The variability of early accent peaks in Standard German

Tamara Rathcke and Jonathan Harrington

This paper is concerned with the relationships between 'early' pitch accents in German and with whether downstep in German is phonological or phonetic. At the core of our analysis is an investigation into the differences between two kinds of pitch accents in which the pitch peak precedes the accented vowel: these are H+!H and H+L* which are claimed to be phonologically contrastive in German. We made use of two experimental procedures: (1) a production experiment in which speakers were asked to imitate synthetically manipulated sentences and (2) a semantic differential experiment in which listeners rated the perceived meaning of those sentences on eight semantic scales. Although both the imitation and perception experiment provided evidence for a distinction between an early and a later (H*) peak accent, the results pointed neither to a three-way distinction, nor to a categorical distinction between H+!H* and H+L*. Finally, we present some results from an analysis of both English and German corpora which suggest that the difference between these two peaks may be phonetic and attributable to the number of syllables following the nuclear accent in the tail. Based on these results and from theoretical considerations, we argue that H+!H* is an inappropriate pitch accent category in the inventory of paradigmatic phonological intonational contrasts of standard German.*

1. Introduction

The autosegmental-metrical (AM) theory of intonation rests on the premise that an f0-contour is derived from a sequence of abstract phonological tone-targets that are systematically aligned with segmental units (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Ladd 1996). The tonal inventory is reduced to a basic distinction between H(igh) and L(ow) tones and a starred notation is used in pitch accents to indicate the phonological association between a tone and a rhythmically strong syllable, which in turns has a major influence on their temporal alignment. In one sense, temporal alignment is clearly phonological because it is the result of the way that the tune is associated with the text, but in another alignment is also context-dependent and phonetic as various experiments have shown (Silverman and Pierrehumbert 1990). The frequency scaling of the High and Low tones is assumed to be predictable from phonetic factors and controlled by a system of rules. As Liberman and Pierrehumbert (1984) have argued, downstep is one of the mechanisms used to justify the adherence to a two-tone model in the face of what are evidently many different f0-levels on the surface. Thus, in the famous stepping contour, the progressive lowering of f0-levels on the strong syllables of successive accented words can be modelled by downstep that is iteratively (and in theory infinitely) applied to a H* tone within a single phrase. This specification of downstep is clearly phonetic.

However, it has often been difficult to place 'downstep' along the divide between phonological categories and phonetic implementation. Downstep was originally invoked in tone languages to express the idea that there is a high tone that is automatically lowered after a low tone in an H-L-H sequence and in the AM model, downstep refers to the lowering of successive high tone targets within a prosodic phrase. Unlike Liberman and Pierrehumbert (1984), Ladd (1996: 97) argues that there is also evidence that downstep is largely an 'orthogonal phonological variable' with the meaning of 'finality' or 'completeness' (cf. Ladd 1996: 90–92; this was also argued in Ladd 1983).

The downstepped pitch accents in American English include monotonal !H* and bitonal L+!H* (Beckman and Ayers 1994). Since they necessarily follow a non-downstepped and often bitonal accent, these downstepped accents cannot occur in phrase-initial position. Our interest lies in a third type of downstepped pitch accent H+!H* which arose in the evolution of the ToBI system, based on the AM model, for transcribing standard American English intonation. In the ToBI system, H+L* came to be replaced by H+!H* (Ladd 1996: 96–97; Gussenhoven 2004: 132) both for phonetic reasons and because the level of the starred tone was not usually at the bottom of the speaker's range, as would be expected if the second component of this bitonal accent were actually L*. Given that one pitch accent essentially replaces another in the ToBI system, there is obviously no sense in which they can be phonologically contrastive in American English. However, the research on intonational phonology by Grice (1995) in Southern British English and by Grice and Baumann (2002) and Grice, Baumann and Benzmüller (2005) in standard German does provide evidence that there may indeed be a contrast between H+L* and H+!H* and it is this aspect that we wish to explore for German using acoustic and perceptual techniques.

