Is there a distinction between H+!H* and H+L* in standard German? Evidence from an acoustic and auditory analysis.

Tamara Rathcke & Jonathan Harrington

Institute of Phonetics and digital Speech Processing
Christian-Albrechts-University of Kiel, Germany
{tkh; jmh}@ipds.uni-kiel.de

Abstract

This paper is concerned with intonation in German and whether there is a phonological distinction between two types of early peaks H+L* and H+!H*. Speech perception and production data are presented to shed light on this issue. The results show little evidence for a phonological distinction between these categories. The results are interpreted in terms of the relationship between downstep and early peak placement in German.

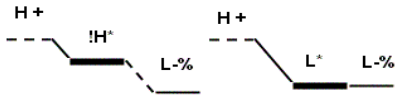
1. Introduction

One of the main aims of intonational phonology is to abstract a set of phonological categories from phonetic variability in order to uncover the linguistically relevant contrasts. A number of methodologies have been developed for this purpose with varying degrees of success (see e.g. [3, 62ff.]).

As far as the phonology of German intonation is concerned, there is no consensus about the number of different tonal categories. Although there is experimental evidence that the distinction between an early (pitch peak precedes the accented vowel) and a mid (pitch peak aligned with the accented vowel) is phonologically contrastive in German (e.g. [4]), the question of whether there are two early peaks is more controversial. In an autosegmental-metrical (A-M) model, Kohler's [4] early/mid peak distinction corresponds approximately to the distinction between H+L* and H* respectively. In the A-M model, a further possible categorical distinction is posited between H+L* and H+!H* and this distinction is adopted in the A-M treatment of German by Grice & Baumann [2]. In both of these categories, there is a pitch peak due to the H+ that precedes the accented vowel: the difference between them is that in H+L* a low pitch target is reached in the accented vowel, whereas in H+!H* there is a downstepped peak that occurs during the accented vowel, so the target step is from high to mid. This difference is illustrated in Figure 1 (due to [2]). In this analysis, the original use of downstep as a syntagmatic process of tone lowering [3, 5] is being used to expand the paradigmatic inventory of pitch accents: that is, the phonological theory of two tones which is postulated to be more transparent and sufficient in comparison with the earlier four tone models of intonation (cf. [6]), is effectively expanded to be a three tone model with high, medial and low tone levels.

There are only a couple of examples from German connected speech which can be interpreted in terms of this tonal distinction (cf. [13]). An acoustic analysis of minimal pairs spoken by Stefan Baumann (e.g. [14]) suggests that the main differences between H+!H* and H+L* concern the magnitude of the pitch fall to the accented vowel and the achievement of speaker's base line values: that is, the F0 range between the height of the early aligned pitch peak and the base line is divided into a medial and a low tonal regions which contrast with each other at the accented syllable. The semantic contrast that is evoked by this tonal distinction can be summarized as that between general or polite statements (H+!H*) and resolute or soothing assertions (H+L*). For example, H+!H* might be used in a context when an adult wishes to reassure a young child that the performers in a circus are not real [13] (accented words are underlined): *'Du brauchst keine Angst haben. Es sind nur Schauspieler.* ('Don't worry. They're only actors.). On the other hand, H+L* might be used assertively to remind an older child of something s/he already knows: *'Du weißt doch, dass in Filmen nichts echt ist. Es sind nur Schauspieler'.* (Come on, you know that films aren't real. They're only actors.).

Figure 1: *Formal differences between two falling nuclear pitch accents: H+!H* L-% (left) vs. H+L* L-% (right). The bold line marks the accented syllable. The continuous lines indicate the obligatory pitch movements, the dashed lines give the variable pitch movements.*



In recent experimental studies of German intonation, no phonological distinction is made by Kohler [4] between two early peaks while Grabe [1] considers the distinction between H+L* and H+!H* to be gradient.

The goal of the present study is to test experimentally a hypothesis about the status of the difference described above using both perceptual and acoustic techniques. We assume there is a phonological distinction between H+!H* and H+L* pitch accents in German based upon the different F0-targets at the accented syllable or vowel, i.e., it is mid for H+!H* (as appropriate for a downstepped peak) but at a low level, near the speaker's baseline for H+L*. The general semantic difference is assumed to be that between general and resolute statements.

We made use of three experimental procedures to assess the evidence for a categorical distinction between H+L* and H+!H*: (1) a semantic congruity test in which listeners were asked to rate the appropriateness of a set of sentences whose intonation had been manipulated to create a continuum between these categories, (2) an AXB discrimination test of tripled tokens from the synthetic continuum and (3) a production experiment inspired by [7] in which speakers were asked to imitate the synthetically manipulated sentences.

