

The Importance of Tonal Cues for Untrained Listeners in Judging Prominence

Stefan Baumann

IfL Phonetik, University of Cologne, Germany

stefan.baumann@uni-koeln.de

Abstract

The paper discusses the question of how untrained listeners perceive and interpret prosodic and non-prosodic features in a prominence rating task. We adapt the method of Rapid Prosody Transcription (Cole et al. 2010) to read German. Results suggest that tonal cues are most relevant for prominence perception, with rising accents reaching the highest prominence scores. Accent position, in particular the position of the nuclear accent, seems to be less important for untrained listeners. Furthermore, non-prosodic factors related to the accentability of a word based on language-specific knowledge and expectations (part-of-speech, word frequency, morpho-syntactic focus marking) also influence prominence ratings but to a lesser extent.

Keywords: prosody, prominence, untrained listeners, perception, nuclear accent, pitch accent type

1. Introduction

The factors determining the manual annotation of prosodic prominence are not yet fully understood. Usually, annotation is carried out by expert transcribers who may produce artefacts since they are biased by top-down knowledge derived from specific theoretical assumptions, e.g. in their treatment of secondary prosodic prominences in relation to a nuclear accent (defined as the last and both structurally strongest and semantic-pragmatically most important pitch accent in an intonation unit). However, the vast majority of prosody ‘users’ is untrained, and the question the present paper poses is how they perceive and interpret prosodic and non-prosodic features when asked to judge the prominence of words in connected speech.

In order to examine this question we broadly apply the method of *Rapid Prosody Transcription* (RPT), developed by Jennifer Cole and her colleagues (e.g. Cole et al. 2010) for collecting coarse-grained prosodic judgements from untrained listeners based on their integrated perception of form and function. That is, the resulting prominence scores for each test word are regarded as primary data that can be interpreted linguistically.

The present explorative study examines to what extent prominence perception is signal-driven (bottom-up) or expectation-driven (top-down) by investigating the influence on prominence judgements of a) phonetic-phonological (prosodic) factors and b) other factors related to a word’s ‘accentability’.

More specifically, the study asks the following questions:

1. Do untrained listeners perceive nuclear accents as prominent due to their – structural – importance for the message?
2. Do untrained listeners perceive secondary accents as prominent in spite of their weaker structural position (i.e. pre- and postnuclear)?

3. Are acoustic cues (F0 height, F0 movement and alignment, duration, intensity) dominant for prominence judgements?

2. Method

2.1. Material

The test material consists of 60 German sentences (containing between 5 and 18 words) read by 14 native speakers. They were selected from various databases and display different focus structures (including second occurrence focus and verum focus structures) as well as various information status categories of referring expressions (new, accessible and given) marked by varying degrees of prosodic prominence. All sentences are declarative and have low final boundary tones. The main selection criterion was the occurrence of various accent types (AT) in different accent positions (AP), according to a consensus annotation of three intonation experts (trained with GToBI; Grice et al. 2005). Fig.1 shows the categories tested, along a scale of assumed perceptual prominence.



Figure 1: *Accent types (AT) and accent positions (AP) tested in the experiment.*

The accent type (AT) classification takes the tonal movement in the vicinity of the accented syllable into account, in particular information about the onglide to the accented syllable. That is, a falling accent has higher pitch immediately before the accented syllable (comprising the ‘early peak’ accents H+L* and H+!H* in GToBI), whereas a rising pitch accent either has a low pitch target before the accented syllable (L+H* in GToBI) or rises from the accented syllable (L*+H). Low (L*) and high (H*) accents indicate tonal targets on stressed syllables without a considerable tonal movement in their immediate vicinity. The category ‘no pitch accent’ comprises complete lack of accent and phrase accents (see below).

Fig.2 indicates that the duration of a syllable differs as a function of the type of accent it carries: low (and falling) accents trigger the longest durations probably compensating for a lack of (rising) tonal movement as the main cue for prosodic prominence. These data confirm results by Röhr and Baumann (2010).

As to accent position (AP), we do not only differentiate between pre- and postnuclear accents but also between final and non-final nuclear accents. This latter distinction is made in order to capture the potentially secondary status of non-final nuclei, since they are often classified as nuclear only due to the occurrence of hesitation pauses or a slow speech rate.