Some examples of this contrast are presented in the German GToBI training materials (Benzmüller, Grice and Baumann URL2). In both H+L* and H+!H*, there is a pitch peak due to the H+ that precedes the accented vowel: the difference between these pitch accents 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 that the target step is from high to mid in H+!H* but not in H+L*. This is further illustrated in Fig. 1 which shows an H+L* pitch accent on *schräg* ('diagonally') in the left panel and an H+!H* pitch accent on *schön* ('pretty') in the right panel. There is a marked step down in pitch from the preceding syllable to a pitch-trough that occurs during the [E] of *schräg*; on the other hand, there is no evidence of a trough in the [2] vowel of *schön* and the pitch does not reach the bottom of the speaker's range as it does for H+L* in *schräg*.

If we allow the system to contrast H+L* and H+!H* pitch accents, then this means that the frequency scaling of the (starred) tones that are associated with the accented syllable is no longer phonetic as argued in e.g., Pierrehumbert (1980), but is implicitly phonological because there is now a paradigmatic contrast between *three* tonal levels: high-star in (L+)H*, low-star in H+L* and effectively a mid-star in H+!H*. Furthermore, H+!H* represents a departure from the phonotactics of downstep discussed earlier, since – in contrast to the other downstepped accents – there is no obligation for the H+!H* to follow a pitch accent (as a result of which, H+!H*, unlike !H* or L+!H*, could be the first and indeed only pitch accent in a phrase).

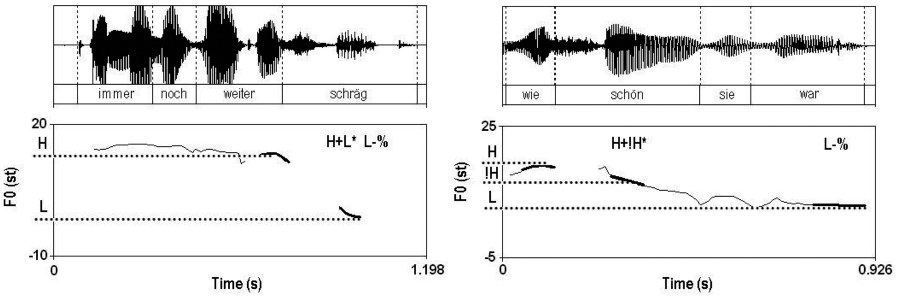

Figure 1. Synchronized time-waveform and f0-trajectory with word labels (top) and GToBI labels of nuclear accents (bottom) taken from Benzmüller, Grice, and Baumann (URL 2) of two German utterances: ... *immer noch weiter schräg.* ('still further along diagonally') and ... *wie schön sie war.* ('how pretty it was').

There is some difficulty in establishing the meaning differences associated with these pitch accents. Baumann (pers. comm.) suggests that H+L* might convey a reassuring, soothing tone of voice in contrast to the greater matter-of-factness or neutrality of H+!H*. In a previous study (Rathcke and Harrington 2006), we assumed the general semantic difference to be that between general/polite

(H+!H*) and resolute/self-evident (H+L*) statements, but we were not able to verify any of these hypothesised meaning-tune associations empirically. Furthermore, our previous study showed that it was very difficult to find any semantic context differring between the usage of H+!H* and H+L*.

In other models of the intonation of standard German (e.g., Féry 1993; Kohler 1997), there is clear evidence for a contrast between H* and H+L* (or H*L vs. H H*L in Féry's analysis), i.e. for a contrast of timing in which the pitch peak is synchronised near the temporal midpoint of the vowel in H* as opposed to its much earlier synchronisation with the immediately preceding syllable in H+L*, but there have been no suggestions to our knowledge that there is a three-way phonological contrast between H* and what would be two types of early peaks, H+L* and H+!H*. We therefore have to consider the possibility that the distinction between H+L* and H+!H* may not be phonological but phonetic, i.e. predictably related to factors such as phrasal position, number of syllables and speech rate.

