli are translated back into stronger prominences in the target items' first or second stressed syllable, then the following conclusions can be drawn from the present results.

F0 peak range extension increases prominence. The same applies to syllable lengthening. In these respects our experiment replicated what was found before in many previous studies on German [6,10,11,12,13] and other languages [2,26]. The new contribution of our study is that its results allow to *quantify* the interplay of the two prominence cues, providing estimates for their relative cueing power. An increase in F0 of less than 1 st is needed in order to outweigh an increase in duration of 30 % on a neighboring syllable. More specifically, it were 0.76 st in the F0-D (downstep) and only 0.49 st in the D-F0 (upstep) condition that counterbalanced a 30 % change in syllable duration. These numbers were surprisingly stable and applied to a large range of F0 and duration values,

independently of the number of unstressed syllables (0-2) in between the two stressed ones. This stability across a large value span in combination with the 61 listeners who participated in our experiment are in favor of the generalizability of our findings and their usefulness as numerical points of reference in phonetic models and applications. Note that the cue power estimates probably include the Gussenhoven-Rietveld-Effect [27]. But, as this effect also occurs in real speech, this is no flaw in terms of generalization or validity.

An increase of 1 st or less is so small that it hardly exceeds the just-noticeable difference for F0 changes in speech [28], whereas a duration change of 30 % is well above this perceptual threshold (of about 10 %), [29]. This fact supports the conclusions of [6,10] and others that F0 is a more powerful cue to perceived prominence in German than duration. Moreover, it is known for a long time that the phonological feature of 'upstep' has a prominence-boosting effect [30]. The constant difference between the F0-D and D-F0 conditions in our results allows us to put a number on this prominence-boosting effect in German: 0.27 st. That is, in combination with 'upstep', F0 increases could afford to be 0.27 st smaller and still counterbalanced a 30 % increase in duration; 0.27 is about one third of 0.76. Thus, our results suggest that 'upstep' adds about one third to the prominence created by a pitch-accent F0 peak.

Of course, these are all still rough estimates. Therefore, the main task of follow-up studies will be to refine these estimates based on more fine-grained F0 and duration manipulations and provide further estimates for other prosodic contexts. For example, we focused on prenuclear accents in the present study. But, as nuclear accents have an inherently higher perceived prominence [14,31], F0 increases on nuclear syllables are likely to require duration increases of more than 30 % (on an adjacent prenuclear syllable) in order to be counterbalanced. Similarly, it has repeatedly been shown that, everything else being equal, different pitch-accent types are associated with different prominence levels in German. We only investigated the most neutral and frequent type H* in the present study. H+L* (or H+!H*) typically creates a lower and L*+H a higher prominence than H* [32]. Thus, compared to H*, duration increases compensating for a given F0 increase probably have to be smaller for H+L* and larger for L*+H.

We have no indications in our data that listener-specific variables like gender or age could also play a role in estimating the relative prominence-cueing power of F0 and duration. However, given the results of [28], listeners' musical training could be worth taking a closer look at in future studies. Then, of course, there is also the interplay between F0 and intensity as well as the interplay of duration and intensity for which relative cue-power estimates needs to be determined. Word class and semantic context [33] could also be taken into account as important secondary factors.

To sum up, in the words of the elephant metaphor of [7], the future challenge of prominence research is not only to bring together the different parts of the same elephant; actually, there is also still a lot of the elephant's anatomy that remains to be discovered and quantified.

5. Acknowledgements

The authors are greatly indebted to all students of the MA seminar on "Contrastive Phonetics: Perception" (winter term 2013, Kiel University) for their help in developing and creating the stimuli and carrying out the perception experiment.

