TWO DIMENSIONS OF PROMINENCE

Petra Wagner   *Thomas Portele*

Institut für Kommunikationsforschung und Phonetik, Universität Bonn

ABSTRACT

Prosody fulfills a variety of functions in dialogues. Our study examines the relationship between different levels of perceived prominence of syllables and the linguistic and paralinguistic categories *accent* and *emphasis* which are conveyed prosodically. It is still unclear, how a notational system might look like that is able to capture the fine–grained differences between both. The notion of *perceptual prominence* — defined as a relational parameter on a scale between 0 and 31 — seems to be a useful phonetic measure to capture *both* the subtle differences and shared characteristics of the phenomena commonly referred to as *linguistic* and *paralinguistic*. Our data indicate that the overall level of prominence within an utterance reflects the level of emphasis, whereas the relative difference of prominences to each other distinguishes between different linguistic accent types.

1.   INTRODUCTION

One prosodic function relates to the expression of (meaning differentiating) *linguistic* categories like prosodic focus or word stress. Others are usually regarded to be closely connected to more *paralinguistic* properties of speech (e.g. conveying the current emotional state of the speaker or signalling different "speaker attitudes"). Various contextual settings (e.g. a piece of information may be "given" or "new") may interact with a speaker's decision regarding his/her opinion on the proper accentuation or deaccentuation of specific words or phrases. Furthermore, paradigmatic load might influence the prosodic pattern (e.g. in a correction or ordering statement). Another influence may stem from emotional involvement influencing speaking style. The future success of automatic dialogue systems involving speech recognition and synthesis (e.g. Wahlster [1993]) may in part depend upon a differentiation of those two levels of prosodic expression. The research presented in this paper is based on a database of short constructed American English dialogues.

2.   DEFAULT ACCENT, CONTRAST AND EMPHASIS

2.1.   'Default' vs. 'Non-default' accent

It has long been argued that there is a need for differentiating between a so-called "default accentuation pattern" and a "non-default accentuation pattern" in prosody, which seems to be at least relevant for most Germanic languages. The differences between both patterns are mostly explained by the former being the result of phonological rules or constraints, whereas the latter can only be explained when contextual issues often referred to as *givenness*, *topicality* or *narrow focus* are taken into account. Even though there has been considerable debate whether there are any phonological prosodic differences between the "default" and the "non-default" accentuation pattern this problem has not finally been solved — some researchers have even argued against any "default"–pattern at all and claimed that accentuation could not be explained on purely syntactic grounds (e.g. Fuchs [1984], Bolinger [1972]). Here, we take the view ague for in [Ladd, 1996, 160ff.] that the "default accentuation pattern" should be the one commonly referred to as conveying "broad focus" or "all new".

2.2.   Differences between 'contrastive' and 'normal' accent

Another dispute concerns the question of whether there are any differences between a "normal accent" and a "contrastive one". Recent results for Dutch indicate that contrastive accents are perceived to be more prominent, but only when they are presented within the utterance context (Krahmer and Swerts [1998]). Thus, contextual configurations seem to play a major role in determining prosodic patterns for specific accent types which cannot be successfully examined in isolation. This is further supported by the phenomenon often referred to as "deaccentuation" of contextually *given* or *presupposed* material.

2.3.   'Emphasis' vs. 'contrast'

The case of *contrastive stress* is closely connected to the combined effects of linguistic and paralinguistic functions on prosody in discourse. Some people have recognised that the terms *contrastive stress* and *emphasis* have been used interchangeably in the literature (see Trask [1996]), especially in those cases where the pragmatic load is high. For example, this is the case in utterances of correction, where the speaker's intention is not only to provide the discourse with new information but also to change what (s)he assumes to be the hearer's beliefs (Wagner [1999]). The disentangling of what is usually referred to as *categorical* vs. *paralinguistic* appears to be extremely difficult in such cases. Ladd illustrates this point by referring to the wide range of (para)linguistic functions ranging from repetition or contrast to surprise, all of which are according to hin related to a property [raised peak] of pitch accents.

