Rhythm influences the tonal realisation of focus

Nadja Schauffler1, Katrin Schweitzer2

1Institute of English Linguistics, University of Stuttgart, Germany
2Institute for Natural Language Processing, University of Stuttgart, Germany

nadja.schauffler@ifla.uni-stuttgart.de, katrin.schweitzer@ims.uni-stuttgart.de

Abstract

Several studies suggest that rhythm affects different aspects in speech production and perception. For example, in German, discourse structure is normally marked by pitch accent placement and pitch accent type, however, there is variation that cannot be explained by purely semantic or syntactic factors. Prosody-inherent factors, like rhythm, can contribute to this variation. This becomes evident in prosodically more complex environments: while the prosody of utterances containing one focused constituent is well investigated and rather clear-cut, the prosodic organisation of multiple contrastive foci is less clear. In double-focus constructions, for example, two focused constituents demand prominence, possibly resulting in the realisation of two pitch accents. If these pitch accents are required on adjacent syllables they conflict with rhythmic preferences. We present a sentence reading experiment investigating the tonal realisation of two focused constituents and how their contours affect each other in different rhythmic environments. Specifically, we tested whether a potential pitch accent clash in a sentence with two corrective foci influences the pitch excursion and the absolute peak height of the accented syllables. The results demonstrate that rhythmic constraints affect the organisation of the tonal marking of corrective focus.

Index Terms: rhythm, prosody, contrastive focus, melodic effects, F0 parametrisation

1. Background

It is well accepted that information structure and prosody interrelate. For instance, in German, information structure can be marked by both accent placement and accent type (e.g. [1, 2, 3]). Focused constituents are (typically) prosodically marked by a pitch accent, like in “Does Mike eat cakes? – No, he BAKES cakes.” – where we expect a contrastive accent on “bakes” (capital letters indicate focus expected to be marked by pitch accent throughout the paper). However, the actual choice of pitch accent type as well as the distribution of pitch accents in a phrase can vary beyond what discourse structure can explain. Several studies suggest that prosody-inherent factors, like rhythm, contribute to this variation in pitch accent placement. Speakers apparently use deaccentuation in order to avoid two accented syllables (e.g. [4, 5]), or two stressed syllables directly following each other (e.g. [6, 5, 7]). Therefore pitch accents may be omitted, or shifted in order to avoid a clash. For instance, [4] found cases in a German radio news corpus where pitch accents are shifted from the noun to the adjective in order to prevent an accent clash due to another accented syllable following the noun.

The purpose of the present experiment is to investigate the prosodic realisation of two adjacent contrastive accents in double-focus environments, like e.g. “Does Mike eat cakes? - No, he BAKES BREAD”. In particular, we test how double-accent realisations are influenced by rhythmic factors, namely the avoidance of pitch accent clashes. Given the findings by the aforementioned studies, we hypothesise that speakers will avoid the realisation of a pitch accent when its production would lead to a clash. To test our hypothesis, an experiment based on a sentence-reading task was designed to investigate whether double-focus constructions always surface with two pitch accents regardless of rhythmic constraints. It is important to note that for this data, there are no pitch accent labels available yet – the current analysis solely concentrates on acoustic features. The parameters we investigate, absolute peak height and pitch excursion, have been shown to be cues used to convey prominence and to correlate with contrastiveness (cf. e.g. [8, 9, 10]). We expect that rhythm affects the use of these cues. In double-focus constructions two foci are expected to be prosodically marked, potentially by two pitch accents. We expect that it plays a role how close together they are.

We are aware that we cannot infer from purely acoustic parameters to perception. Perception will be included in future work in order to get a comprehensive picture of contrastiveness. Nevertheless, the current analysis gives insight into how the rhythmic environment influences the tonal organisation of prosodic focus marking. In our test-cases the foci were on sentential objects which were corrections to previously mentioned objects. It has been debated how an accent marking contrastive focus looks like phonetically and phonologically (see e.g. [11, 12]). Our study investigates if prosody-inherent factors, like rhythm, need to be taken into account in addressing such questions.