Our basis for assessing whether the distinction between two pitch accents is categorical is firstly that native listeners can hear and reproduce the distinction and secondly that it is associated with semantic differences. Since, for the reasons outlined above, it is difficult to be certain about the type of meaning difference that is likely to be associated with the pitch accent distinctions being investigated, the present paper draws upon a semantic differential test for this purpose (Osgood, Suci, and Tannenbaum 1957; Uldall 1964; Dombrowski 2003; Ambrazaitis 2005), in which listeners were asked to make judgements on various semantic scales. We also carried out an imitation test of the kind used in Pierrehumbert and Steele (1989) in which subjects imitate tokens from a continuum spanning the tonal categories under investigation.

We used the H* pitch accent as a control stimulus and created a continuum including f0-trajectories of the three pitch accents: H*, H+!H* and H+L*. If the distinction between H+!H* and H+L* is categorical, then listeners should be more likely to perceive, and to produce, between- rather than within-category differences. On the other hand, subjects' responses will be bimodal due to a contrast between H* vs. a collapsed H+!H*/H+L* category, if the latter are just phonetic variants of a single 'early peak' category. Similarly, we would expect three different types of responses to the continuum in the semantic differential test if all three pitch accents are phonologically distinct.

2. Experiment

2.1. Stimuli

We created a synthetic continuum in the f0-scaling domain spanning three pitch accent categories, H*, H+!H* and H+L* with the H* end of the continuum as control stimuli. We chose to make our continuum span three rather than just two (H+!H*, H+L*) categories, because we wanted to make sure that there was at least one unequivocal contrast represented in the stimuli (i.e., H* vs. H+L*).

The test sentence *Sie mag Bananen* ('she likes bananas') was read several times by a female speaker as a simple statement with a nuclear accent on the medial rhythmically strong syllable *-na-* in *Bananen*. The production in which she produced the final weak syllable *-nen* with a schwa and in which the pitch accent was clearly H+L* was chosen as the basic item for resynthesis purposes. This was done primarily because our pilot study (Rathcke and Harrington 2006) had shown that tokens of *Bananen* with no final schwa vowel were often imitated with a creaky voice extending from some part of the accented syllable of *Bananen* to its offset, thereby making f0 difficult to measure.

The construction of the continuum was based on the idea, proposed by Grice *et al.* (Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005), that the main differences between H*, H+!H*, and H+L* are in the *f0-height of the accented vowel*. The continuum was therefore designed to extend over three categories from H* through H+!H* to H+L* by incrementing f0-height variations in the accented vowel. More specifically, we created a stylised H+L* contour from the subject's actual production of H+L* by making the f0-contour fall over the accented syllable /naː/ in a straight line from the onset of the initial [ν] at 210 Hz to 140 Hz at a point 1/3 of the way into the vowel. For this stylised H+L* (st_{16} in Figure 2), the f0 was kept level at 140 Hz to the end of [aː]. Two points (at the onset of the nuclear syllable and at the offset of the following nasal) remained fixed in time and frequency for all synthesised tokens. The continuum was synthesised by raising the f0-level in fifteen 0.5 semitone steps over the middle 1/3 section of the vowel and then by connecting the beginning and end of this raised section with straight lines as shown in Figures 2 and 3. This procedure resulted in 16 stimuli. We assumed that the first stimuli (st_{01}-st_{06}) corresponded to H* given that the f0 drop from the initial /ν/ of *Bananen* (time point *ref1* in Figure 2) to the midpoint of /ɑː/ was less than 3 st, i.e. generally less than that required for an tonal fall to be perceived ('t Hart 1981). The stimuli st_{07}-st_{11}represented clear falls of between 3 and 5 st from the preceding high peak at ref1 in Figure 2 to a mid ranged f0-area in the accented vowel (i.e., H+!H*). The stimuli st_{12}-st_{16} corresponded to the acoustic form of the H+L* pitch accent defined as a steep