> [raised peak] covers some of what has often been called "contrastive stress", but [...] the distinction involves

much more than mere logical contrast. [Ladd, 1983, 736]

Wide agreement exists regarding the issue that a paralinguistic property of speech like 'emphasis' should be represented on a gradient rather than a categorical scale. This attitude has its roots in Bolinger [1961]. Categorical distinctions should be reserved for purely linguistic functions. However, representations are not always categorical in phonology. Instead, numerical measures have been introduced into phonology via a notation known as *metrical grids* reflecting the relative prominence of syllables in a utterance in terms of column height. Metrical grids have been developed for a description of purely linguistic entities like word or sentence stress. On each relevant linguistic level, a metrical description imposes a categorical relative *weak* vs. *strong* relation onto two neighbouring linguistic entities. Metrical descriptions have already been used for descriptions of gradient measures. For Dogil [1979], they served as a methodological frame in a study of emphasis in English and Polish (Dogil [1979]). For some very special phenomena like 'correction contrast', it seems to be widely accepted that a typical accentuation pattern involves a process that can be described as postfocal deaccentuation on a metrical (Féry [1988], Dogil [1979]) and perceptual level (Wagner [1999]). Given the usefulness of metrical grids and their perceptual correlates — syllable prominences — for both categorical linguistic and gradient paralinguistic descriptions, they appear to be good candidates for solving the problem mentioned above.

2.4.   Questions

From these introductory considerations, several questions follow which are going to be addressed in the study described subsequently:

1. Are the different accentuation patterns reflected perceptually as prominence patterns?

2. Can the contextually bound perceptual difference between "default accents" and "contrastive accents" be modelled/represented adequately?

3. Is there a unitary representational model for capturing the interactions and differences of linguistic and paralinguistic (emphatic) prosodic properties of speech?

3.   MEASURING PERCEPTUAL PROMINENCE

3.1.   Labelling procedures

Our approach towards manual labelling of perceptual prominence is based on the work of Fant and Kruckenberg [1989]. They used prominence ratings on a scale between 0 and 30 for a study of Swedish prosody. The method showed high inter-labeller correlation and was successfully adapted for a prosodic database of German (Heuft et al. [1995]) and American English (Elsner et al. [1998]).

3.2.   Acoustic correlates of prominence

A major advantage of the prominence–based labelling schema is the nicely defined interface to acoustic phenomena. These have been extensively studied for Swedish (Fant et al. [1998]), German (Heuft et al. [1995]) and American English (Portele [1998]). The approach was successfully implemented in a German speech synthesis system (Portele and Heuft [1997]). Building upon acoustic as well as linguistic information, prominences can be predicted automatically and very reliably on the basis of a CART–tree (Portele [to appear]). The resulting prominence ratings showed a high correlation ($cc = 0.87$) between predicted and observed prominence. The possibility of an automatic labelling scheme is a further advantage of a prominence–based approach to the study of prosody.

4.   THE DATABASE

The prominence–labelled database under investigation consists of short dialogues spoken by one female and one male native speaker of American

English (Elsner et al. [1998]). The speakers were instructed to read lively and specific parts of the dialogues were printed in capitals as indication for the speakers to emphasize those parts of the dialogue. An emotional coloring of several utterances was also introduced via emotion conveying verbs or contexts. These, however, where left out of the study presented here.

4.1.   Default vs. Non–default Accents

The database has been examined as to whether the prominence patterns provide any answers to the questions formulated in section 2.4. Since the aim of this study was to analyse the interaction of different accentuation patterns and prominence patterns, the first step involved a labelling of contexts matching a default accentuation pattern and those matching a non–default one. The non–default cases were either those where the context would lead to an accent on a syllable not expected in an "all new"–environment, or where a word or phrase was marked for emphasis in the text to be read aloud by the speakers which also would not have been accented in the default case.

Example 1: Default accent in question

A: Where is that BUtton?
B: It's on the left hand side of the panel.

Example 2: Non–default accent in question

A: Press the orange button!
B: Where IS that button?