2. Method

2.1. Speech material and stimulus generation

The test sentence *'Sie mag Bananen'* ('she likes bananas') was produced by a male speaker with experience in prosody with

an H* L-L% falling contour and with the H* on *'na'* of nuclear accented *'Bananen'*. This test sentence was used for resynthesis purposes to create a continuum from H* to the downstepped H+!H* to H+L*.

The resynthesis was accomplished as follows. Two points were fixed in time and frequency for all resynthesized utterances: one at the nasal midpoint (at 110 Hz) and the other at the offset of the accented vowel (at 70 Hz). The F0-peak in the original utterance which was at 120 Hz was progressively lowered in 10 Hz steps. The lowest manipulation point was limited by speaker's base line, so that the tonal area between the H*-value of the original and the value of speaker's base line was divided by 5 equal steps resulting in 6 stimuli. The F0-trajectories of the stimuli are shown in Figure 2. The first stimulus corresponds to the original contour of the test sentence.

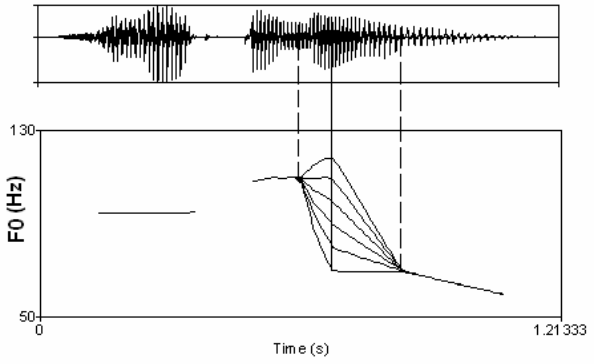


Figure 2: *Synchronized time-waveform and the six tokens of the synthetic continuum. The dashed lines mark the first fixed reference point preceding the accented vowel and the second reference point at the vowel offset, respectively. The solid vertical line shows the manipulation point in the accented vowel.*

According to the assumptions made in 1., the stimuli of continuum can be related to pitch accent categories shown in Table 1.

Table 1: *F0-values in the accented vowel and the pitch accent categories to which the F0-manipulations are assumed to correspond (column 3)*

stimulus number	*F0 at the vowel (Hz)*	*pitch accent (target at the vowel)*
1	120	H*
2	110	(high target)
3	100	H+!H*
4	90	(mid target)
5	80	H+L*
6	70	(low target)

2.2. Subjects

10 speakers of standard North German, 5 M and 5 F of between 20 and 40 years of age with no known speech or hearing disorders participated in all tests. Some subjects were trained students of phonetics at the IPdS Kiel, the others had no experience in phonetics. None of the subjects were told of the purpose of the experiment.

2.3. Semantic congruity test

With the assistance of Stefan Baumann (pers. communication), two contexts were created that were considered to be appropriate for the production of H+!H* and H+L* respectively (the test sentence is underlined):

(1) (the H+!H* context) *'Katrin hat einen ganz normalen Geschmack. Sie mag gern Bananen. Aber sie mag auch anderes Obst.'* (Katrin's taste in food is typical. She likes bananas. But she likes other fruit too).

(2) (the H+L* context) *'Wenn man Katrin einen Korb mit Kirschen, Bananen und Erdbeeren hinstellen würde, wüsste ich genau, was sie auswählen würde. Sie mag gern Bananen. Anderes Obst verabscheut sie.'* (If you present Katrin with a basket of cherries, bananas, and strawberries, I know exactly what she'd choose. She likes bananas. She has a dislike for other fruit.)

A tape containing 30 randomized stimuli with 5 seconds of silence between them were created from 5 repetitions of each of the 6 stimuli. The tape was presented to the subjects who had to judge, by marking their answers on a sheet of paper containing the written contexts, whether the stimulus was most appropriate for the H+!H* or the H+L* of neither of these contexts. They did this for each stimulus during the intervening 5 s pause. The stimuli were presented from a computer via headphones in a silent room at the IPdS Kiel.

2.4. Discrimination test

As discussed in 2.1, there were 6 stimuli that were assumed to correspond to the three categories in Table 1. For the discrimination test, the following cross-category pairings were presented in both AXB and BXA-orders (with four possible combinations AAB, ABB, BAA, BBA): 1&3, 2&4, 3&5, 4&6. Additionally, the following within-category pairings: 1&2 (as AAB); 3&4 (as ABB); 5&6 (as BBA) as well as the cross-category pairing of 1&6 (as BAA) were used as control stimuli.

