Alignment of Medial and Late Peaks in German Spontaneous Speech

Oliver Niebuhr[1] *& Gilbert Ambrazaitis*[2]

[1]IPdS, University of Kiel, [2]Centre for Languages and Literature, Lund University
on@ipds.uni-kiel.de, Gilbert.Ambrazaitis@ling.lu.se

Abstract

Starting from a corpus of German spontaneous speech, the phonetic realisations of the two KIM categories medial and late peak were investigated in prenuclear position. The results show that, for both categories, the onset of the rising F0 movement (L) is comparably aligned around the accented-syllable onset, whereas the F0 maximum (H) is independently aligned and predominantly located before the accented-syllable offset or after the onset of the following unaccented syllable, respectively. The data further suggest that also from the AM point of view the two prenuclear rises are different at the phonological level. Finally, the possibility is pointed out that the alignment patterns found for prenuclear rises in other studies are to some extent due to a combination of categories like the medial and late peak.

1. Introduction

The alignment of F0 peaks with the segmental string has been under investigation for a variety of languages, within a variety of theoretical frameworks and under a variety of hypotheses, e.g. [1, 2, 3]. In the last decade, a number of studies within the autosegmental-metrical (AM) framework have been concerned with measurements of F0 minima and maxima in so-called bitonal pitch accents, consisting of two tonal targets, L and H, cf. [4, 5] for recent examples. The results of these studies are relevant for evaluating phonological models of intonation. A consistent alignment of tonal targets with segmental landmarks, resulting in contextually determined contour shapes, is sometimes regarded as evidence against a contour-based model. Furthermore, within the AM framework the results are used to model the association of tonal targets with the segmental string: Which tone in an accent is the starred one? Is a given accent bitonal at all? Are the tones in a bitonal accent aligned independently of each other; or is only the starred tone aligned with the segmental string, while the other one leads or trails the starred tone at a constant temporal interval? Pierrehumbert originally suggested the trailing/ leading tone approach [6].

An independent alignment, however, is supported by a number of recent studies on a variety of languages concentrating on prenuclear L+H accents. Studies on Greek, Dutch, and English commonly found that the L target appears to be aligned at the onset of the accented syllable, e.g. [3, 7, 8]. The H target appears to be aligned independently, most often at the onset of the postaccent vowel, but the studies have yielded variable results. In a study on German [4], the L (and in part the H, too) were found to be aligned significantly later than in Greek and English, with southern German being even later than northern German. Atterer and Ladd argue that the investigated L+H accents in the relevant studies on Greek, Dutch, English, and German are instances of the same phonological category – a *rising prenuclear accent*. The detailed differences in alignment across languages or dialects, then, are due to different phonetic realisations of the same phonological category. A possible criticism, also discussed by Atterer and Ladd themselves, is that their approach does not account for language-specific alignment contrasts, i.e. for phonologically different prenuclear rises. However, Atterer and Ladd find support for their claim in the fact that the experimental conditions in the relevant studies were equivalent, and the test sentences were comparable. They also take into consideration that there are no phonological contrasts within prenuclear rises (at least for the languages compared).

We take up both the possible criticism and the supporting argument here and suggest that indeed a common shortcoming in this kind of studies lies in focussing only on the connection of phonological form (L+H) and phonetic content (the alignment in ms). Functional aspects have to be taken into account, too. First, these may differ across languages. Secondly, functionally different rising accents may exist in one language. For example, the medial and the late peak in German form a well-established functional and phonological distinction, which is integrated in the Kiel Intonation Model (KIM) [9]. The medial peak can be regarded as a default pattern for conveying new information, while the late peak additionally evaluates the information as unexpected. Both patterns include a rising movement linked with the accented syllable.