The default accents were labelled as [+default], the non–default ones consequently as [-default]. In some cases, the speakers failed to emphasize the intended words as expected. Those cases were not included in the following study.

4.2.   Emphatic Accents

Furthermore, for the male speaker of the database those phrases spoken emphatically according to the orthographic instruction

were labelled as [+ emphatic], independent of its being a *default* or *non–default* accent. Due to time restrictions, examinations of the female speaker have not been finished and cannot be reported but are currently in progress.

4.3. Further Issues

Because of the labelling procedure explained above, there is a considerable overlap between [+emphatic,-default] cases.

Further automatic annotations were made regarding whether a syllable was uttered before or after the maximal prominence in the prosodic phrase. Several utterance configurations and their related prosodic features were to be tested and abbreviations are used in order to refer to them. To follow the statistical analyses presented below, it is essential to thoroughly understand the descriptions given in Table 1.

maxprom	the syllable which has been assigned the highest prominence value within an utterance is called *maxprom syllable*. The prominence value of that syllable is referred to as *maxprom*.
mean prominence	the mean prominence within an utterance not taking into account the maxprom syllable
premax	the mean prominence value of all syllables preceding *maxprom* within an utterance. A *premax syllable* is a syllable preceding maxprom.
postmax	the mean prominence value of all syllables following *maxprom* within an utterance. A *postmax syllable* is a syllable following the maxprom syllable.

Table 1: The prosodic features that were tested

5.   STATISTICAL ANALYSES

Three working hypotheses were tested:

- There is a correlation between the maxprom value and the mean prominence perceived within the same utterance. The expected relationship involves some kind of deaccentuation resulting in lower mean prominence.
- The type of accentuation pattern [±default] somehow influences the perceived prominence pattern in terms of strength of maximal prominence or its surroundings.
- The dimension of "emphasis" has an influence on the pattern of perceived prominence.

5.1. Maximal Prominence and Deaccentuation

There seemed to be no correlation at all between the perceived strength of the maxprom syllable and the mean prominence within an utterance. Thus, our data indicate that given a high maxprom value, no deaccentuation has to be present, at least not on a level of perception. Knowing this, it was tested whether some correlation between the maxprom syllable and its average difference to the mean prominence could be detected. Taking

into account this average difference, we found a high correlation (male speaker: $cc = 0.807$, $p < 0.0001$; female speaker: $cc = 0.76$, $p < 0.0001$) between both measures. This can be read as "The greater the prominence value of the syllable perceived as maximally prominent, the greater the difference to the mean syllable prominence in the remaining utterance". Apparently, mean prominence does not vary depending on the value of the maxprom syllable. This indicates that the strength of the maxprom syllable does not affect the perception of the surrounding ones. They are in fact perceived as quite stable. Instead, single syllables can obviously be marked as extremely prominent by the speaker. Since the difference to the maxprom syllable appeared to be a good indicator for the relative prominence pattern in an utterance, it was used as a measure in the subsequent study. It was
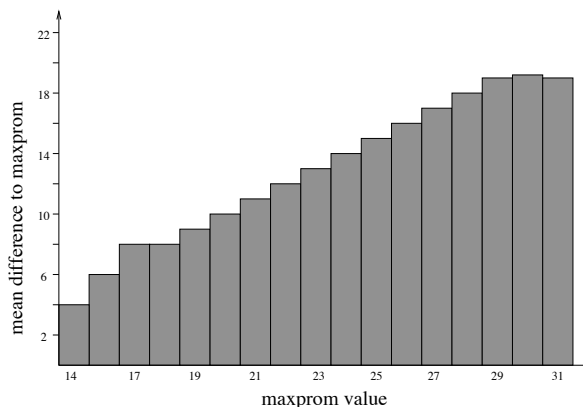
Figure 1: Correlation between maximal prominence and mean difference to contextual ones

further taken into account whether premax syllables are affected in a different way by the maxprom value than postmax syllables. There is an overall tendency for premax to have a larger difference to maxprom than for postmax (t–test, $p < 0.0001$). This may indicate a kind of "inertia" of prominence perception or reflect the general tendency of Germanic languages that the main stress is expected towards the end of an utterance.