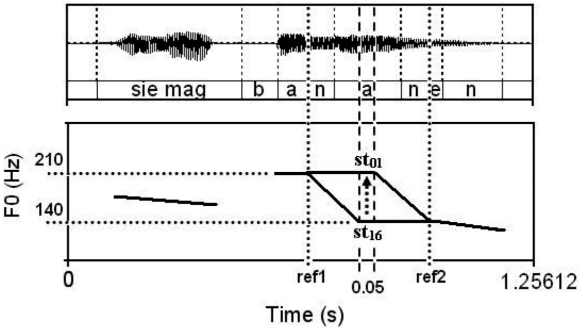

Figure 2. Time-waveform with word and segment labels of the test sentence (top) synchronized with f0-trajectories of the first (st_{01}) and the last (st_{16}) stimulus of the continuum. For all tokens from the continuum, the time and frequency points marked by *ref1* and *ref2* on the time axis were fixed. The vertical dashed lines mark the interval over the middle 1/3 section of the vowel that was raised in sixteen 0.5 semitone steps.

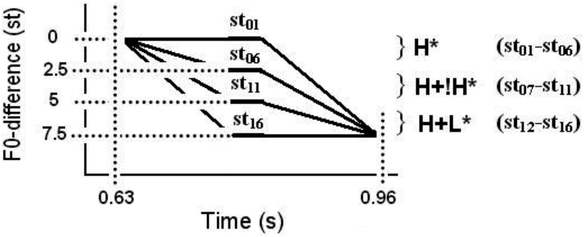

Figure 3. Schematic outline of the components that make up the stimulus continuum and their presumed association with the pitch accents shown on the right.

fall of between 5.5 and 7.5 st to the very low f0-values near the bottom of the speaker's range (see also Figure 3).

2.2. Semantic scales

The semantic differential measures subjects' reactions to stimuli in terms of ratings on bipolar scales typically defined with contrasting adjectives at each end. The list of semantic scales resulting in a semantic differential is primarily chosen based on hypotheses about the meaning of the categories under investigation. As has been shown in various studies (e.g., Heise 1970), ratings on bipolar scales tend to be correlated, and the three basic dimensions of response (labelled 'evaluation', 'potency' and 'activity') account for most of the co-variation in ratings. By using a few pure scales like ‘good-bad’ for evaluation, ‘powerful-powerless’ for potency, and ‘fast-slow’ for activity, reliable measures can be obtained of subjects’ overall attitudinal responses to different stimuli.

In contrast to the earlier semantic differential studies (e.g., Uldall 1964) which were characterised by an *a posteriori* factor analysis for the purposes of extracting the underlying semantic dimensions, a complete semantic differential including the basic semantic dimensions was constructed *a priori* (see e.g., Dombrowski 2003 and Ambrazaitis 2005 for more details). For the present investigation, eight scales were created. The choice of scales was based on the following considerations:

1. Scales 1–3: These correspond to meaning contrasts that have already been demonstrated for distinctions analogous to H* vs. H+L* (e.g., Kohler 1987; 2005). So, these scales were selected as the control condition.
2. Scale 4: The choice of this scale was inspired by the frequency code (Ohala 1984): perhaps H+!H* is used paralinguistically to avoid very low f0-values (which are assumed to be typical for H+L* pitch accents). As is well known (e.g., Gussenhoven 2002), low tones tend to convey speakers’ dominance or to signal unalterable facts in a great number of languages.
3. Scale 5: As outlined in the introduction, this scale’s semantic distinction has already been hypothesised as one of the ways that meaning differences between H+!H* and H+L* may be conveyed.
4. Scales 6–8: These were added to complete the semantic differential with the three basis dimensions (evaluation, potency, activity, respectively). A relevant factor for the choice of contrasting adjectives for scale 6 is that Grice *et al.* (Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005) consider ‘politeness’ to be one of the meanings associated with the H+!H* vs. H+L* distinction. The semantic contrasts chosen for the scales 7 and 8 were thought to be good exponents of ‘potency’ and ‘activity’ in dialogues.