These combinations resulted in 20 pairings and each of these 20 pairings was repeated 5 times resulting in 100 experimental stimuli. For each AXB trial, the listeners had to judge which pair of stimuli were the same and to mark their answers on a sheet of paper. The stimuli were randomized and played from a computer via headphones in a silent room at the IPdS Kiel.

2.5. Imitation test

The six stimuli detailed in 2.1 were each repeated to the subject 10 times resulting in 60 items. These 60 items were randomised and then each presented twice with a preceding beep. After the second presentation of each stimulus, there was a pause during which the subject was instructed to imitate it paying particular attention to copying the melody as closely as possible. In the event of a hesitation or speech error, the item was repeated. No time limit was imposed for responses. The imitation experiment was carried out in a sound treated recording studio at the IPDS Kiel.

The raw F0-contours were smoothed over the extent of /nan/ in 'Bananen' using the first few coefficients of the discrete-cosine-transformation (DCT). A DCT decomposes any time signal of length *N* points into a set of N-point cosine waves of increasing frequency 0, 1, 2, .. *N*-1 radians/sample such that, when these are summed, the exact signal is reconstructed. If the signal is reconstructed starting from the lowest frequency cosine waves, then the signal is smoothed, and the fewer the cosine waves that are used in this summation, the smoother the signal. We found that summing the first 6 coefficients (with frequencies k = 0, 1, 2…5 radians/sample) produced a suitably smooth contour showing clear F0-peaks and valleys but without deviating too much from the raw signal as shown in Figure 3. Since the DCT

coefficients encode the shape of the signal, they can also be used to classify shape differences: for the first 6 DCT coefficients each F0-trajectory is represented as a point in a six-dimensional space and two trajectories that differ significantly in their shape will be positioned in different coordinate regions in this multidimensional space. The lowest 3 DCT coefficients are related to the shape of the original signal in the following way: they encode the mean (from the onset to the offset of the signal), the linear slope, and the curvature of the signal (see [11] and [12] for formulae and further details).

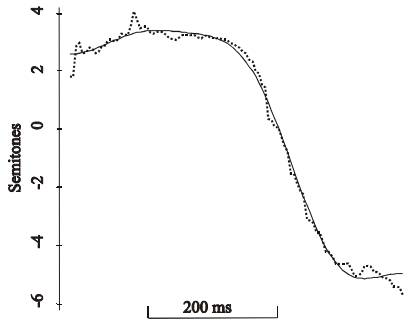


Figure 3: *A smoothed F0-contour derived by summing the first six DCT coefficients (solid) superimposed on a raw F0-contour (dotted) of [nan] from 'Bananen' produced by a male speaker. 0 st = 87.3 Hz = the speaker's F0-mean across all [nan] tokens.*

Prior to DCT-smoothing, any 0 values close to zero frequency in the signal that were due to pitch-tracking errors were readjusted by linear interpolation and the raw values in Hz, f_{Hz}, were converted into semitones, f_{St}, using the formula:

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

where k is speaker-dependent constant equal to the average F0-value in Hz across all of the frames of all of the speaker's [nan] tokens. The above formula sets each speaker's mean F0-value to 0 st and thereby acts as a (crude) form of speaker-normalization of the F0-contours.

3. Results

The results of the semantic congruity test (Figure 4) showed that subjects divided the six-point continuum into two categories: the division was between stimuli 1-2 on the one hand and 3-6 on the other, that is into some form of mid (1-2) as opposed to early (3-6) peak with little evidence of a semantic discrimination between the latter. The medial peaks (1-2) were predominantly associated with the semantic context assumed to be appropriated for the H+L* realisations, whereas the early peaks (3-6) were matched with the context created for the H+!H* stimuli. Only very few stimuli were judged to fit into neither of the two semantic contexts. A repeated measures ANOVA with independent variables *stimulus* and *context* showed that the context alone had a significant influence on the *'matching'* judgements ($F = 19.86$, $p < 0.001$), whereas the interaction between context and stimulus was not significant ($F = 2.71$). This statistical result was affected by response variation between subjects: there were three female subjects who divided the continuum in the same way (1-2 vs. 3-6), but who showed the reverse pattern of contextual associations (1-2 as matching for the H+!H*-context, 3-6 as matching for H+L*-context).