5.2. Default vs. Non–default accents

In a next step, the [+default] accentuation patterns were compared with the [-default] ones regarding their prominence patterns and relationship to the maxprom syllable.

Here, no significant differences could be found regarding the strength of maxprom or the premax or postmax values. The mean differences to maxprom were almost identical (see Figure 2) comparing [+default] with [-default] contexts.

Differences between the [+default] and [-default] accentuation patterns showed up when the premax and the postmax syllables were regarded in isolation concerning their mean difference to maxprom. It could be detected that the mean difference to maxprom in premax contexts is significantly higher in [-default] accentuation contexts (male speaker: t–test, $p < 0.0001$; female speaker: (t–test, $p < 0.01$)). Thus, there is a tendency for prominences to be perceived less intense prior to [-default] accents (see Figure 3). In postmax contexts, however, the picture changes completely. Here, the mean difference to maxprom is higher in
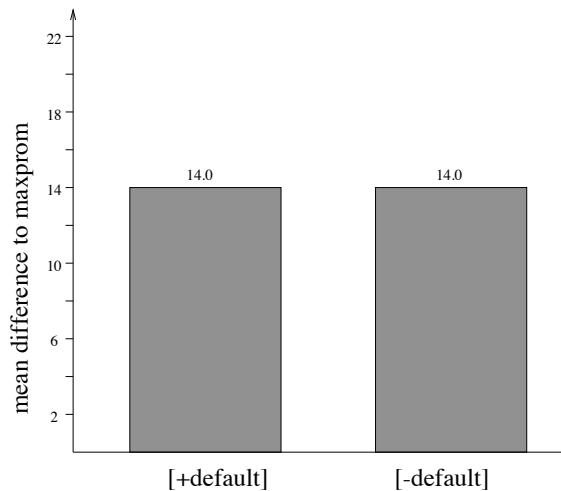
Figure 2: Differences from mean prominence to maxprom in default and non–default accent patterns

[+default] utterances (male speaker: t–test, $p < 0.05$; female speaker: (t–test, $p < 0.05$)), which means that here [-default] contexts appear to be perceived as more prominent relative to maxprom (see Figure 4).
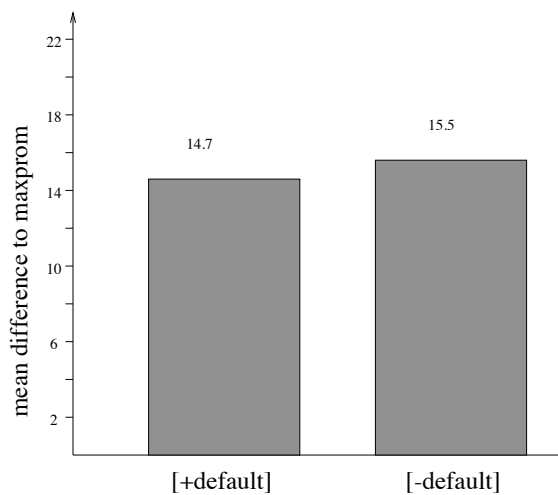
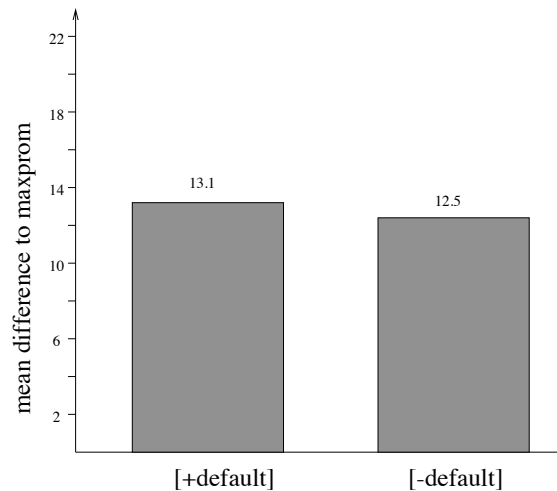**Figure 3:** Difference to maxprom in premax contexts