6. References

- [1] G. Kochanski, E. Grabe, J. Coleman, and B. Rosner, "Loudness predicts prominence: Fundamental frequency lends little", *J. Acoust. Soc. Am.*, 118, pp. 1038-1054, 2005.
- [2] D.B. Fry, "Experiments in the perception of stress," *Language and Speech*, 1, pp. 126-152, 1958.
- [3] T. Gay, "Physiological and acoustic correlates of perceived stress," *Language and Speech*, 21, pp. 347-353, 1978.
- [4] A.C.M. Rietveld and C. Gussenhoven, "On the relation between pitch excursion size and prominence," *Journal of Phonetics*, 13, pp. 299-308, 1985.
- [5] C. Gussenhoven, B.H. Repp, T. Rietveld, H.H. Rump, and J. Terken, "The perceptual prominence of fundamental frequency peaks," *J. Acoust. Soc. Am.*, 102, pp. 3009-3022, 1997.
- [6] K.J. Kohler, "The perception of prominence patterns," *Phonetica*, 65, pp. 257-269, 2008.
- [7] P. Wagner, A. Origlia, C. Avesani, G. Christodoulides, F. Cutugno, M. D'Imperio, D. Escudero Mancebo, B. Gili Fivela, A. Lacheret, B. Ludusan, H. Moniz, A. Ni Chasaide, O. Niebuhr, L. Rousier-Vercruyssen, A.-C. Simon, J. Šimko, F. Tesser, and M. Vainio, "Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence," *Proceedings 18th International Congress of Phonetic Sciences, Glasgow, Scotland*, 2015, pp. 1-5.
- [8] H. Mixdorff, "Speech Technology, ToBI, and Making Sense of Prosody," *Proceedings 1st International Conference of Speech Prosody, Aix-en-Provence, France*, 2002, pp. 25-29.
- [9] D.O. Johnson and O. Kang, "Automatic prominent syllable detection with machine learning classifiers," *International Journal of Speech Technology*, 18, pp. 583-592, 2015.
- [10] B. Andreeva, W.J. Barry, and I. Steiner, "Producing phrasal prominence in German," *Proc. 16th International Congress of Phonetic Sciences, Saarbrücken, Germany*, 2007, pp. 1209-1212.
- [11] S. Baumann, O. Niebuhr, and B. Schroeter, "Acoustic Cues to Perceived Prominence Levels: Evidence from German Spontaneous Speech," *Proceedings 8th International Conference of Speech Prosody, Boston, USA*, 2016, pp. 711-715.
- [12] B. Andreeva, W.J. Barry, and M. Wolska, "Language differences in the perceptual weight of prominence-lending properties," *Proceedings 13th Interspeech Conference, Portland, USA*, 2012, pp. 1-4.
- [13] O. Niebuhr, "F0-based rhythm effects on the perception of local syllable prominence," *Phonetica*, 66, pp. 95-112, 2009.
- [14] S. Baumann, K. Eckart, and A. Riester, "Quantifying Prosodic Prominence for Research in Information Structure," *Proceedings Workshop on Prosody and Information Structure in Stuttgart (PINS), Stuttgart, Germany*, 2016.
- [15] D.B. Fry, "Duration and intensity as physical correlates of linguistic stress," *J. Acoust. Soc. Am.*, 27, pp. 765-768, 1955.
- [16] D.H. Klatt, "Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception," *J. Acoust. Soc. Am.*, 53, pp. 8-16, 1973.
- [17] G. Fant, and A. Krukenberg, "Preliminaries to the study of Swedish prose reading and reading style," *Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh.*, 2, pp. 1-83, 1989.
- [18] Ch. Jensen, and J. Tøndering, "Choosing a Scale for Measuring Perceived Prominence," *Proceedings 8th Interspeech, Lisbon, Portugal*, 2005, pp. 2385-2388.
- [19] H. Mixdorff, and C. Widera, "Perceived Prominence in Terms of a Linguistically Motivated Quantitative Intonation Model," *Proceedings Eurospeech Conference, Aalborg, Denmark*, 2001, pp. 403-406.
- [20] D. Arnold, P. Wagner, and B. Möbius, "Evaluating different rating scales for obtaining judgments of syllable prominence from naïve listeners," *Proceedings 16th International Congress of Phonetic Sciences, Hong Kong, China*, 2011, pp. 252-255.
- [21] D. Arnold, B. Möbius, and P. Wagner, "Comparing word and syllable prominence rated by naïve listeners," *Proceedings 12th Interspeech Conference, Florence, Italy*, 2011, pp. 1877-1880.
- [22] P. Boersma, "Praat, a system for doing phonetics by computer," *Glott International*, 5, pp. 341-345, 2001.
- [23] D.H. Klatt, "Synthesis by rule of segmental duration in English sentences," In B. Lindblom, S. Öhman (eds.), *Frontiers of Speech Communication Research* (pp. 287-299). London/New York: Academic Press, 1979.
- [24] K.J. Kohler, "A model of German intonation," *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung*, 25, pp. 295-360, 1991.
- [25] D. Deutsch, R. Lapidis, and T. Henthorn, "The Speech-to-Song Illusion," *J. Acoust. Soc. Am.*, 124, p. 2471, 2008.
- [26] B. Andreeva and W.J. Barry, "Fine phonetic detail in prosody: Crosslanguage differences need not inhibit communication," In O. Niebuhr (ed.), *Prosodies - context, function, and communication* (pp. 259-288). Berlin/New York: de Gruyter, 2012.
- [27] D.R. Ladd, J. Verhoeven, and K. Jacobs, "Influence of adjacent pitch accents on each other's perceived prominence: two contradictory effects," *Journal of Phonetics*, 22, pp. 87-99, 1994.
- [28] J. t'Hart, "Differential sensitivity to pitch distance, particularly in speech," *J. Acoust. Soc. Am.*, 69, pp. 811-821, 1981.
- [29] S.G. Nootboom and G.J. Doodeman, "Production and perception of vowel length in spoken sentences," *J. Acoust. Soc. Am.*, 67, pp. 276-287, 1980.
- [30] S. Baumann, M. Grice, and S. Steindamm, "Prosodic Marking of Focus Domains - Categorical or Gradient?" *Proceedings International Conference of Speech Prosody, Dresden, Germany*, 2006, pp. 301-304.
- [31] G. Ayers, *Nuclear accent types and prominence: Some psycholinguistic experiments*. PhD thesis, Ohio State University, USA, 1996.
- [32] S. Baumann and C. Röhr, "The perceptual prominence of pitch accent types in German," *Proceedings 18th International Congress of Phonetic Sciences, Glasgow, UK*, 2016, pp. 1025-1028.
- [33] F. Kleber and O. Niebuhr, "Semantic-context effects on lexical stress and syllable prominence," *Proceedings 5th International Conference of Speech Prosody, Chicago, USA*, 2010, pp. 1-4.