The similar experimental conditions – namely reading lists of non-related, pragmatically rare sentences – in the alignment studies cited by Atter and Ladd may indeed have elicited the same phonological category across the studies and languages, including German. It is, however, not clear whether this prenuclear rise corresponds to a late or a medial peak. On the one hand, one would expect laboratory speech to lack any affective characteristics. Considering the oddity of the experimental situation and the test sentences, on the other hand, it is not self-evident that the subjects have used the default pattern, i.e. the medial peak. Finally, it could even be the case that both patterns have occurred, but merged in the analysis into the single category of the prenuclear rise. However, without any knowledge on how the prenuclear rise relates to the well-established medial–late contrast in German, it is difficult to evaluate Atterer and Ladd's interpretations.

Due to the theoretical background of KIM, the medial-late contrast and its acoustic correspondence was primarily founded on functional analyses and perception experiments. Acoustic analyses, working out the phonetic realisations of these categories were not given special emphasis, although, of course, there are some studies dealing with this issue, e.g. [10].

It was therefore the aim of our study to gain further insights into the phonetic properties of the medial and late peaks in prenuclear position, and to relate these insights to the findings for the prenuclear rises in German and other languages. Regarding the functions of the two peaks, we expected that particularly the late peak can best be observed in natural conversation. Consequently, our analysis is based on a corpus of (Standard) German spontaneous speech. It has been prosodically labeled using the PROLAB system [11], which is based on KIM. The decisions concerning the choice of labels have been made primarily perceptually-based and discussed by at least two labellers. Considering that peak position and shape

as well as intensity and duration of the underlying segments were found to be involved in the perception of German medial and late peaks (ongoing research by the first author), the alignment patterns appearing in our analysis cannot be regarded as an artifact of the labeling process.

2. Method

Our study is based on a (Standard) German spontaneous speech corpus, referred to as the *Lindenstrasse* corpus. It was elicited using the scenario described in [12]. In this, pairs of subjects were selected due to two criteria: (1) they are friends in private life and (2) they fancy the most famous soap opera on German television: "*Die Lindenstrasse*". The subjects were seated in separate rooms and each was presented a 15 minute video, a manipulated episode based on single scenes from earlier episodes. The two videos were generally similar but differed slightly concerning the selection or ordering of scenes and deliberately integrated interferences such as short scenes from completely different broadcasts. After having watched the video twice, the subjects were connected via headset and instructed to discuss the differences in content between their respective versions of the episode. The whole corpus contains six such dialogues, i.e. 12 speakers, and has a total length of 80 minutes or 13.000 words [13]. The main advantage of the scenario and the corpus is that they provide highly natural spontaneous speech, viz. everyday conversation on a common topic between friends.

The disadvantage of spontaneous speech in general is, of course, the lack of experimental control. However, with a sufficiently large corpus, it is possible to select tokens by adequate criteria and thereby control the material for a given study. In order to design a database of medial and late peaks that would be as comparable as possible to the common denominator of the databases in the most recent alignment studies (e.g. [4]), the following criteria were defined for the selection of medial and late peaks from the *Lindenstrasse* corpus:

- The token is labelled either as a medial or as a late peak.
- The token is a prenuclear rise, i.e. it is not the last pitch accent of a prosodic phrase (no distinction between intermediate and intonational phrases is made in KIM or PROLAB).
- The direct segmental environment of the rise is *phonetically* voiced, i.e. even phonologically voiceless, but assimilated voiced segments are allowed. The direct environment is defined as starting at one segment before the onset of the accented syllable, and ending at the vowel offset in the postaccent syllable.
- The selected medial and late peaks are concatenated with the following accents by an indentation in the F0 course.
- In order to reduce influences of adjacent pitch accents, there is at least one (and indeed more than two in most cases) unaccented syllable between the token and any pre- or succeeding pitch accents.

For the selected tokens, the points in time and the F0 values for the starting and ending points of the rise (L and H) were measured manually in the F0 curve. As in the most recent alignment studies, no attempt was made to consider any micro-intonational effects. The points in time of the segmental onsets were obtained from the existing segmental labelling for the following segments: the vowels of the accented syllable and the postaccent syllable: **V0** and **V1**; all intervening consonants (whether belonging to the accented syllable or the postaccent syllable); and all initial consonants (if any) of the accented syllable. Those segment onsets that coincided with a syllable boundary where labelled **S0** (accented syllable), or **S1** (postaccent syllable), respectively.

