The phonetics and phonology of high and low tones in two falling f0-contours in Standard German

Tamara Rathcke, Jonathan Harrington

Institute of Phonetics and Speech processing, University of Munich, Germany

`tra@phonetik.uni-muenchen.de, jmh@phonetik.uni-muenchen.de`

Abstract

The present paper reports the results of an imitation experiment developed to evaluate empirically the validity of the AM-analyses given for two falling f0-patterns in German by different researchers. We look at the phonetic realisations of temporal alignment and frequency scaling of high and low tonal targets in varying syllabic environments. The effects of two phonetic factors were tested: (1) syllable structure of the postnuclear part of a phrase and (2) syllabic structure of the nuclear syllable. The results show that scaling and alignment are affected by the investigated factors in an unexpected way, so that predictions from different AM-analyses could not be confirmed by the data. We discussed the implications of the results in the light of the proposed analyses.

Index Terms: autosegmental-metrical phonology, pitch accent, temporal alignment, frequency scaling

1. Introduction

The autosegmental-metrical theory of intonation assumes that an f0-contour can best be described as a sequence of tonal targets, the **H***(igh)* and **L***(ow)* tones (e.g., [10], [14]). Phonologically, the tones are defined (1) at the syntagmatic level with regard to their function as edge-marking (*boundary tones*) or prominence-lending (*pitch accents*) and (2) at the paradigmatic level in terms of association between tones and rhythmically strong syllables. Phonetically, tonal targets vary in *alignment* and their *scaling* or location in frequency. For one thing, alignment and scaling are influenced by the phonological association of the corresponding tonal targets. For another, characterising the phonetics of f0-events is important for establishing the validity of tonal phonological categories. The problem of empirical validity of a phonological analysis has been widely discussed ([1], [11]) suggesting that there is a need for the empirical investigation of postulated AM-categories in a given language in order to understand how tonal association and its phonetic implementation are related. The aim of our study is to test the AM-analysis of one phonological contrast in German. So far, alignment and scaling properties have been studied for a single pitch accent under varying phonetic factors (cf. [2], [16], [18], and [19]). In the present investigation, we compare the phonetic realisations of two pitch accents in order to shed more light on the essentials of their phonologic structure.

German falling f0-patterns show at least two phonological pitch accents, often called *'early'* (f0-peak precedes the accented vowel) and *'medial'* (f0-peak aligned with the accented vowel) (see [5], [7], [9], and [13]). At present, there is no consistent AM-analysis of these f0-patterns. According to GToBI ([7]), medial peaks are analysed as monotonal H*, early peaks are in contrast bitonal H+L* pitch accents (we discussed the nature of the starred tone in this pitch accent elsewhere, see [17]). The phonetic difference of peak alignment between medial and early peaks in GToBI is analysed as a difference between unassociated and associated H-tone showing L as a starred tone in an early peak and as a boundary tone in a medial peak. According to the AM-analysis given in [5], both pitch accents have the phonological structure H*L. In this model, early peaks are indicated by an additional unassociated H-tone preceding the base H*L structure. In the following, we use the more common GToBI-analysis referring to both pitch accents under investigation.

Our aim in the present study, then, is to investigate temporal alignment and frequency scaling of high and low tones in German falling f0-contours with a view to evaluating the competing phonological analyses or to suggesting an empirically more appropriate analysis of these f0-patterns. Two phonetic factors were taken into consideration: (1) syllable structure of the postnuclear part (none vs. one vs. two syllables after the nuclear accented syllable) and (2) syllabic structure of the nuclear syllable (CV: vs. CVC). The following hypotheses were formulated:

1.a: alignment of H: The main difference between both pitch accents lies in the alignment of H: it is early for H+L* and relatively late for H* (e.g., [9]). We expect that the absence of unstressed syllables in German nuclear accented word leads to left-displaced, i.e. earlier alignments of H-targets (cf. [19]). Following a main premise of AM-phonology (see [14], and [19]) we expect to find starred H-tones aligned in a systematic way with the accented syllable despite the structural differences in open (CV:) vs. closed (CVC) syllables.

