# ORDER EFFECTS ON THE PERCEPTION OF RELATIVE PROMINENCE

*Nils Jagdfeld & Stefan Baumann*

I*f*L Phonetik, University of Cologne, Germany
{nils.jagdfeld;stefan.baumann}@uni-koeln.de

## ABSTRACT

The paper reports on a perception experiment in German investigating the relative prominence of target words in two positions in a sentence. Listeners judged for each position which of four prominence levels were perceived as accented or unaccented. While subjects clearly distinguished between strongly accented and unaccented (i.e. postlexically stressed and unstressed) words in both positions, we found an order effect for weak pitch accents. They were not perceived as accents in prenuclear but they were in nuclear position. In general, listeners proved to be less sensitive to categorical prominence judgements in prenuclear position (also reflected in longer reaction times) than in nuclear position, which confirms the special functional status of nuclear pitch accents.

**Keywords**: Relative prominence, perception, order effect, reaction time, pitch accent.

## 1. INTRODUCTION

It is generally agreed on that prosodic prominence is an intrinsically gradient phenomenon. However, the annotation of speech often requires the classification of intonational events into categories, forcing a choice e.g. between accent versus lack of accent, or between specific accent types. Furthermore, prosodic prominence is relational in nature, in that an item's prominence can only be determined with respect to neighbouring items. Although this syntagmatic aspect adds important information as to the perception of an item's prominence, it does not (at least not necessarily) solve the problem of making an appropriate paradigmatic (i.e. categorical) choice.

In fact, the kind and number of available categories depends on the theoretical framework chosen. In terms of intonational prominence, e.g., it has been established already in the British tradition (e.g. [9]) but also in autosegmental-metrical approaches (e.g. ToBI) that the nuclear syllable is the last pitch accented and most prominent syllable of an intonation unit. This definition implies on the one hand that prenuclear items are structurally less prominent than nuclear ones, and on the other hand that a pitch accent *after* the nucleus is ruled out.

The motivation for the present study arose from actual labeling exercises using the German ToBI system [4]. We noticed that annotators marked syllables in prenuclear position frequently as accented, whereas similarly prominent syllables were *not* marked as accented if they followed a pitch accent (which was perceived as nuclear). That is, the (presumably) same degree of prosodic prominence on an item seemed to be analyzed as two different intonational categories just by virtue of the item's relative position in a phrase, i.e. either before or after a clearly pitch accented item.

In the experiment reported below we want to find out whether the relative position of an item, which is acoustically marked by a specific level of prominence (strong pitch accent (S), weak pitch accent (W), stress (ST), no stress (noST)) has an influence on the item's perception as carrying an accent or not. In addition to these categorical ratings we will measure (gradient) reaction times.

There is evidence for the four prominence levels in the literature. Primary (i.e. strong, S) and secondary (i.e. weak, W) pitch accents have widely been discussed (e.g. in [8]). Prominences which are associated with postlexically stressed syllables (ST) have been proposed e.g. in the concept of *phrase accent* [5]. Empirical evidence for this prominence level has been found as a marker of *second occurrence focus*, which is predominantly expressed by increased duration in comparison with unstressed syllables (noST) marking background information [2].

We derived three hypotheses from our observations, which will be tested on three groups of listeners: naïve, partly trained, and experts.

***Hypothesis 1*** We expect that words which carry fully-fledged pitch accents (S,W) will be perceived as accented in both prenuclear and nuclear position, while unstressed words (noST) will not be perceived as accented in both positions (i.e. pre- and postnuclear). In contrast, we hypothesize that words marked by a postlexical stress (ST), i.e. by segmental lengthening but not by tonal movement, will be perceived as accented in prenuclear position but not in postnuclear position.

***Hypothesis 2*** Prominence levels at the extreme ends of the scale, such as strong accentuation (S) and no stress (noST), are expected to be judged more easily and more quickly than the in-between levels of weak accents (W) and post-lexical stresses (ST). This should be reflected in longer reaction times for the latter two levels.