Small breaks like these trigger the insertion of intermediate phrase boundaries in most ToBI systems (which renders the last pre-boundary accent nuclear by definition), although they often do not appear to be intended by the speaker. Prenuclear accents are fully-fledged pitch accents, while phrase accents are not. They are associated with a stressed syllable in postnuclear position and often lack a considerable tonal movement (see also Grice et al. 2000).

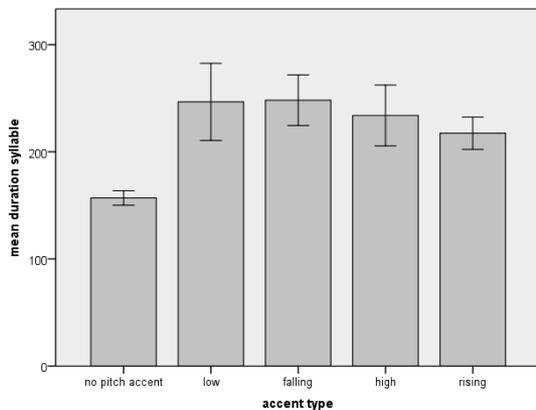


Figure 2: Differences in mean duration of syllables as a function of accent type.

The duration differences given in Fig.3 show that syllables carrying nuclear accents are generally longer than syllables marked by pre- and postnuclear accents. As expected, unaccented syllables are the shortest.

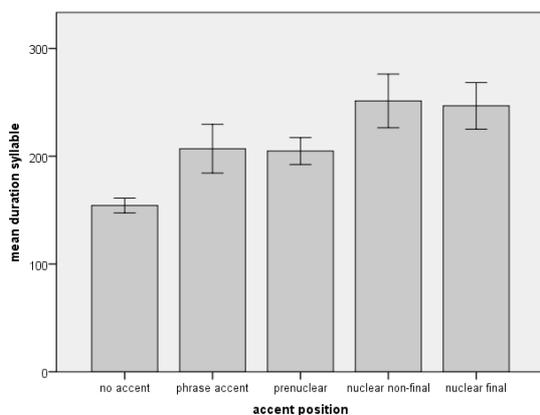


Figure 3: Differences in mean duration of syllables as a function of accent position.

The non-phonetic-phonological factors which have been claimed to affect the predictability and, in turn, accentability of a word, and which are tested in the experiment are word frequency (based on the *SubtLex* corpus, see Brysbaert et al. 2011), part-of-speech and morpho-syntactic focus marking (i.e. presence or absence of a focus-sensitive particle like *nur* ('only') or *sogar* ('even')).

2.2. Procedure

In a self-paced perception test with 28 prosodically untrained native speakers of German (aged between 18 and 58, with a mean of 24.8 years), the 60 utterances were presented over headphones in a silent room in pseudo-randomized order. According to the *Rapid Prosody Transcription* method, only a brief instruction without examples was given. After that, the

subjects had to judge which words they perceive as prominent by underlining them on a sheet of paper displaying the written text. Capitalization and punctuation marks were removed in order to avoid structural cues that may influence the subjects' judgements. The actual instruction was spelled out as follows:

[...] „Ihre Aufgabe besteht nun darin, sämtliche Wörter, die Sie in einer Äußerung als betont / hervorgehoben / wichtig wahrnehmen, auf dem Transkript zu unterstreichen.“
 ([...] *Your task is now to underline all the words on the transcript which you perceive as stressed/ highlighted/ important.*)

The judgements result in a probabilistic prominence score (*p-score*) for each word expressing the proportion of subjects who labelled a word as prominent.

3. Results

The statistical analyses in this pilot study are restricted to descriptive measures (i.e. percentages/p-scores and correlations), since the utterances chosen and thus the data points analysed were not counterbalanced for all factors reported here. A follow-up study will investigate a strictly controlled data set that allows to explore interdependencies between the prosodic and non-prosodic variables, as well as transcriber-specific strategies in judging prominence.

3.1. Inter-transcriber agreement

The agreement among the 28 untrained annotators reached an intermediate value (Fleiss' kappa $\kappa=0.53$; cf. Fleiss 1971, Geertzen 2012). This can be regarded as an acceptable basis for further analyses, since it shows that transcriber-specific variation was moderate.