1.b: alignment of L: Alignment of Ls in bitonal pitch accents has been shown to be stable (see [1], [4]), so we expect to find relatively invariant alignments of L* with the accented syllable without any effects of the phonetic factors under investigation. Alignment of the low boundary tone is expected to be at a fixed temporal interval preceding the phrasal edge and unaffected by varying syllable count (cf. discussion in [2]).

2.a: scaling of H: So far, scaling variations of high tones have only been shown in relation to paralinguistic factors like different degrees of emphasis or surprise ([8], [12]) or to be caused by the downstep from one high pitch accent to the next (e.g. [10]). Since these factors were fixed (i.e., did not vary) in the experiment, we do not expect any changes of H-tone scaling.

2.b: scaling of L: For L*, we expect to find undershoot of low boundary tones in phrase-final nuclear syllables (see [6]). They should be much more undershot in phrase-final nuclear syllables in the CV:- compared with the CVC-structure.

2. Method

In the present experiment, we made use of the imitation task inspired by [15] and modified the procedure for the purposes

of our research questions. We presented subjects with two sets of materials. Firstly, a single audio stimulus of an unambiguous production by a trained phonetician of H+L* and H* pitch accents with low boundary tones. Second, visual stimuli consisting of a number of sentences each written on a single sheet of paper. The subject's task was to read each of these sentences, and to produce them with the perceived (i.e., H+L* L-% or H* L-%) melody of the audio stimulus. The auditory and visually presented stimuli differed only in their nuclear accented words. The same auditory stimulus was played to the subject prior to each presentation of all the visual stimuli. Further details of these two kinds of stimuli are given below.

2.1. Speech material

1. Visual stimuli: We designed carrier sentences and test words inspired by the test materials in [6]: proper names provide a base for highly controllable variation of segmental and syllabic structure of the nuclear word and still sound natural to speakers. The carrier sentence *'Das war Herr X'* ('That was Mr. X') was chosen for the recordings. Table 1 gives an overview of the test items used (i.e., of X).

2. Audio stimuli: The sentences for audio stimuli had the same structure as the test stimuli presented visually except for the accented word: *'Das war Herr Neumann.'* ('That was Mr. Newman'). Since *Neumann* has only voiced segments and two syllables, we could assume that the intended pitch accents can easily be realised and perceived in these contexts. The stimuli were spoken by a trained male speaker of Standard German (Northern variety). Each prosodic boundary was produced with a low boundary tone which can be labeled as L-% in GToBI-system (see [7]).

Table 1. *Overview of test items selected as visual stimuli for the imitation task.*

Syllable count after the nucleus	Syllable structure	
	CVC	CV:
none	*Linn*	*Lie*
one	*Linner*	*Liener*
two	*Linnerer*	*Lienerer*

2.2. Procedure, recordings and subjects

As explained above, the experiment used auditory and visual stimuli. The audio stimuli were presented through headphones of good quality and the test items for imitation were presented visually on small sheets of paper. Each subject was placed in a sound treated room in front of a microphone. The subject was asked to listen closely to the melody of the sentence 'That was Mr. Newman' and to produce the read sentence with the melody that had been perceived. The subjects were asked to repeat any imitation that they thought to have been produced incorrectly. Each visual stimulus was presented three times, in a random order. The recordings took place at the Institute of Phonetics and digital Speech Processing (IPdS) Kiel.

14 speakers of standard North German (6 M and 8 F) participated in the experiment. All subjects were between 21 and 38 years of age and none of them reported any speech or hearing disorder. Three subjects were students of phonetics at the IPdS Kiel, the others had no experience in phonetics.