Figure 5 shows the results of the discrimination test for all subjects. There was an overall greater than 50% discrimination within the AXB-triples. The within-category (within 1&2; within 3&4; within 5&6) stimuli were distinguished slightly less often than between categories. Discrimination was greater in the AXB than in the BXA sequences which is compatible with other findings (e.g. [10]) showing a time-order error. The results of a repeated measures ANOVA showed that the type of pairings (control vs. AXB- vs. BXA-pairings) had a significant influence on the *'different'* judgements ($F = 8.803$, $p < 0.01$).

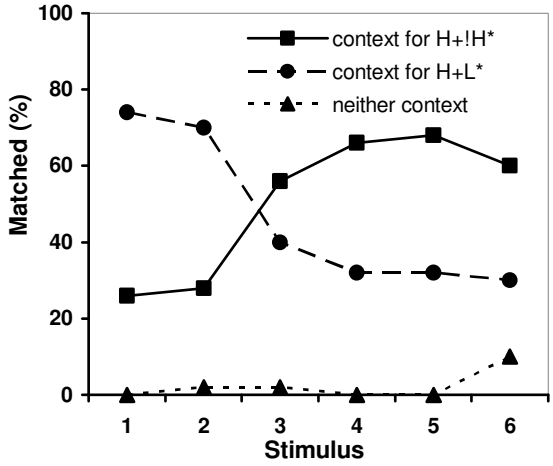


Figure 4: *Results of semantic congruity tests: percentage of 'matching' judgements for 6 stimuli depending on the context (n = 10).*

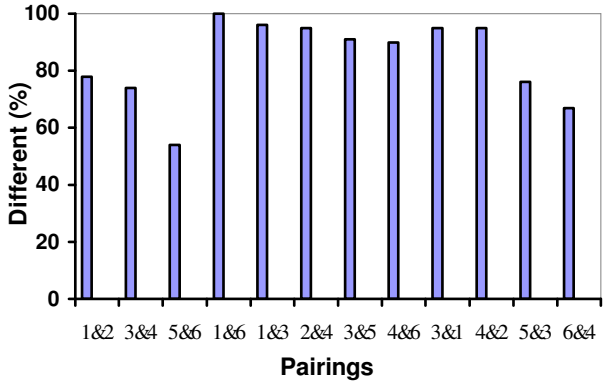


Figure 5: *Results of AXB-discrimination test: percentage of 'different" judgements for control pairings (1&2, 3&4, 5&6, 1&6) as well as test item pairings in both AXB (1&3, 2&4, 3&5, 4&6) and BXA (3&1, 4&2, 5&3, 6&4) orders (n = 10).*

Turning now to the production data, we used t-tests (with a conservative, Bonferroni adjusted significance level) to assess the extent of acoustic differences in the imitated contours. Pairwise comparisions were made between successive stimulus numbers: that is, we compared the imitation to stimulus n with the imitation to stimulus n+1 (n = 1, 2, 3, 4, 5). The dependent variables were the 6 DCT-coefficients which, as described earlier, encode the shape of the F0-trajectory. The results of these pairwise comparisions showed

significant effects only for DCT0 (the mean of the contour over [nan]) in comparing stimulus 1 and 2 ($p < 0.001$) as well as in comparing stimulus 2 and 3 ($p < 0.01$). There were no significant differences for any other stimulus numbers nor for any of the other DCT-coefficients.

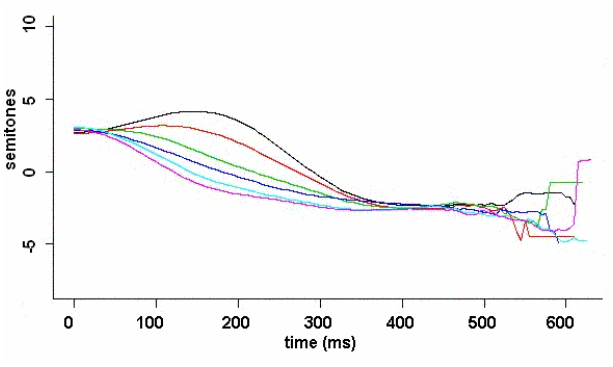


Figure 6: *DCT-smoothed F0-contours of imitated [nan] in 'Bananen' averaged across all speakers separately for each of the six stimulus numbers (stimuli 1-6 are from top to bottom respectively). The contours were synchronized at the onset of the first [n] in [nan] (t = 0 ms) prior to averaging.*