5.3. Emphatic Contexts

When comparing [+default] vs. [-default] contexts, no significant distinction was detected concerning the difference from the mean prominence to the maxprom value (cf. Figure 2). However, different patterns were isolated taking into account the premax and postmax contexts (cf. Figures 3 and 4).
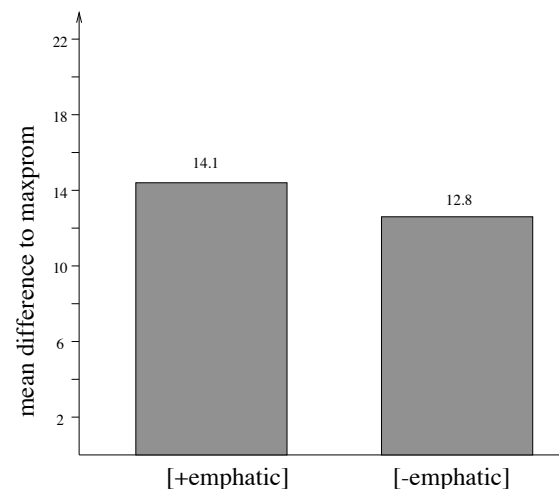
In contrast, in [+emphatic] contexts the mean prominence was closer to maxprom (t–test, $p < 0.0001$) than in the [-emphatic] contexts. This tendency remained stable independent of premax or postmax contexts. Maxprom itself, however, could not be shown to be significantly higher than in [-emphatic] contexts. This indicates that in emphatic contexts, there is a general in-

Figure 4: Difference to maxprom in postmax contexts

crease in prominence but no specific effect on maxprom. Due to the considerable overlap in the data between [-default] and [+emphatic] utterances analyses were repeated, leaving out the [-default] accent patterns. The previously observed tendencies remained present.

The picture becomes more complex when comparing the [-default] cases with the [+default] ones within emphatic environments. It can be shown that maxprom is significantly higher (t–test, $p < 0.0001$) *and* that the difference from the mean prominence to maxprom is higher (t–test, $p < 0.0001$) in the [-default] cases. This effect is identical to the premax pattern we found before, when emphasis was not taken into account explicitly. The postmax effect of being closer to maxprom, however, vanishes. Apparently, given an overall tendency to a somewhat higher prominence due to emphasis, the [-default] accentuation pattern requires a prominence increase on the maxprom syllable in order to retain its perceptual dominance compared to the preceding (and here also succeeding) context.

Figure 5: Difference to maxprom in emphatic and non–emphatic contexts

5.4. Discussion

First, it could be shown by the data that a maxprom syllable can be salient while the mean prominence value does not vary. It is very similar to mean prominence in utterances where the maxprom value is less strong. This was indicated by the high correlation between maxprom syllable and mean difference to the prominence of the contextual syllables: with increasing maxprom value, the difference to the mean prominence also increased. Besides, a general tendency for all utterances was that postmax syllables are perceived as more prominent than premax ones relative to maxprom. This tendency was confirmed by a somewhat lower prominence difference to maxprom in postmax syllables. Besides, this effect was evident throughout the database — for the [-default] and [+emphatic] cases as well. There are several potential explanations for this phenomenon — one is the possible interaction between final lengthening and utterance final perception of somewhat higher prominence. It could also be an effect of the so–called *Nuclear Stress Rule* (NSR), tracing back to Chomsky and Halle [1968]. This rule claims that in English there is a tendency for the main stress to be located near the right phrase boundary. At least our data indicate that people are more likely to perceive higher prominences towards the end. Such an interaction with the NSR may also account for the fact that in [-default] cases, this effect was even stronger. Since in the [-default] contexts, the element which ought to receive the primary accent in the default case *follows* the maxprom syllable, there still seems to be a tendency towards perceiving the rest of the phrase as rather salient. This outcome also goes hand in hand with a perception experiment based on synthetic speech described in Wolters and Wagner [1998], where subjects were able to perceive utterance final narrow focus much more reliably than utterance initial focus.