3.2. Prosodic factors

3.2.1. Accent type and position

Fig.4 shows the interaction between accent type and position indicating that high and especially rising utterance-final nuclear accents received the highest p-scores. In comparison with (high and) rising accents, the scores for low accents are considerably lower in both nuclear and prenuclear position. Falling accents, however, were judged as more prominent than high accents in nuclear non-final and in prenuclear position.

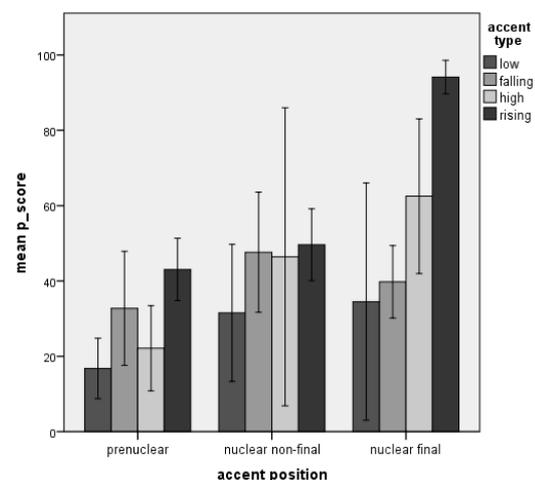


Figure 4: Prominence judgements of combinations of accent type and position.

This finding suggests a crucial role of tonal movement for the given task. Support for this view comes from the fact that subjects hardly recognised phrase accents as prominent, which generally lack a pronounced pitch movement. Interestingly, but as a plausible consequence, in utterances without perceptually salient accents many listeners did not mark any word as prominent. The strong influence of tonal movements also becomes obvious in the example illustrated in Fig.5. Here, both prenuclear rising accents on *Bekannten* ('friend') and *gute* ('good') trigger high p-scores, whereas the nuclear but low accent on *Empfehlung* ('recommendation') is not perceived as prominent.

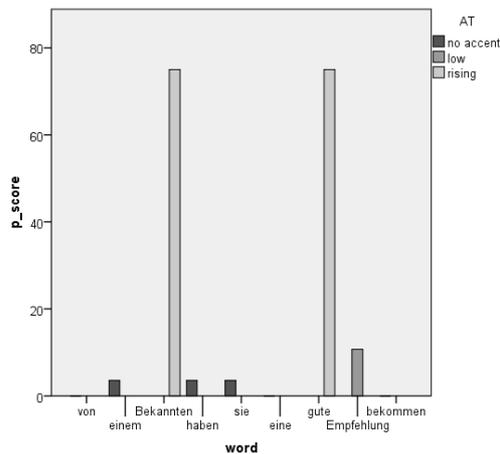


Figure 5: Prominence judgements of different accent types in the test sentence *Von einem Bekannten haben sie eine gute Empfehlung bekommen* ('They got a good recommendation from a friend') (small capitals indicate accented syllables).

3.2.2. Duration

Results for duration measures of vowels in all stressed syllables show the tendency for a linear correlation with prominence judgements: the longer the lexically stressed syllable of a given word, the more prominent the word is perceived. The Pearson's correlation coefficient was found to be slightly higher for phonologically long vowels ($r = 0.40^{**}$; see scatter plot in Fig.6) than for phonologically short vowels ($r = 0.31^{**}$).

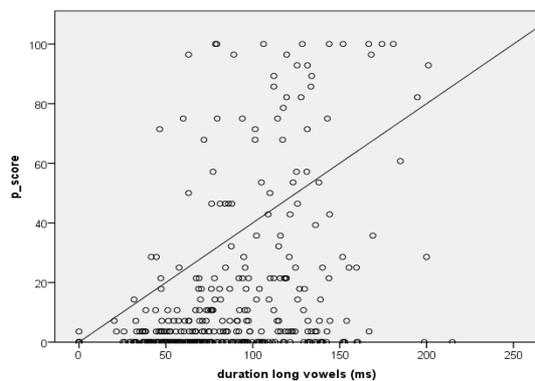


Figure 6: Prominence judgements as a function of vowel duration in lexically stressed syllables (phonologically long vowels only).

3.2.3. Intensity

A similar tendency as for duration could be observed for intensity: the louder the peak intensity (RMS) of a lexically stressed syllable, the higher the p-score. Results reveal a moderate positive correlation between the two variables (Pearson's $r = 0.42^{**}$).