***Hypothesis 3*** We expect that experts are generally faster in making prominence judgements than partly trained listeners, and that they are in turn faster than naïve listeners. The same ranking should be found in listeners' recognition of the prominence levels presented.

## 2. METHOD

### 2.1. Stimuli

As a first step into stimulus creation we recorded three repetitions of the utterances in (1) and (2), spoken by a trained male phonetician, with three prominence levels on the target words (strong high accent, weak high accent, no accent/stress). Both test words (*Banane* 'banana', *Melone* 'melon') were three-syllabic and had penultimate stress. However, they differed in the vowel quality of the lexically stressed syllable.

(1) Ich habe eine **Banane** und eine **Melone** umgetauscht. (*I (have) exchanged a banana and a melon.*)

(2) Ich habe eine **Melone** und eine **Banane** umgetauscht. (*I (have) exchanged a melon and a banana.*)

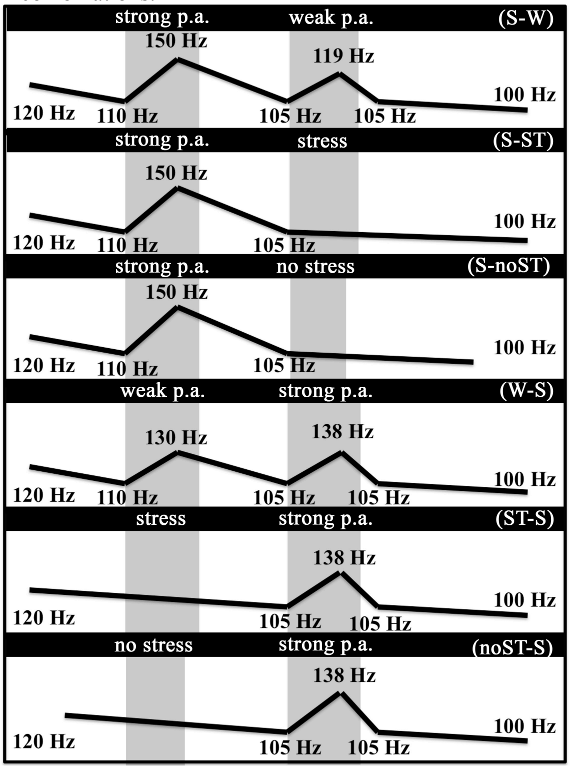
For both target words in both positions we measured the mean values of the parameters *duration*, *F0 peak alignment, peak height* and *peak excursion* for the three prominence levels produced. For noST words, only the duration of the lexically stressed syllable was measured.

In a second step, the test words were manipulated to create four distinct levels of prominence ('stress' in addition to the three produced levels) on the basis of the measurements. According to the actual mean values, we set the duration of the lexically stressed syllable to 250 ms for S, W and ST, and to 200 ms for noST. Peak timing and height were manually manipulated with Praat [3]. The beginning of the tonal rise for weak and strong accents was set to 0% and the F0 peak was reached at 90% of the duration of the accented syllable. The peak height for strong accents was 150 Hz and 130 Hz for weak accents in prenuclear position (Fig.1).

Since the perception experiment was meant to test an order effect in the comparison of items of *acoustically identical* prominence, we created only a single item of a specific prominence level per target word and copied it to the respective position in the test sentence (cross-splicing). In order to take declination into account, we lowered the height of an accentual peak in position 2 (P2) by 1.5 semitones in comparison to the same accent type in position 1 (P1) (following the formula described in [7]). For unstressed and stressed syllables, the F0 value for P1 was the result of interpolation between the start of the utterance and the beginning of the rise of the accent in P2. For unstressed and stressed syllables in P2, the F0 was flat from the beginning of the lexically stressed syllable and continuously falling to the end of the sentence.

The beginning and the end of the utterances were kept constant. For the beginning of each utterance, we set the F0 height to 120 Hz and the end to 100 Hz. In order to avoid the impression of an intra-sentential phrase break, we controlled the fall after the first accent in that it declines linearly to the beginning of the second accent. Finally, the utterances were informally tested to make sure they sounded natural. An overview of the stimuli design is given in Fig.1.