3. Results

The data of 4 subjects (3 M, 1 F) were not taken into account for two reasons: Firstly, because two of the male speakers produced an excessive amount of creaky voice making measurements difficult; and secondly because the other two subjects were unable to produce any difference between two intended pitch accents (*H+L** and *H**). For these reasons, only the data from 10 subjects (3 M, 7 W) were analysed here.

Segment edges of each test item as well as high and low turning points in their f0-trajectories (H and L targets) were labelled manually using the EMU speech database system (see [3]). Examples of both pitch accents (realised by the same female speaker) are given in Figures 1 and 2.
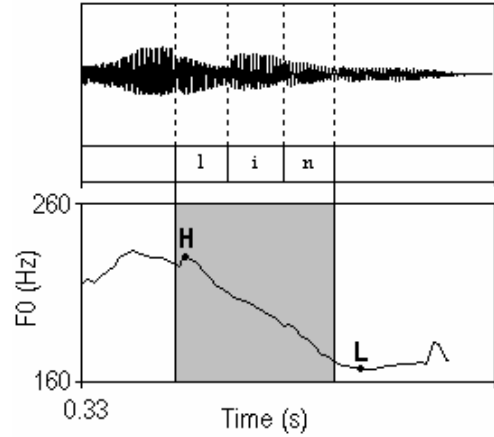

Figure 1: *Waveform, segment labels of the accented syllable and f0-contour of "Herr Linner" realised with H+L* L-%, H and L indicate the turning points in the f0-contour. The accented syllable is shown in grey.*
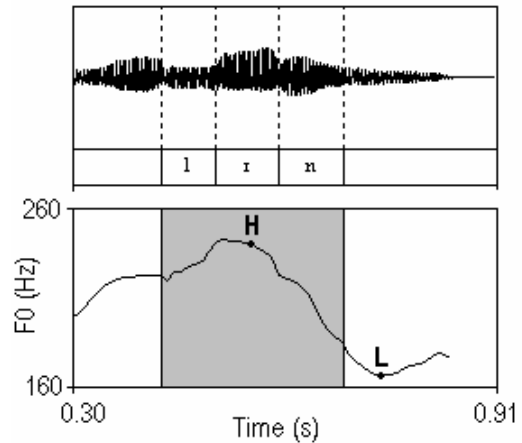

Figure 2: *Waveform, segment labels of the accented syllable and f0-contour of "Herr Linner" realised with H* L-%, H and L indicate the turning points in the f0-contour. Accented syllable is shown in grey.*

For each test condition, the measurements of f0-scaling and temporal alignment of the marked up H and L targets were obtained. For scaling measurements, all f0-values in Hz, f_{Hz}, were converted into semitones, f_{St}, using the formula:

$$f_{St} = 12(\log_2 f_{Hz} - \log_2 k) \tag{1}$$

where k is a speaker-dependent constant equal to the average f0-value in Hz across all of the frames of all of the speaker's test tokens. The above formula sets each speaker's mean f0-value to 0 st and thereby acts as a speaker-normalization of the f0-data. The f0-values of the H and L tones are given in Table 2.

There is some debate about how to measure the alignment of tonal targets. One common procedure is to define alignment as an actual temporal interval (in ms) between a segmental landmark and the target (e.g., [2], [16]). In this

case, the segmental landmark must be immediately adjacent to the tonal target, because the greater the time interval, the greater the alignment variance (cf. [18]). Alternatively, alignment can be defined as a proportion of the duration of a given segment ([19]). The latter approach has been shown to reduce the variance caused by different segmental durations. Since the main precondition for using actual alignment is a temporal proximity between a tonal target in question and a segmental landmark ([18]), the first approach can be effectively used only for comparisons of alignment realisations of the same pitch accent. In the present study, we wanted to compare alignment realisations of two different pitch accents and two different syllable structures. So, we decided to use proportional alignment (2):

$$t_T = (T - S_{on}) / (S_{off} - S_{on}) \qquad (2)$$

From (2), the onset of the nuclear accented syllable (S_{on}) is set to 0 and its offset (S_{off}) to 1: thus, (2) can be interpreted as a linear temporal normalization of the alignment data. The results of these proportional alignment-measurements are shown in Figures 3 and 4. Additionally, we measured the time interval from right phrasal edge to each L-turning point (i.e., actual alignment in ms, see Table 3).