3.3. Non-prosodic factors

3.3.1. Part-of-speech

A word's part-of-speech (POS) was found to influence prominence perception as well. Highest p-scores were assigned to content words, in particular nouns and adjectives. Fig.7 shows the mean prominence values for POS in the experiment, following the categorisation of the *Stuttgart Tübingen TagSet* (STTS; Schiller et al. 1999).

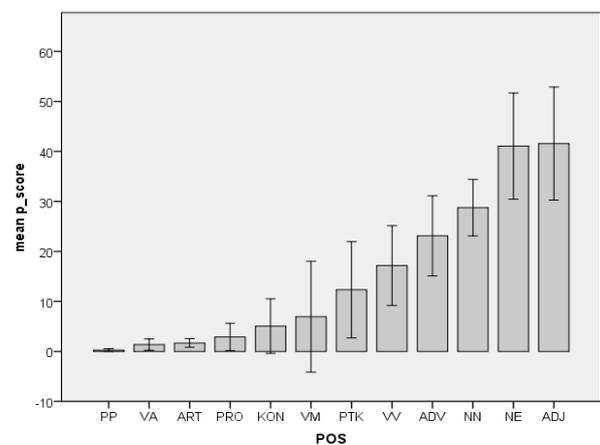


Figure 7: Prominence judgements as a function of part-of-speech.

However, the prominence of different POS was not judged independently of accentuation. Thus, if function words such as pronouns or particles (also modal verbs) carried a (nuclear) pitch accent in our data, this function word was perceived as particularly prominent (cf. Calhoun 2010a and discussion).

3.3.2. Word frequency

As to word frequency, results show the expected tendency for a negative correlation: the less frequent a word is (we used the *SubtLex* corpus as a reference) the more prominent it is perceived (Pearson's $r = -0.33^{**}$).

3.3.3. Morpho-syntactic focus marking

The test material contained words carrying phrase accents and nuclear accents which followed the focus markers *nur* ('only') and *sogar* ('even'). An example of a second occurrence focus marked by a phrase accent on *Bahber* is the construction *Auch eine Bachblütenkur kann nur Doktor Bahber machen* 'Also a cure with bach flowers can only be done by Dr. Bahber'. Both phrase accented and nuclear accented words were judged as more prominent if they were preceded by a focus particle, as illustrated in Fig.8.

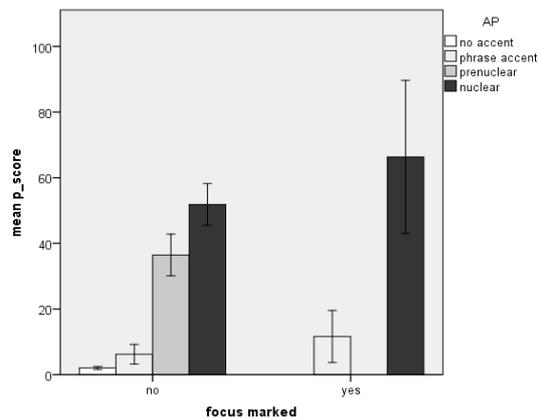


Figure 8: Prominence judgements of words following a focus particle as a function of accent position.

4. Discussion and conclusion

Our results suggest that untrained listeners base their judgements primarily on prosodic features, in particular F0 movement and alignment, encoded here in different accent types (rising accents being rated as most prominent). The position of an accent in the phrase also plays a role, as well as – to a small degree – duration and intensity, which are not fully independent of accent type and position, though (cf. Figs. 2 and 3). Secondary accents due to their prenuclear position are perceived as less prominent than nuclear accents, but their p-score crucially depends on the type of pitch movement – since both nuclear and prenuclear accents are fully-fledged pitch accents. In contrast, postnuclear secondary accents (phrase accents) are perceived as less prominent due to their lack of tonal movement. In general, bottom-up acoustic features seem to attract most attention in untrained listeners when judging prominence.