The scales are shown in Table 1. Taking into account the meaning distinctions associated with the frequency code (Ohala 1984) and what is known about how meaning differences are conveyed by the distinctions between H*, H+!H* and H+L* (Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005; Kohler 1987, 2005), the negative pole of each scale was assumed to be more appropriate for H+L* pitch accents, the positive pole in scales 1–3 for H*, and the positive pole of scales 4–8 for H+!H* pitch accents. Scales 1–3 are associated primarily with differences of linguistic meaning, while the other scales express predominantly differences of paralinguistic meaning (cf. Ladd 1996: 33–36).

The scales were embedded in a carrier sentence *‘The speaker sounds like . . . ’.* The instructions to the subjects for judging the scalar values made it clear that

Table 1. Eight pragmatic scales chosen for the perception test with semantic differential. German terms give the exact meaning of the scale poles used in the test.

Scale	Negative pole (−)		Positive pole (+)	
no.	German term	English translation	German term	English translation
1	*wissend*	knowing	*erkennend*	realising
2	*abschließend*	concluding	*diskussions-eröffnend*	opening a discussion
3	*akzeptierend*	accepting	*kontrastierend*	refusing
4	*resolut*	out of the question	*kompromissbereit*	ready to compromise
5	*beruhigend*	calming	*ermunternd*	encouraging
6	*unhöflich*	impolite	*höflich*	polite
7	*sicher*	certain	*unsicher*	uncertain
8	*gelangweilt*	bored	*interessiert*	interested

we were interested in potential meaning differences evoked by the speaker's melody. Although semantic judgements can be influenced by the morpho-syntax and lexis of an utterance, we decided to use only one sentence for the semantic differential test in order to limit the length of the test to 30 min.

2.3. Procedures

A tape containing 128 stimuli was created from eight repetitions of each of the 16 stimuli. Thus each stimulus was paired once with each scale and presented in randomised order. The tape was played to the subjects who had to mark their answers on a sheet of paper containing the scales. They also had to rank the utterance between ±3, i.e. between the extreme negative and positive ends of the scale. They were told to interpret these values as follows: 0 was 'neutral', ±1 was 'slightly' (e.g., assigning -1 on scale 8 corresponds to 'slightly bored'), ±2 was 'quite', and ±3 'extremely'.

The subjects were instructed to wait for the auditory prompt preceded by a beep and to judge spontaneously on the given scale how they thought the speaker sounded. They did this for each stimulus separated by a 4 s pause. The perception task was preceded by a trial session including eight stimuli presentations (one for each scale) to familiarise the subjects with scales and with the procedure. After the trial session the subjects had the opportunity to ask questions. The stimuli were presented from a CD-player in a sound treated room at the IPDS Kiel.

For the imitation test, the sixteen stimuli detailed above were each copied ten times resulting in 160 items. These 160 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 stimuli were presented to subjects via headphones. The imitation test was carried out in a sound treated recording studio at the IPDS Kiel.

Twelve speakers of standard North German, five M and seven F between 20 and 22 years of age without known speech or hearing disorders participated in the experiment. All subjects were beginner students of phonetics at the IPDS Kiel. None of the subjects had any experience in prosody nor was told the purpose of the experiment. All subjects were paid for participation. The experiment (containing both semantic and imitation test) took about one hour.

2.4. Results

2.4.1. Semantic Differential

If there is a separate range of meanings conveyed by H+!H* as opposed to H* and H+L*, then we would expect to find a marked judgement profile for stimuli 7–11 on any of the scales 4–8: that is, these stimuli should get high positive or negative scale values whereas stimuli 01–06 as well as 12–16 should be labelled ‘neutral’ or with scale values opposite to those from stimuli 7–11.

The evaluation of the 16-point continuum is shown in Figure 4: the x-axis represents the stimulus number (01–16) and the y-axis is the scale value (seven points). The data for each scale were analysed separately with a repeated measures analysis of variance using SPSS (Brosius 2002). The f0-level in the accented vowel (i.e., stimulus number) was the independent variable and the judgement score the dependent variable. We used a repeated measures ANOVA because the subjects judged more than one stimulus within one analysis. The Greenhouse-Geisser correction for non-homogenity of variances was applied (Leonhart 2004). These statistical results are summarised in Table 2. The significance level (alpha) was set at 1% since the sample was only n = 12.