Table 2. *Averaged f0-values of labeled H and L turning points (in semitones) for different test conditions (n=10).*

Test conditions		H-values (st)		L-values (st)	
Syll. structure	Syll. count	H+L* L-%	H* L-%	H+L* L-%	H* L-%
CV:	none	2.7	3.0	-3.8	-2.6
	one	2.9	3.0	-3.6	-2.9
	two	3.0	3.4	-3.3	-2.7
CVC	none	2.8	2.9	-3.5	-2.7
	one	2.7	3.1	-3.5	-2.6
	two	3.2	3.5	-3.2	-2.7

Table 3. *Averaged time intervals between the right phrasal edge and labeled L turning points (in ms) for different test conditions (n=10).*

Syll. strucure	CV:			CVC		
Syll. count	none	one	two	none	one	two
H+L* L-%	64	125	226	92	130	244
H* L-%	43	120	235	68	115	223

Five repeated measures ANOVAs were run to test the effects of two independent variables: (1) the count of postnuclear syllables (none vs. one vs. two) and (2) the syllabic structure of the nucleus on both dependent variables (i.e., on the alignment and on the scaling of H- and L-targets). The following statistically significant results were found:

- *Alignment of H* is significantly later for H* than for H+L* ($F_{1,9}$=96.1, p<0.001). Increasing the count of postnuclear syllables leads to later alignments of both pitch accents ($F_{2,18}$=76.9, p<0.001). H-tones are consistently later aligned in CV:-syllables compared to CVC-syllables ($F_{1,9}$=98.9, p<0.001). Syllable structure and syllable count also interact ($F_{2,18}$=17.3, p<0.001).

- *Alignment of L* is significantly later with an increasing number of syllables following the nucleus ($F_{2,18}$=284.9, p<0.001). There is also a strong effect of syllabic structure ($F_{1,9}$=143.5, p<0.001) showing that Ls are aligned later in CV: than in CVC-syllable words. The syllabic structure and syllable count also interact ($F_{2,18}$=15.6, p<0.01). The alignment of L is only influenced by the count of

postnuclear syllables ($F_{2,18}$=159.1, p<0.001), when it is measured as a time interval to the phrasal edge.

- *Scaling of H* is significantly affected by the syllable count ($F_{2,18}$=7.9, p<0.01): this shows that the f0 of H-tones increases with an increasing number of syllables following the nucleus.

- *Scaling of L* is strongly influenced by the pitch accent ($F_{1,9}$=28.0, p<0.001): on average, the f0 of L-tones are 0.5-0.9 st lower in H+L* than in H* pitch accents.
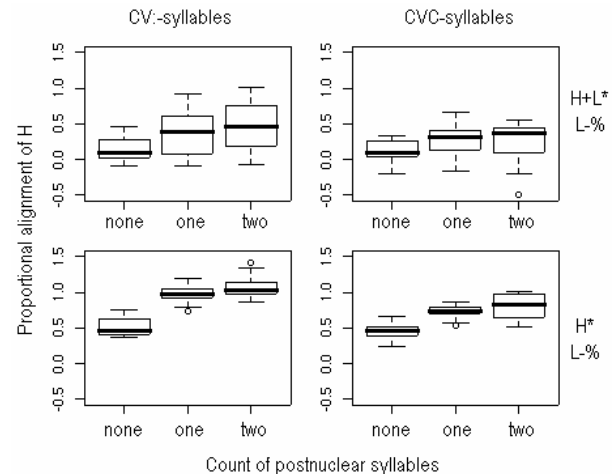
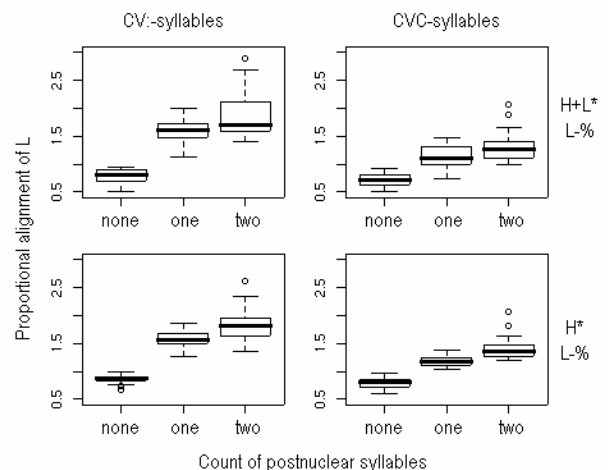Figure 3: *Proportional alignment of H-turning points for different test conditions (n=30).*