**Figure 1:** Stimuli design for the perception experiment. Shaded areas indicate lexically stressed syllables (position 1 and 2). 'p.a.' stands for 'pitch accent'. Abbreviations in brackets indicate prominence level combinations.




### 2.2. Subjects and procedure

25 right-handed native speakers of German took part in the experiment, which were divided into three groups: a group of ten prosodically 'naïve' students of the University of Cologne (8f, 2m, mean age 21,2 years), a group of seven 'partly trained' listeners which participated in at least one course in intonation at the Universities of Cologne or Düsseldorf (5f, 2m, mean age 28 years), and a group of eight 'experts' in prosody and labeling intonation within the GToBI framework (4f, 4m, mean age 26,3 years).

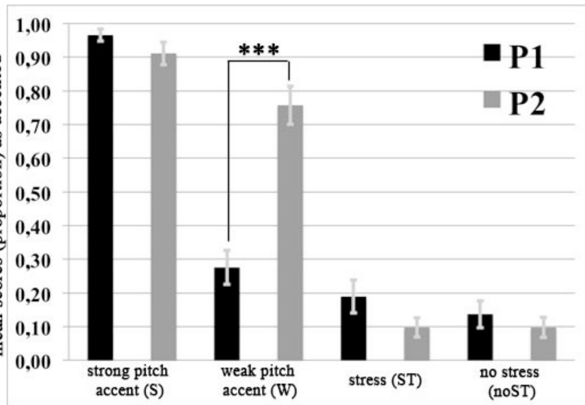
For the perception test, we used the SuperLab Pro 2.04 software [1] and a Cedrus x830 response pad to collect both prominence ratings and reaction times (RT). Subjects had to decide in a forced-choice task for each of the two target words (*Melone*, *Banane*) whether they perceived the word as "accented" or not. The response pad displayed four buttons: for each of the two positions in the utterance (P1: left hand, P2: right hand) there was one button for "accent" (to be pushed with the forefinger) and one button for "no accent" (to be pushed with the thumb). The stimuli were presented in five blocks with two repetitions of each stimulus in each block (in sum 10 repetitions of each stimulus). The order was randomized for each block and participant. The first block (two repetitions) was the training block and thus not analyzed. Subjects listened to the stimuli over headphones at a comfortable volume.

## 3. RESULTS AND DISCUSSION

### 3.1. Prominence judgements (Hypothesis 1)

We calculated the mean accent ratings for every word in each *prominence level* per *position*. A repeated measures ANOVA revealed a highly significant main effect ($F(3,72) = 111.71$; $p<0.001$). Fig.2 gives the mean accent scores showing a general decrease from strongly accented words to unstressed words.

**Figure 2:** Mean proportional ratings as "accents" for the different prominence levels and for all subjects pooled; ratings for P1 in black bars, for P2 in grey bars. For reasons of clarity, only the significant differences between positions within one prominence level are marked.




Hypothesis 1 was partly confirmed, since strong accents were almost always perceived as accented in both positions, whereas unstressed (and stressed) tokens were only rarely perceived as accented. Thus, as expected, the extreme end points of the prominence scale are clearly distinguished (significant difference between strong pitch accents and no stresses in both positions: $p<0.001$). For both prominence levels, there was no significant difference between P1 and P2.

The hypothesis was neither confirmed for (a) words carrying a weak pitch accent (W) nor for (b) postlexically stressed words (ST).