Nevertheless, prominence perception also is to some extent based on non-prosodic factors related to top-down knowledge, or rather intuition, about a word's accentability. This knowledge triggers language-specific expectations which are structural or conceptual in nature, e.g. expectations on the appropriate metrical and information structure, on the alignment of the nucleus with the last argument near the right edge of an intonation phrase in German, or on the correspondence between semantic weight and perceived prominence, e.g. in terms of POS. Our data only show few hints at expectation-matching prominence ratings, such as the general prominence-ranking of POS and a slight effect of focus particles. A stronger influence on the p-scores could be observed in cases of *mismatch* with a listener's expectations. Mismatches attract attention, demanding increased processing effort (Cole et al. 2010). They may either be triggered by the occurrence of unpredictable words (low frequency or non-derivable from the discourse) or by the difference between expected prominence (due to structural knowledge of an utterance's metrical structure including the position of the nuclear accent) and evaluated prominence. This difference affects the interpretation of an utterance's information structure (see Calhoun 2010a,b). A number of nuclear accents on function words in the test material which were perceived as highly prominent serve as evidence for this effect.

The results generally confirm the *dual nature of prosody* as reflecting the interdependency between acoustic patterns on the one hand and higher-level structures that account for these patterns on the other (Shattuck-Hufnagel and Turk 1996:196, Smith 2013:6). The relevance of both aspects for the

perception of prominence and phrasing has also been confirmed to a large extent by the RPT data by Cole et al. (2010) on American English. However, it is important to point out that the wording of the instructions may strongly influence the subjects' annotation of prosody. Smith (2013) showed in an RPT study on French that more prominences were marked under 'meaning-based' instructions than under 'acoustically-based' instructions. Our instructions were deliberately designed to leave both interpretations open, giving the options 'important' (meaning-based) next to 'stressed' and 'highlighted' (rather acoustically-based). Given the outcome of our study, the instructions may have been interpreted by most subjects as acoustically-based.

Finally, the data suggest that the main difference between trained and untrained listeners may be that the former attribute a special status to the nucleus (defined as the structurally most prominent and communicatively most important element in most common intonation theories), whereas the latter do not, at least not to the same extent. Instead, they determine the location of a prominent word according to their perception of prosodic, especially tonal, salience.

5. References

- Brysbaert, M., M. Buchmeier, M. Conrad, J. Bólte, and A. Böhl, (2011). "The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German." In: *Experimental Psychology* 58(5), pp. 412–424.
- Calhoun, S. (2010a). "The centrality of metrical structure in signaling information structure: A probabilistic perspective." In: *Language* 86, pp. 1–42.
- Calhoun, S. (2010b). "How does informativeness affect prosodic prominence?" In: *Language and Cognitive Processes*, 25 (7), pp. 1099–1140.
- Cole, J., Y. Mo, and M. Hasegawa-Johnson (2010). "Signal-based and expectation-based factors in the perception of prosodic prominence." In: *Laboratory Phonology* 1, pp. 425–452.
- Fleiss, J.L. (1971). "Measuring nominal scale agreement among many raters." In: *Psychological Bulletin* 76, pp. 378–382.
- Geertzen, J. (2012). "Inter-Rater Agreement with multiple raters and variables." Retrieved March 19, 2014, from <https://mlnl.net/jg/software/ira/>
- Grice, M., S. Baumann, and R. Benzmüller (2005). "German Intonation in Autosegmental-Metrical Phonology". In: S.-A. Jun [Ed.] *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, pp. 55–83.
- Grice, M., D.R. Ladd, and A. Arvaniti (2000). "On the Place of Phrase Accents in Intonational Phonology". In: *Phonology* 17 (2), pp. 143–185.
- Röhr, C., and S. Baumann (2010). "Prosodic Marking of Information Status in German." In: *Proceedings Speech Prosody 2010*, 100019:1–4.
- Schiller, A., S. Teufel, C. Stöckert, and C. Thielen (1999). "Guidelines für das Tagging deutscher Textcorpora mit STTS." Technischer Bericht, Universitäten Stuttgart und Tübingen. <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf>.
- Shattuck-Hufnagel, S., and A. Turk (1996). "A prosody tutorial for investigators of auditory sentence processing." In: *Journal of Psycholinguistic Research* 25, pp. 193–247.
- Smith, C. (2013). "French listeners' perceptions of prominence and phrasing are differentially affected by instruction set". In: *Proceedings of Meetings on Acoustics*, Vol. 19, 060191 (Acoustical Society of America [DOI: 10.1121/1.4799041], 21st International Congress on Acoustics, Montreal.)