Figure 4: *Proportional alignment of L-turning points for different test conditions (n=30).*

4. Discussion and Conclusions

In the present study, we investigated the behavior of high and low f0-targets in realisations of two pitch accents under varying phonetic conditions. Before we can make conclusions about implications of our results for the AM-analysis of early and medial peaks in German, we will consider in turn the hypotheses set up in 1.

Consistently with *1.a*, early and medial peaks mainly differ in the alignment of H: it is aligned in the first half of the syllable for an early peak and in the second half for a medial peak. In both cases, a leftward displacement of high f0-targets occurs under time pressure caused by decreasing the amount of segmental material available in the tail: that is, the tendency shown for other languages (see [16], [18], and [19]) is also confirmed for German. Contrary to the expectation

formulated in *1.a*, there are significant differences in the alignment of high targets depending on the syllabic structure of the nucleus: they show consistently later alignment in open syllables with long vowels.

Contrary to *1.b*, alignment of Ls is unstable both in early and in medial peaks. Furthermore, the alignments of both pitch accents do not differ from each other, which is surprising if we assume that an L-tone has a different status depending on whether it is an associated tone in a bitonal pitch accent or a boundary tone preceding a phrase edge. However, the interval between the right-hand phrasal edge and the L-target is not fixed: as discussed in [2], a steady interval between the target and the boundary is typical of boundary tones. In our data, the L-alignment shows some displacement effects similar to those of Hs being aligned earlier in closed syllables and under time pressure.

Contrary to *2.a,* the scaling of Hs was shown to be affected by the count of postnuclear syllables: the highest peaks occur in words with several following postnuclear syllables and the lowest peaks are more common in word-final nuclear accented words. This result is explicable in terms of the topline declination defined as a global downtrend during a sentence in e.g. IPO model of intonation ([20]). But it cannot be explained in terms of the AM model of intonation, since the downtrend is assumed to be a result from a local downstep operating from one high pitch accent to the next in a sequence of several high tones ([10]).

Contrary to *2.b*, there is no undershoot of the L-target in any condition. Instead, we found that L-targets are scaled significantly lower in early peaks.

The results presented here suggest that there are some difficulties with the two kinds of AM-analyses ([5, 7]) discussed earlier. On the one hand, our results might favor [5] for two reasons: (1) in both pitch accents, the high tone is aligned with the accented syllable in the same systematic way that has been shown for other languages, (2) the alignment of the low tone is more variable: it is not temporally anchored to the phrase boundary nor is it systematically different for the two pitch accents. This would seem to support the argument for H*L for both pitch accents. The analysis of an early peak as H H*L is further supported by an observation (not investigated here) that early peaks in these data are always preceded by a long high plateau. However, such a solution ignores the important finding that H-alignment is very different in early and medial peaks (see results for *1.a*).