2. Data elicitation

The data for the current study was elicited via a reading production experiment.

2.1. Participants

Sixteen (5 men, 11 women) German native speakers participated in the experiment. Their mean age was 27.25 years (range: 19 to 33) and none of them had known speech or reading disorders. All participants were naïve as to the purpose of the experiment. They were paid for their participation.

2.2. Stimuli

In order to test the prosodic realisation of double-focus environments, a set of stimuli was constructed containing question-answer pairs. The mini-dialogues were designed in such a way that two objects introduced in the question needed to be corrected in the answer. To investigate influences of the rhythmic environment, two conditions per sentence type were included:
one eliciting pitch accents on successive syllables, which therefore directly follow each other (clash condition) and one condition in which the potential pitch accents are separated by an unaccented intervening syllable (no clash condition). Examples are given in (1); lexically stressed syllables are underlined.

Since phrase length matters in the distribution of accents (cf. [13]), the stimuli were controlled for number of syllables (8 words and 13 syllables starting from the embedded clause). In the double-focus conditions, the first object's final syllable was the one bearing lexical stress. The second object always had four syllables with lexical stress on the initial syllable in the clash condition and lexical stress on the second syllable in the no-clash condition. In order to avoid a segmental influence on the tonal marking, which would be especially expected for stops, there were only continuants in the coda of the first object and no voiceless stops in the onset of the second object. Additionally, all "contrasting pairs" (first objects and second objects) were controlled for word form frequency which were taken from the Leipzig Wortschatz corpus [14].

2.3. Procedure

Twenty sentences per condition (clash, no clash, F1, F2) were distributed over 4 lists using a Latin Square Design. The experimental sentences in each list were pseudo-randomised for each participant so that the first 3 mini-dialogues were fillers and that sentences of the same condition were not successive. One list contained 20 experimental sentences and 40 filler sentences. The context questions of each question-answer pair were pseudo-randomised for experimental sentences in each list were pseudo-randomised for

Two control conditions were added to each sentence type, in order to see the tonal realisations of single-focus constructions. Two control conditions were added to each sentence type, in order to see the tonal realisations of single-focus constructions.

Additionally, all "contrasting pairs" (first objects and second objects) were controlled for word form frequency which were taken from the Leipzig Wortschatz corpus [14].

The recordings were automatically segmented into words, syllables and phonemes [15]. The analyses were carried out on the stressed (and potentially accented) syllable of the target words (underlined syllables in examples (1) and (2)).

3. Data processing and analyses

The recordings were automatically segmented into words, syllables and phonemes [15]. The analyses were carried out on the stressed (and potentially accented) syllable of the target words (underlined syllables in examples (1) and (2)).

3.1. PainEt: Parametric Intonation Modelling

We captured details of the shape of the tonal contour on the syllables of interest using the PainEt model [16], which approximates a peak in the smoothed F0 contour by employing a model function operating on a 3-syllable window. There are 6 free parameters in the function term which are set by the model so that the actual F0 shape is fit best. They are linguistically meaningful:

- parameter $b$ locates the peak within the 3-syllable window,
- parameter $d$ encodes its absolute height.
- The remaining parameters specify the amplitude and (amplitude normalised) steepness of the rise before, and the fall after the peak (parameters $c_1$ and $c_2$ for the amplitude and $a_1/a_2$ for the steepness).

Figure 1 illustrates the function.