3. Results

For the prosodic environments defined, 13 medial and 18 late peaks were found in the *Lindenstrasse* corpus (see http://www.ipds.uni-kiel.de/on/onga.html for the references to the 31 tokens of the two samples and a link to the *Lindenstrasse* corpus). Both samples comprise open and closed accented syllables with either phonologically long or short vowels. Furthermore, each sample contains syllables with the accented vowel in initial position (i.e. without a consonantal syllable onset). So, with regard to possible effects of syllabic factors, the sets of measurements taken from the samples were approximately counterbalanced. Moreover, an impressionistic analysis shows a considerable variation of speech tempo in each sample. This variation was acoustically estimated by the speaking rate in syllables per second (syl/s), based on the whole prosodic phrase in which the peak contour was found. Syllables were counted with regard to the canonic representation of the word forms in the phrase. It turned out that the two samples show a comparable variation in speaking rate, ranging from 5.6 to 8.9syl/s in the medial peak sample and from 5.0 to 8.8syl/s in the late peak sample. Considering the small sample sizes, a non-parametric *U* test was performed revealing that the speaking rates of the phrases of the two samples do not differ significantly ($z_{[0.05]}$= 1.96>0.58; p=0.562). Hence, potential influences of this factor can be regarded as controlled in the following contrastive description of the L and H alignment in medial and late peaks.

Figures 1 and 2 display the distributions of L alignment of the two peak categories relative to common segmental landmarks, i.e. the onset of the accented syllable (S0) and its vowel (V0). The corresponding means and standard deviations are given in Table 1. Both figures show a substantial overlap in the L alignments of medial and late peaks relative to S0 and V0. Except for one L in the medial peak sample, all L were found closely around the accented syllable onset (Fig. 1) before the syllable nucleus (Fig. 2). According to Table 1, the mean distance between L and the onset segment is 15ms in the medial peak and only 7ms in the late peak sample. A *U* test yielded no significant difference between the L alignment of medial and late peaks relative to both segmental landmarks, S0 ($z_{[0.05]}$=1.96>1.26; p=0.207) and V0 ($z_{[0.05]}$=1.96>1.92; p=0.055). However, the large standard deviations as well as the distributions in Figure 1 and 2 point to a certain variability, which is more pronounced for the L of the late peak (compared with medial peaks the measurements of late peaks generally showed a higher variability, cf. Tab. 1).

		L to S0	L to V0	H to S1	H to V1
medial	x (ms)	15	-29	-32	-68
peak	s (ms)	20	18	33	45
late	x (ms)	7	-54	21	-21
peak	s (ms)	37	39	54	43
		LH dur (ms)		LH range (st)	
medial	x (ms)	103		5	
peak	s (ms)	35		3	
late	x (ms)	183		6	
Peak	s (ms)	66		2	

Table 1: *Means (x) and standard deviations (s). At the top: the L alignment relative to the accented syllable onset (S0) and to its vowel onset (V0) as well as for the H alignment relative to the onset of the postaccent syllable (S1) and to its vowel onset (V1). At the bottom: LH rise duration and range of the rise (in semitones, st).*

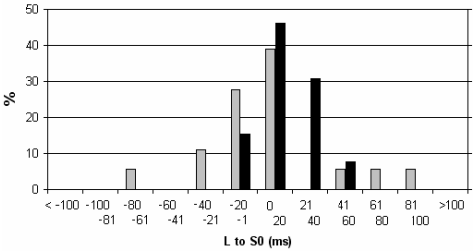

Figure 1: *L alignment (in %) for medial peaks (dark bars, n=13) and late peaks (light bars, n=18) relative to the accented syllable onset (S0), divided into classes of 20ms.*