(a) Interestingly, there was a different distribution over the two positions for weak accents in comparison to the three other levels. In prenuclear position, weakly accented words were mostly perceived as unaccented (mean score 0.276), whereas in nuclear position they were mostly perceived as accented (mean score 0.757). This distribution of prominence ratings between both positions revealed a strong order effect ($p<0.001$), which can be explained by the different phonological status of nuclear and prenuclear accents. Since the nuclear position (P2) is both functionally more important (with respect to information structure) and structurally stronger (representing the DTE in terms of Metrical Phonology) than the prenuclear position (P1), listeners were more sensitive to pitch prominence in P2. In other words, listeners *expected* an accent on the final argument in the utterance, so that a pitch accent in P2 can be regarded as the default pattern; as long as this pattern is met, it does not seem to be crucial whether the element in P1 carries an accent or not. This led to fewer accent ratings for weak accents in P1.

In addition, the large number of "accent" ratings for W accents in P2 may be supported by 'downstep enhancement' (see [11] on lexical accents in Japanese) describing the effect that the *perceived* value of an accent in P2, which is considerably lower than a preceding accent, is higher than its *physical* value. Alternatively, the high accent score for weak accents simply reflects a declination effect saying that lower accents in P2 are perceived as equally prominent as a higher accent in P1 (cf. [10],[6]). However, the weak accents in P2 were designed as being considerably lower than a strong accent in P2, so that a downstep enhancement effect seems to be more plausible.

(b) We expected stressed words to be perceived as accented in prenuclear but not in postnuclear position. However, no significant difference between the positions could be found. Still, at least a tendency in this direction was observed (mean accent score P1: 0.189, P2: 0.097). Thus, the perceived prominence of stressed words in P1 was similar to the value for weakly accented items in P1, since the default pattern was met. In P2, again, listeners were more sensitive and perceived the stressed word as clearly not carrying a nuclear accent (in which case we have to talk about *de*accentuation, since an expected default accent is missing).

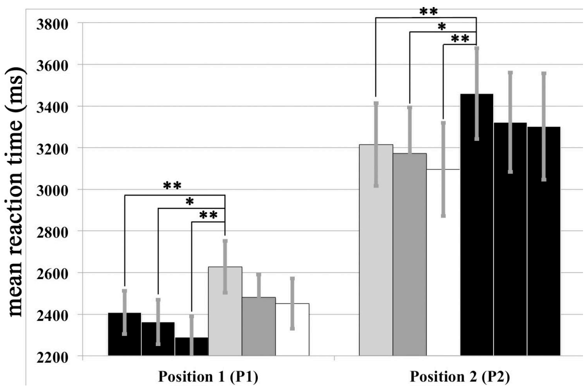
Furthermore, a main effect of *word* could be found ($F(1,24) = 11.615$; $p<0.01$), since the test word *Banane* ('banana') was significantly more often perceived as accented than *Melone* ('melon'). This effect is probably due to a difference in vowel quality, since open vowels like the /a/ in the lexically stressed syllable of *Ba**na**ne* are more sonorant than close-mid vowels like the /o/ in *Me**lo**ne*. Additionally, we suppose that the open vowel is more sensitive to prosodic events (here: pitch accents), as described in [12] for prosodic boundaries.

### 3.2. Reaction times (Hypothesis 2)

For each subject, we determined the median of the RTs in relation to the beginning of each test sentence. RTs outside ± three standard deviations were not taken into account. In contrast to the analysis of prominence judgements we could not compare RTs across the two positions, since the judgement for P2 was always given *after* P1, so that the RTs were necessarily longer for P2. In fact, we found that listeners generally rated both positions after the end of the whole utterance.

We performed two repeated measures ANOVAs, one for each position, with the factor *prominence level* (see Fig.3). A significant main effect was found for both positions ($F(5,60) = 6.027$; $p<0.001$ for P1 and $F(5,60) = 4.927$; $p<0.001$ for P2).

**Figure 3:** Reaction times for the four prominence levels on test items in P1 and P2. Black bars indicate *strong accents*, light grey bars *weak accents*, dark grey bars *stresses* and white bars indicate *no stresses*.