As mentioned before, we expect potential differences in pitch excursion and absolute peak height. For pitch excursion, the relevant PainEt parameter varies according to the pitch contour: In rising contours, $c_1$ is expected to be higher than $c_2$, whereas in falling contours, $c_2$ is expected to be higher. Thus, the higher value of either of them was taken. That is, the analysis captures changes in pitch excursion independent of pitch accent type. For absolute peak height we directly employed parameter $d$. Since we are looking at adjacent syllables in clash and F1, the size of the PainEt approximation window (3 syllables) is potentially problematic. If a peak is found on any of these syllables, its values will be encoded for each of the syllables: the values capturing the peak's shape are expected to be the same, only the temporal alignment within the 3-syllable window will differ. Therefore, we reduced our dataset in such a way, that only syllables were retained, where the $b$ parameter was between 0 and 1.2 ensuring that the peak is not realised further into the next syllable than 20% of that syllable's duration. This reduced the dataset considerably (148 of the 638 test items, i.e. 23% were removed). However, the reduction ensures that the $c$ and $d$ values employed match the syllable in question.

3.2. Statistical analyses

For each parameter under investigation, we performed a linear mixed effects analysis to investigate the relationship between that parameter and the rhythmic and semantic context. In all analyses, we tested the following factors as fixed effects: the experimental condition (clash, no clash, F1, F2), the type of the object (first object vs. second object), the interaction of these two factors, and the trial number (to control for possible learning effects). As random factors, we tested intercepts...
The reference between the two objects is encoded in the trajectory of interception (i.e. the reference condition), is about 58 Hz. The difference between the two objects is about 34 Hz greater on OBJ2 than on OBJ1. For no clash, the objects are realised with a similar excursion (β = -1.874, SE = 5.160, t = -0.363).

For the first object (OBJ1, circles on the left side of the plot), the results show that all conditions having a focus differ significantly from the unfocused condition (F2): no clash by about 26 Hz (SE = 5.976, t = 4.301), clash by about 24 Hz (SE = 6.218, t = 3.935), and F1 by about 33 Hz (SE = 6.036, t = 5.200), while they do not differ from each other. That is, peak excursion on OBJ1 reflects the difference between unfocused first objects (F2) and focused ones (F1, clash, no clash).

For the second object (OBJ2, circles on the right side of the plot), the unfocused condition F1 differs from all other conditions significantly in that pitch excursion is greater in all the other cases: by about 13 Hz (SE = 5.923, t = 2.14) for no clash, by about 21 Hz (SE = 6.042, t = 3.514) for F2, and by about 30 Hz (SE = 6.053, t = 4.998) for clash. Interestingly, among the focused conditions on OBJ2, only the clash and the no clash condition differ significantly from each other with respect to their pitch excursion. In the clash condition, pitch excursion is about 18 Hz higher than in the no clash condition (SE = 5.352, t = 3.278). That is, peak excursion on OBJ2 reflects the difference between the two conditions differing only in their rhythmic make-up: If the two potential focus markings are on syllables directly following each other (condition clash), pitch excursion is greater when there is an intervening syllable.

4.2. Absolute peak height

Figure 3 illustrates the effect that object and condition have in the model predicting absolute peak height of the tonal contour realised on relevant syllable. The difference between the two
objects is encoded in the trajectory of the lines in the plot. The two single-focus conditions, F1 and F2, realise OBJ1 significantly different from OBJ2: compared to the focused OBJ1 (where the peak reaches about 220 Hz) the peak on the unfocused OBJ2 is realised about 31 Hz lower in condition F1 (SE = 6.754, t = -4.694). In cases where OBJ2 is focused (F2), the peak is realised about 29 Hz higher than in the unfocused OBJ1 (SE = 7.876, t = 3.690) where it reaches about 193 Hz. Both double-focus conditions realise OBJ1 and OBJ2 with similar peak height.

For OBJ1 (circles on the left side of the plot), the results show that absolute peak height only significantly differs between the unfocused condition F2 and the focused ones (F1, clash, no clash): the peak in condition F2 is at about 193 Hz and thereby about 22 Hz lower than in the no clash condition (SE = 5.779, t = 3.814), about 24 Hz lower than in clash (SE = 5.926, t = 4.027) and about 27 Hz lower than in F1 (SE = 5.811, t = 4.668). The two double-focus conditions clash and no clash do not differ from each other (t = 0.481).