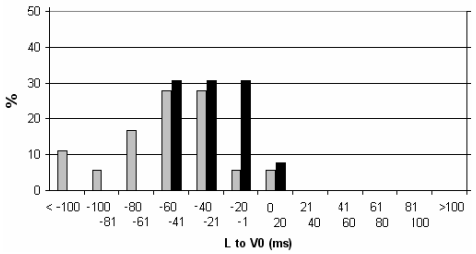

Figure 2: *L alignment (in %) for medial peaks (dark bars, n=13) and late peaks (light bars, n=18) relative to the accented vowel onset (V0), divided into classes of 20ms.*

While the L alignment is comparable for medial and late peaks, the two peak categories differ significantly in the duration of the rising LH movement ($z_{[0.05]}$=1.96<3.40; p<0.001). Table 1 shows a mean duration of 103ms for the medial peak rises, as against a mean duration of 183ms for the rises of late peak. Despite the difference in rise duration, the range of the rise is comparable for the two peak categories, showing values of about 5-6 semitones (cf. Tab. 1). A *U* test yielded no significant differences between the range of medial and late peak rises ($z_{[0.05]}$=1.96>0.94; p=0.347). Following from these results, we also found no significant correlation between the duration and the range of the rise in the medial and late peak sample (medial peak: r=0.454, df=11, p=0.12; late peak r=0.011, df=16, p=0.96).

Moreover, the comparable L alignment of medial and late peaks, combined with a significantly different rise duration, resulted in considerable deviations of the H alignment. Figures 3 and 4 show the alignment of H relative to the onset of the postaccent syllable (S1) or to the vowel onset of the postaccent syllable (V1), respectively. It can be seen that almost all H of the medial peaks were found within the accented syllable, whereas the H of late peaks are predominantely aligned after the onset of the postaccent syllable (S1, cf. the means in Tab. 1) and occurred in many cases even after the vowel onset of the postaccent syllable (V1). Accordingly, *U* tests revealed a significantly different H alignment for medial and late peaks relative to both segmental landmarks (H to S1:$z_{[0.05]}$=1.96<2.52; p=0.012 and H to V1: $z_{[0.05]}$=1.96<2.84; p=0.005). Furthermore, we found that the duration of the LH rise is positively correlated with the duration of the accented syllable (S1-S0) and with the interval from the accented-syllable onset to the vowel onset of the postaccent syllable (V1-S0). For medial peaks, the correlation coefficients are r=0.622 (df=11; p=0.023) in case of S1-S0 and r=0.724 (df=11; p=0.005) in case of V1-S0. The corresponding correlation coefficients for late peaks are r=0.572 (df=16; p=0.013) and r=0.673 (df=16; p=0.002).

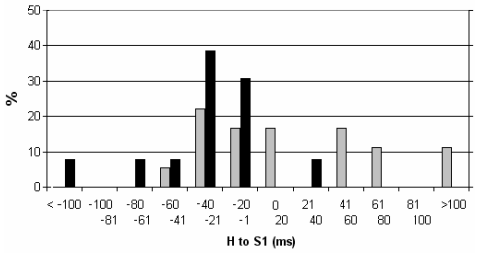

Figure 3: *H alignment (in %) for medial peaks (dark bars, n=13) and late peaks (light bars, n=18) relative to the onset of the postaccent syllable (S1), divided into classes of 20ms.*

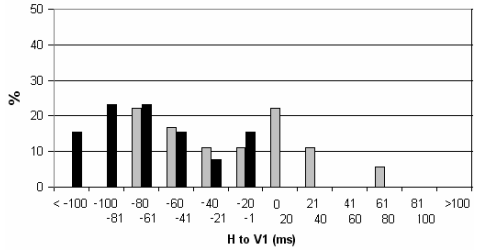

Figure 4: *H alignment (in %) for medial peaks (dark bars, n=13) and late peaks (light bars, n=18) relative to the vowel onset of the postaccent syllable (V1), divided into classes of 20ms.*

4. Discussion

In summary, a different phonetic realisation of the two peak categories was found in the present acoustic analysis. Starting from a comparable alignment of the onset of the rise (L), the two peak categories differ significantly in the duration of the rise and, correspondingly, also in the alignment of the F0 maximum (H). For the medial peak, the H was mainly reached immediately before the end of the accented syllable, whereas for the late peak, H was predominantely located in the following syllable or its vowel, respectively.