Hypothesis 2 was partly confirmed in that strong accents (S) in P1 were recognized faster than other prominence levels, and that no stresses (noST) were (in tendency) recognized faster than stresses (ST) and weak accents (W). However, *significant* differences could only be found between strong and weak accents in the same position. In P2, on the other hand, noST was recognized fastest. In this position, it was preceded by a strong accent, which apparently was the easiest prominence level combination to judge. That is, the greater the difference between the prominence levels of P1 and P2, the shorter was the RT.

At first sight, it is surprising that RTs were *longer* for strong accents in P2 compared with the other prominence levels. However, Fig.3 shows that the *relation* between the RTs for P1 and P2 remained constant for the six combinations. In other words, the differences in RT depended on the perceived prominence level in P1.

Apparently, it was most difficult for subjects to judge the prominence of weak accents (as hypothesized), showing the longest mean RT. This difficulty is in accordance with the finding discussed above, namely the variability in the prominence judgement of weak accents between P1 and P2 (Fig.2). Although weak accents were mostly perceived as unaccented in P1, subjects were obviously not as sure about their decision as with the other prominence levels. Nevertheless, RTs did not differ significantly from RTs for stressed and unstressed items, which were also judged as unaccented. These longer RTs for W, ST and noST again suggest that listeners were less sensitive for the prominence in P1, after having perceived a nuclear accent in P2 (since they made their judgements after the end of the utterance). In contrast, a strong accent in P1 is rated faster, in particular if it is perceived as the nucleus.

The variable *word* did not reveal a significant effect in terms of RTs.

### 3.3. Between-group ratings (Hypothesis 3)

Hypothesis 3 could not be confirmed, since the ANOVAs did not show a significant effect of the between-subject factor *group*. That is, we neither found differences in categorical prominence judgements nor in RTs between the three groups (naïve, partly trained and expert subjects).

## 4. CONCLUSION

While the data obtained by our perception experiment showed clear results for the extreme poles of the prominence scale, the results for the intermediate levels were somewhat surprising. We found an order effect for weak pitch accents but not for postlexical stresses (in contrast to our hypothesis). That is, weak accents were not perceived as accents in prenuclear but in nuclear position, whereas stresses were generally judged as not accented. Thus, listeners proved to be less sensitive to categorical prominence judgements (for weak accents and stresses) in prenuclear position, which is also reflected in longer reaction times in P1. In nuclear position, however, we observed a clear division in the ratings between pitch accents and no pitch accents, which confirms the special status of nuclear pitch accents in the prosodic hierarchy.

## 5. REFERENCES

[1] Abboud, H. 1991. *SuperLab*. Wheaton, MD: Cedrus.

[2] Baumann, S., Mücke, D., Becker, J. 2010. Expression of Second Occurrence Focus in German. *Linguistische Berichte* 221, 61–78.

[3] Boersma, P., Weenink, D. 2005. Praat: doing phonetics by computer (Version 5.2). *http://www.praat.org/*.

[4] Grice, M., Baumann, S., Benzmüller, R. 2005. German intonation in autosegmental-metrical phonology. Jun, S. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 55-83.

[5] Grice, M., Ladd D.R., Arvaniti, A. 2000. On the Place of Phrase Accents in Intonational Phonology. *Phonology* 17 2, 143-185.

[6] Gussenhoven, C., Repp, B.H., Rietveld, A., Rump, W.H., Terken, J. 1997. The perceptual prominence of fundamental frequency peaks. *JASA* 102, 3009-3022.

[7] Hart, J. 't, Collier, R., Cohen R. and A. 1990. *A perceptual study of intonation.* Cambridge: CUP.

[8] Ladd, D.R., 1980. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

[9] Palmer, H. 1922. *English Intonation with Systematic Exercises*. Cambridge: CUP.

[10] Pierrehumbert, J. 1979. The perception of fundamental frequency declination, *JASA* 66, 363-369.

[11] Shinya, T. 2005. Lexical accent effects on the perception of fundamental frequency peaks in Japanese. *150th ASA Meeting*. Minnesota: Minneapolis.

[12] Tabain, M. 2003. Effects of prosodic boundary on /aC/ sequences: articulatory results. *JASA* 113, 2834-2849.