For OBJ2 (circles on the plot’s right side), the results show that, as in OBJ1, only the condition in which OBJ2 is unfocused (F1), differs from the ones, where OBJ2 is focused. The peak on OBJ2 is at about 193 Hz and thereby about 27 Hz lower than in no clash (SE = 5.704, t = 4.667) and than in clash (SE = 5.721, t = 4.782), and about 34 Hz lower than in F2 (SE = 5.502, t = 6.113). The double-focus conditions clash and no clash do not differ significantly from each other (t = 0.717).

To sum up, while absolute peak height is generally influenced by the focus structure, the differences in rhythm have no effect.

5. Discussion and Conclusion

We presented stimuli elicited from a production study where we compared different focus constructions. The parameters investigated were pitch excursion and absolute peak height, derived from a parametric intonation model which approximates potential peaks in the contour. The results have to be interpreted carefully for two reasons. Firstly, the analysis is purely acoustic, i.e. perception is not taken into account and we cannot automatically infer from acoustics to perception. It has been claimed, however, that a Hertz scale reflects listeners’ prominence judgement the best [10]. Secondly, the analysis is based on purely tonal features, i.e. other parameters known to influence prominence and, most importantly, pitch accenting, such as intensity and duration are not taken into account. Nevertheless, this study presents an insight of how our data is realised and how double-focus constructions differ, when their rhythmic pattern is manipulated. The results also clearly show, that focused and unfocused expressions differ in the analysed parameters. Focus is reflected by a greater pitch excursion, i.e. greater amplitude of the rising or falling contour on the syllable (compared to its immediate context), and by a higher peak height, possibly reflecting a pitch accent. That is, the acoustic parameters employed are suitable to investigate the tonal realisation of focus.

The two conditions that were designed to test rhythm as a prosody-inherent factor potentially influencing the tonal realisation of focus, differed in pitch excursion. If accentuation of both foci in double-focus constructions would lead to a clash, i.e. if the second focus semantically requires prosodic focus marking on a syllable directly following the stressed syllable of the first focus, pitch excursion was higher on the second object. That is, in cases of a rhythmic clash, speakers alter the tonal realisation of the focus. Presumably, when the two required prominences to mark focus are directly adjacent, speakers add prominence-lending cues to the second one, in order to mark it distinctly. The difference between the pitch excursions on the first and second object in the clash condition (about 19 Hz) is remarkable: it is about the same difference as between the two objects in the single focus condition F1 (about 20 Hz). It is highly likely that F1 was realised with a pitch accent on OBJ1 (induced by the focus), which results in the difference to the unfocused (and therefore unaccented) OBJ2. The question is how the same difference is perceived when it is realised on two foci, in the clash condition. First results from a prominence-judgement experiment using the recordings of this study show that listeners judge the first object in the clash condition sig. more often as not prominent compared to the no-clash condition. This suggests that the increased excursion on the second object in the clash condition perceptually removes prominence from the preceding object.

In the no clash condition, speakers produce the two objects with similar pitch excursions. The intervening unaccented syllable seems to be enough to realise two distinct contours. In the above mentioned prominence-judgement task, listeners perceived prominence on both objects more often than in the other conditions. Further perception experiments are planned to investigate whether rhythm will affect the acceptability of the productions in the double-focus constructions.

The results for absolute peak height clearly show that it is used to mark focus tonally, as has been shown before (e.g. [23]). This parameter is employed in single as well as double-focus conditions. However, it does not seem to be influenced by the rhythmic environment. That is, the peak height of the contour on the two objects is the same. Future work will investigate whether speakers preferably produce both foci in one or in two phrases by analysing intonation boundaries and the length of potential pauses. Additionally, we will investigate the timing of the peak and how it is affected by rhythmic manipulations. Preliminary results suggest that peaks are realised later in the clash condition than in the no clash condition which corresponds to the current findings, since delayed peaks are also prominence lending parameters [24].

To conclude, our study demonstrates that the rhythmic environment needs to be taken into account in mapping semantic categories to phonological categories and their phonetic implementation.
6. References