Our results therefore clearly show that prenuclear rises cannot be treated equally. Instead, there are functional differences within prenuclear rises, like the ones expressed by the medial and the late peak, which systematically affect the alignment of structurally outstanding F0 points. In the present case, this concerns the F0 maximum representing H.

In KIM, the medial and the late peak represent a phonological distinction. By contrast, Atterer and Ladd [4] suggest that there are no phonological differences within prenuclear rises. In this connection, it has to be pointed out that our data support the phonological distinction between the two prenuclear rises also from an AM point of view. This correspondence is not self-evident, since the phonological distinctions in KIM are based on completely different criteria.

On the one hand, it has to be pointed out that the alignment difference found here is a statistical difference, i.e. Figures 3 and 4 show a considerable overlap in the H alignment of medial and late peaks relative to the offset of the accented syllable (S1) and to the onset of the following unaccented vowel (V1). On the other hand, however, it is possible that this overlap is mainly due to the various syllabic structures involved in the analysis. For instance, it was mentioned before that the accented syllable as well as the following syllable contained either phonologically long or short vowels (cf. [8]). Furthermore, the number of intervening consonants between the accented and the following unaccented vowel was variable in both samples. Simultaneously, the results show a high and significant correlation between the duration of the rising movement and the duration of certain segmental

strings. On this basis, our data can be interpreted in the way that H shows a different association in the medial peak and the late peak sample. Correspondingly, the two prenuclear rises are phonologically distinct. This means that the class of prenuclear rises is, at least for (Standard) German, not only phonetically and functionally, but also phonologically inhomogeneous.

As regards the nature of this phonological distinction, the L and H of both prenuclear rises seem to be independently aligned with the segmental string. This is indicated by the positive correlation between the durations of the LH rise on the one hand, and the durations of the accented syllable or of the interval from the accented-syllable onset to the onset of the following unaccented vowel, on the other. Since the range of the rise was found to be comparable for the two prenuclear rises, an independent alignment means that the slope of the rise is determined by the syllabic structure. An independent alignment if L and H is consistent with the findings for prenuclear rises in other studies and languages (cf. Introduction). It is at the same time incompatible with the trailing/ leading tone approach by Pierrehumbert [6]. Hence, our results give further rise to reconsider this traditional phonological structure of pitch accents within the AM framework (cf. [14]). At this point, it must also be pointed out that the results of our analysis are compatible with KIM, although it is a contour model. The contour approach in KIM means in the first place that the intonational categories are perceptually defined and thought of as holistic patterns. In this, it is not a prerequisite for the F0 movements of intonational categories to show a specific slope or to remain at a constant slope, cf. [7].

It is possible that the alignment patterns received in studies dealing with prenuclear rises are to some extent due to a combination of categories like the medial and late peak. For instance, it can be seen in the data of some studies that the alignment of H relative to S1 and/or V1 shows a greater variation within and across speakers than the alignment of L relative to S0. Moreover, regarding the data of different studies and languages shows that L generally occurs around the beginning of the accented-syllable onset, whereas the alignment of H is much more variable and difficult to relate to a single segmental landmark. The latter led to approaches in which H is linked to articulatory events.

Therefore, to ensure that the data set is homogeneous, studies dealing with alignment should start from an approach which considers the functions of the intonation patterns within the language investigated and include these considerations in the survey. In this connection, it seems problematic, e.g., to use reading lists of isolated sentences, since such sentences are not restricted to a single functional interpretation and can thus be produced by the subjects with a great diversity of intonation patterns.