But the smaller prominence difference to maxprom *after* the maxprom syllable appears to be not the only indicator for a [-default] accentuation pattern. The examination of the premax contexts showed that prominences differ more markedly from the maxprom value when they appear in [-default] accentuation patterns. Therefore, the prominence–based analysis could isolate two different patterns: Given a [+default] accentuation, a specific difference to the maxprom value is perceived preceding the maxprom syllable and this difference is lower succeeding it. But the main indicator for a [-default] accentuation pattern seems to be the rather low premax value, since this effect remains present even in [+emphatic] contexts while the postmax effect dissapeared.

Emphatic contexts did not exhibit a single extremely high maxprom value. Rather, they could be characterized by a general tendency of higher mean prominence. This indicates that emphatic utterances are being perceived as somewhat more prominent. The [-default] effect of a larger premax difference to maxprom is preserved by a higher maxprom value. [+emphatic, -default] contexts are the only cases, where the maxprom values were significantly higher.

6. CONCLUSION

Due to the limited amount of data and the fact that emphasis was only examined for one speaker, any general conclusions should be avoided. However, it appears to be the case that the perceptual measure of prominence is useful for describing both *linguistic*
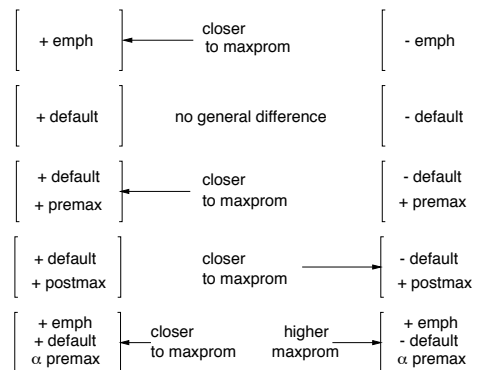

Figure 6: Overview of results

and *paralinguistic* phenomena. First results indicate a possibility to separate the domain of *emphasis* from the domain of *linguistic accentuation*.

'Emphasis' is the overall level of prominence of a stretch of speech relative to its environment. This could be shown by the following facts. The salience of a singular syllable does not generally affect the prominence perception within the remaining utterance (cf. Figure 1). Also, with specific words or phrases marked for emphasis, prominence was perceived to be higher throughout the entire utterance. Speakers appear to be able to emphasize single syllables or words or phrases without necessarily affecting the perceived prominence of the context. However, our male speaker also exhibited the tendency to use emphasis as a property of a larger stretch of speech. Apparently, emphasis can be deliberately added to any stretch of speech in varying degrees. This finding corresponds to the commonly held view of emphasis being a relative property of speech rather than a categorical one. In our view, the possibility of any stretch of speech to be marked paradigmatically as more or less prominent without necessarily interfering with the contextual prominences is the *first* dimension of prominence.

A prominence–related phenomenon of a different kind was detected in the comparison of the [+default] vs. the [-default] accentuation patterns. Here, further evidence was retrieved that linguistic categories such as *default accent* vs. *non–default accent* (or *contrastive accent*) do no differ in terms of their absolute prominence. This finding holds for maxprom syllables and mean prominence values in an utterance. This supports the view that non–default accents do not appear more prominent in isolation (Krahmer and Swerts [1998]). But there are indeed differences between the two configurations. Even though the difference to maxprom stays identical for the mean prominence, distinctions could be detected looking at the premax and postmax contexts. These subtle prominence relationships within the phrase appear to differentiate between the type of accent: speakers utter syllables keeping a relatively high difference to the maxprom value prior to it thus indicating the presence of a [-default] accent. Therefore, maxprom values are higher relative to the preceding part of the utterance. This explains that even though [-default] accents are not perceived as more prominent on an absolute scale, they appear different when presented with context. But as long as these relationships are intact, there seems to be no need to make the maxprom syllable more salient — perhaps due to economical reasons — not even in [+emphatic] utterances. This syntagmatic relationship between prominences is the *second* dimension

of prominence. However, maxprom may vary under the influence of emphasis in order to sustain those prominence relations which have a categorical function (i.e. the ones characterizing the [-default] accentuation). In this case — where an emphatic environment results in an overall higher prominence — there is the need to adjust the pattern of [-default] accents by making the maxprom syllable more salient.