4. Discussion

The results of the semantic-congruity test show that the 6-point continuum was divided into two categories: a mid-peak which is likely to correspond to H* and an early peak which collapses across H+!H* and H+L*. Against expectations, the stimuli with H* tones were predominantly judged to be congruous with the H+L* context. It is of course possible that we did not manage to devise semantic contexts appropriate for the H+!H* vs. H+L* distinction. The results of the congruity test were, however, supported by the imitation experiment which suggests that stimuli 2 (H*) and 3 (H+!H*) span a tonal category boundary. But neither the congruity nor the imitation experiments suggest a phonological tonal category within any of stimuli 3-6.

One of the reasons why the speakers might have perceived and realized a difference between stimuli 1 and 2 is that they were hearing a gradient difference in prominence rather than a difference in phonological tonal categories (see e.g. [8]).

The experiments of the discrimination test were the least conclusive of the three types of analyses. They suggest that pitch changes can be discriminated perceptually even though these perceived differences might not represent differences in phonological categories.

Our results so far seem to provide reasons for dispensing with H+!H* as a tonal category at least in German (and quite possibly in English) in the absence of any experimental evidence supporting a phonological category distinction between H+!H* and H+L*. We do not, of course, wish to rule out the possibility of a *phonetic* distinction between these two types of tones: further research needs to uncover whether there are prosodic contexts (such as the syllable-count and position in the phrase) that give rise to an H+!H*-like F0-shape on the one hand vs. an H+L*-like F0-shape on the other. If, as we suspect, it turns out that this distinction is indeed phonetic and not phonological/categorical, then there is no reason to distinguish between them at the level of tonal phonology, any more than there is to give different category labels to the context-induced variation in the tonal alignment of a H* in English that has been shown to be predictable from factors such as the presence of a prosodic boundary and tonal-clash (e.g. [9]).

Beyond the specifics of the alignment of peaks in German, there is a more general issue to be considered: by including H+!H* in the phonological analysis of tones, the concept of downstep becomes unclear. On the one hand, downstep represents a tone lowering due to a *syntagmatic* relationship (with a preceding high tone in the same prosodic phrase). On the other hand, the inclusion of a tone like H+!H* means that downstep is being used *paradigmatically* to expand the inventory of available pitch-accent contrasts. There seem therefore to be good reasons for a re-evaluation of the whole concept of downstep in the A-M model of intonation.

5. Conclusions

Presented experiments suggest that there is a distinction between a mid (H*) and an early (H+L*) pitch-accent in German, but there is no further phonological distinction within the early category.

6. References

[1] Grabe, E., 1998. Comparative Intonational Phonology: English and German. MPI Series in Psycholinguistics 7. Wageningen: Ponsen und Looijen.

[2] Grice, M., Baumann, S., 2000. Deutsche Intonation und GToBI. *Linguistische Berichte 181,* Hamburg: Helmut Buske Verlag, 1-33.

[3] Gussenhoven, C., 2004. The Phonology of Tone and Intonation. Cambrigde: University Press.

[4] Kohler, K. J., 1987. Categorical pitch perception. *Proceedings of the XIth International Congress of Phonetic Siences.* Tallin, 331-333.

[5] Ladd, D. R., 2001. Intonational Phonology. Cambrigde: University Press.

[6] Pierrehumbert, J. B., 1980. The Phonology and Phonetics of English Intonation. PhD Diss. MIT.

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

[8] Rietveld, A. C. M., Gussenhoven, C., 1985. On the relation between pitch excursion size and prominence. *Journal of Phonetics 13,* 299-308.

[9] Silverman & Pierrehumbert, 1990. The timing of prenuclear high accents in English. *Laboratory Phonology 1,* 72-106.

[10] Schiefer, L., Batliner, A., 1988. Intonation, Ordnungseffekt und das Paradigma der kategorialen Wahrnehmung. In *Intonationsforschungen,* Altmann, H. et al. (eds.). Thübingen: Niemeyer.

[11] Watson, C. I., Harrington, J., 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America, 106 (1),* 458-468.

[12] Harrington, J., in press. Community influences on adults during sound change. *Journal of Phonetics (special edition), 'Variation & Cognition'*, Hey, J., Jannedy, S. (eds.).

[13] http://www.coli.uni-sb.de/phonetik/projects/Tobi/index_training.html

[14] http://www.uni-koeln.de/phil-fak/phonetik/gtobi/minpairs.html