On the other hand, the finding that the L-scaling is lower in early peaks as well as perceptual investigations (e.g., [13], [17]) are consistent with an analysis of the early peaks as L*. As has been shown in [13], the perceptual difference between medial and early peaks depends predominantly on the perception of the pitch fall: relatively steep falls through the accented syllable (as a result of which the low target is reached faster than in medial peaks) are typical of early peak accents. In other words, early peaks in German sound *low* not high. Furthermore, acoustic analyses of early peak sequences (see e.g., the examples in [17]) shows that in prenuclear early peaks, the f0-trough from the preceding high peak is always reached in the accented syllable. In conclusion, we favour the phonological analysis of early and medial peaks as proposed in GToBI ([7]), although the empirical results of the present study do not directly support this analysis. Our results show that there is much more to be learned about the nature of the phonological association and how it is realized phonetically.

5. References

[1] Arvaniti, A., Ladd, D., R., 1995. What is a starred tone? In Broe, M., Pierrehumbert, J. (eds.) Papers in Laboratory Phonology V, Language Acquisition and the Lexicon. Cambridge: Cambridge University Press, 119-131.

[2] Arvaniti, A., Ladd, D., R., Mennen, I., 1998. Stability of tonal alignment: the case of Greek prenuclear accents. Journal of Phonetics 26(1), 3-25.

[3] Bombien, L., Cassidy, S., Harrington, J., John, T., Palethorpe, S., 2006. Recent developements in the EMU speech database system. Proceedings of the 11th Australasian International Conference on Speech Science and Technology. Auckland

[4] Caspers, J., Heuven, V. van, 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. Phonetica 50, 161 - 171

[5] Féry, C., 1993. German Intonation Patterns. Tübingen: Max Niemeyer.

[6] Grabe, E., 1998. Pitch accent realisations in English and German. Journal of Phonetics 26, 129-143.

[7] Grice, M., Baumann, S., Benzmüller, R., 2005. German intonation in autosemental-metrical phonology. In S.-A. Jun (ed.) Prosodic typology. The phonology of intonation and phrasing. Oxford: University Press, 55-83.

[8] Gussenhoven, C., Rietveld, T., 2000. The behavior of H* and L* under variations in pitch range in Dutch rising contours. Language and Speech 43(2), 183-203.

[9] Kohler, K., J., 1987. Categorical pitch perception. Proceedings of the XI ICPhS, Tallin, 331-333.

[10] Ladd, D., R., 1996. Intonational Phonology. Cambrigde: University Press.

[11] Ladd, D., R., 2000. Bruce, Pierrehumbert, and the elements of intonational phonology. In Horne, M. (ed.) Prosody: theory and experiment. Netherlands: Kluwer, 37-50.

[12] Ladd, D., R., Morton, R., 1997. The perception of intonatinal contrast: continiuos or categorical? Journal of Phonetics 25, 313-342.

[13] Niebuhr, O., 2003. Perceptual study of timing variables in f0 peaks. Proceedings of the XV ICPhS, Barcelona, 1225-1228.

[14] Pierrehumbert, J., B., 1980. The Phonology and Phonetics of English Intonation. Ph.D. dissertation, MIT. [Published by Indiana University Linguistics Club, Bloomington].

[15] Pierrehumbert, J., B., Steele, S., A., 1989. Categories of tonal alignment in English. Phonetica 46, 181-196.

[16] Prieto, P., van Santen, J., Hirschberg, J., 1995. Tonal alignment patterns in Spanish. Journal of Phonetics 23, 429-451

[17] Rathcke, T., Harrington, J., accepted. The variability of early accent peaks in standard German. Papers in Laboratory Phonology 10.

[18] Schepman, A., Lickely, R., Ladd, D., R., 2006. Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. Journal of Phonetics 34, 1-28.

[19] Silverman, K., Pierrehumbert, J., B., 1990. The timing of prenuclear high accents in English. Papers in Laboratory Phonology 1, 72-106.

[20] 't Hart, J., Collier, R., Cohen, A., 1990. A perceptual study of intonation: an experimental-phonetic approach. Cambrigde University Press.