Prominence–based descriptions may shed light on the difficult interaction between linguistic and paralinguistic functions of prominence. Both the distinctions and interactions between both levels of speech can be followed by one easily obtainable measure which also provides us with a well–defined interface to acoustic representations.

References

D.L. Bolinger. Generality, gradience, and the all-or-none. *Janua Linguarum*, 1—46, 1961.

D.L. Bolinger. Accent is predictable (if you're a mind reader). *Language*, 48:633—644, 1972.

Noam Chomsky and Morris Halle. *The Sound Pattern of English*. New York: Harper & Row, 1968.

Grzegorz Dogil. *Autosegmental Account of Phonological Emphasis*, volume 7 of *Slavic Linguistics*. Edmonton, Canada, 1979.

Anja Elsner, Maria Wolters, Thomas Portele, Monika Rauth, and Gerit Sonntag. Designing and labelling a prosodic database for American English. In *Proceedings on the Workshop for Language Resources, Granada*, pages 1079—1082, 1998.

G. Fant, A. Kruckenberg, and J. Liljencrants. Acoustical correlates of perceived prominence. A study of Swedish prosody. preliminary manuscript, November 1998.

Gunnar Fant and Anita Kruckenberg. Preliminaries to the study of Swedish prose reading and reading style. *STL–QPSR 2/1989*, pages 1—68, 1989.

Caroline Féry. Rhythmische und tonale Struktur der Intonationsphrase. In H. Altmann, editor, *Intonationsforschung*, pages 41—64. Niemeyer, 1988.

Anna Fuchs. 'Deaccenting' and 'Default accent'. In Dafydd Gibbon and Helmut Richter, editors, *Intonation, Accent and Rhythm. Studies in Discourse Phonology*, volume 8 of *Research in Text Theory*, pages 134—164. The Hague: de Gruyter, 1984.

B. Heuft, T. Portele, F. Höfer, J. Krämer, H. Meyer, M. Rauth, and G. Sonntag. Parametric description of f0–contours in a prosodic database. In *Proceedings of the International Conference on Spoken Language Processing, Stockholm*, volume 2, pages 378—381, 1995.

Emiel Krahmer and Marc Swerts. Reconciling two competing views on contrastiveness. In *Proceedings of ICSLP 98, Sydney*, 1998.

D.R. Ladd. Phonological features of intonational peaks. *Language*, 59:721—759, 1983.

D.R. Ladd. *Intonational Phonology*. Number 79 in Cambridge Studies in Linguistics. London: Cambridge University Press, 1996.

Thomas Portele. Perceived prominence and acoustic parameters in American English. In *Proceeding of ICSLP 98, Sydney*, 1998.

Thomas Portele. Just concatenation — a corpus–based approach and its limits. In *Selected Papers of the 3rd Speech Synthesis Workshop at Jenolan Caves*. (Working Title), to appear.

Thomas Portele and Barbara Heuft. Towards a prominence–based synthesis system. *Speech Communication*, pages 61—72, 1997.

R.L. Trask. *A dictionary of phonetics and phonology*. Routledge: London and New York, 1996.

Petra Wagner. The synthesis of German contrastive focus. In *Proceedings of ICPhS 99, San Francisco*, 1999.

Wolfgang Wahlster. Verbmobil: Übersetzung von Verhandlungsdialogen. Technical report, DFKI GmbH, 1993.

Maria Wolters and Petra Wagner. Focus perception and prominence. In W. Hess B. Schröder, W. Lenders and T. Portele, editors, *Computers, Linguistics, and Phonetics between Language and Speech. Proceedings of the 4th Conference on Natural Language Processing KONVENS-98*, volume 1 of *Computer Studies in Language and Speech*, pages 227–236. Peter Lang, 1998.