Finally, the alignment patterns found for the medial and late peak in the present analysis are slightly deviating from the ones found for similar syllabic environments in a previous analysis of the two KIM categories by Gartenberg and Panzlaff-Reuter [10]. While the alignment pattern of the H is comparable to the present study, i.e. the F0 maxima in [10] were either located within the vowel of the accented syllable or in the preceding syllable, L in [10] also shows a considerable alignment difference. This difference is due to the L of late peaks, which was in all cases aligned after the accented-vowel onset, whereas, as in the present study, the L of medial peaks is found around the onset of the accented syllable.

There are several possibilities to explain the different position of the L of late peaks in the present study and the study of Gartenberg and Panzlaff-Reuter. First, it has to be considered that Gartenberg and Panzlaff-Reuter analysed a corpus of read speech produced by trained speakers, whereas the present study is based on spontaneous speech from naive subjects. Another possibility is that the two acoustic analyses show instances of phonologically different variants of the late peak category. The need to introduce a further distinction within the late peak category was recently suggested by the first author on the basis of functional arguments. This suggestion is further supported by a perception experiment by Kohler [15]. Using a semantic differential, he was able to sketch a comprehensive functional profile within an alignment continuum, in which the whole peak pattern was shifted across the accented syllable and part of the following unaccented syllable.

The latter possibility suggests that even more than two prenuclear rises have to be distinguished (in Standard German). In this connection, it is interesting that Atterer and Ladd [4] found that, deviating from other findings for prenuclear rises, the L was not aligned closely around the accented syllable onset, but within the accented vowel. They interpret this finding as a language-specific variation within the same phonological category, the prenuclear rise. From our perspective, it seems equally possible that the contrast in the L alignment goes back to different prenuclear rises.

5. References

[1] Bruce, G., 1977. *Swedish word accents in sentence perspective.* Lund: Gleerup.

[2] Kohler, K.J., 1987. Categorical pitch perception. *Proc. 11th ICPhS, Tallinn, vol. 5,* 331-333.

[3] Arvaniti, A., D.R. Ladd and I. Mennen, 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26, 3-25.

[4] Atterer, M. and D.R. Ladd, 2004. On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32, 177-197.

[5] Dilley L., D.R. Ladd and A. Schepman, 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33, 115-119.

[6] Pierrehumbert, J., 1980. *The phonology and phonetics of English intonation*. PhD thesis, M.I.T.

[7] Ladd D.R., D. Faulkner, H. Faulkner and A. Schepman, 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *JASA* 106, 1543-1554.

[8] Ladd, D.R., I. Mennen and A. Schepman, 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *JASA* 107, 2685-2696.

[9] Kohler, K.J., 1991. Prosody in speech synthesis: the interplay between basic research and TTS application. *Journal of Phonetics* 19, 121-138.

[10] Gartenberg, R. and C. Panzlaff-Reuter, 1991. Production and perception of F0 peak patterns in German. *AIPUK* 25, 29-115.

[11] Kohler, K.J., 1997. Modelling prosody in spontaneous speech. In Y. Sagisaka, N. Campbell, and N. Higuchi (eds.), *Computing Prosody. Computational models for processing spontaneous speech.* N.Y.:Springer, 187-210.

[12] Peters, B., 2001.*'Video Task' oder 'Daily Soap Scenario'.* www.ipds.uni-kiel.de/kjk/forschung/lautmuster.de.html

[13] Peters, B., 2003. *Die Datenbasis 'The Kiel Corpus'.* www.ipds.uni-kiel.de/kjk/forschung/lautmuster.de.html

[14] Arvaniti, A., D.R. Ladd and I. Mennen, 2000. What is a starred tone? Evidence from Greek. In M. Broe and J.B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, Cambridge University Press.

[15] Kohler, K.J., 2005. Timing and communicative functions of pitch contours. *Phonetica* 62, 88-105.