There were significant differences in subject responses on five out of eight scales, i.e. on scales 1–3 as well as on scales 5 and 8 (Figure 4). As expected, scales 1–3 showed changes of judgement that are compatible with the semantic differences for the distinction between ‘medial’ and ‘early’ peaks (Kohler 1987 and 2005), corresponding roughly to the H* vs. H+L* difference in an AM analysis. Scale 5 shows similar results for scales 1–3, i.e. the meaning ‘soothing’ shows differences primarily between the beginning and end of the continuum, and hence between H* and H+L* and not – as hypothesised – between the part of the continuum corresponding to the H+!H* vs. H+L* distinction (in which case, we would have expected a major difference between the middle and the end of the continuum). The same is true for scale 8 although the profile of judgments

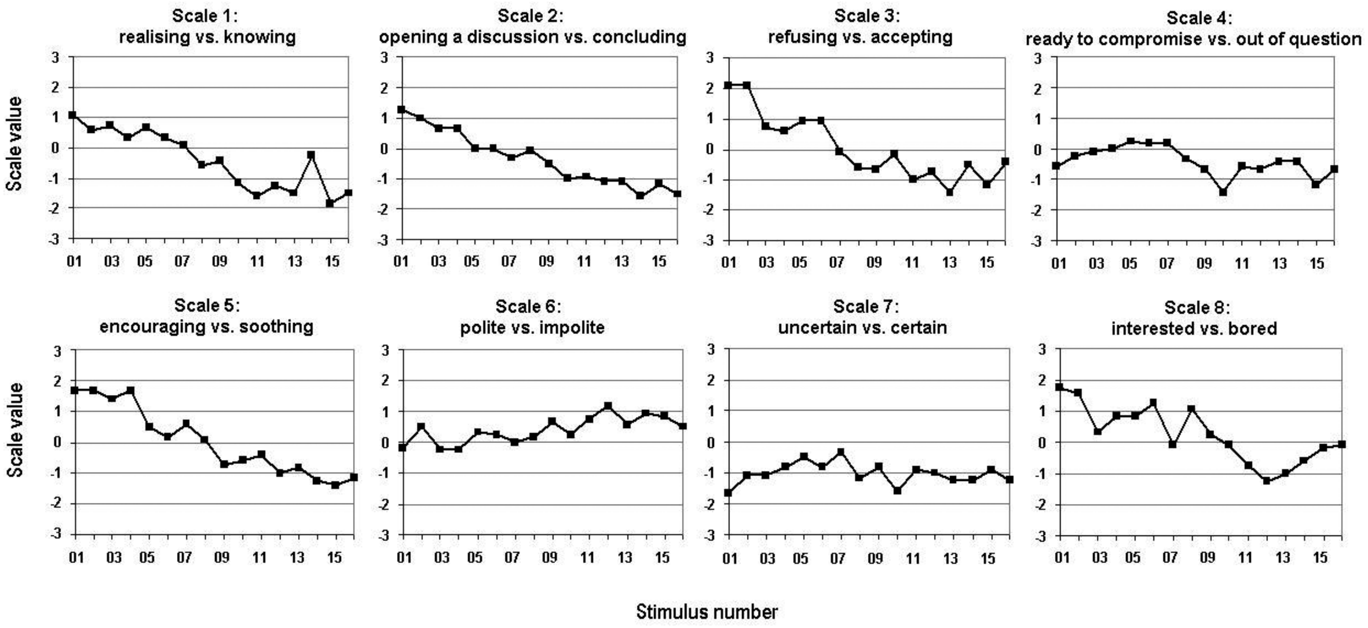

Figure 4. Mean judgements (n = 12) for the stimuli st_{01}–st_{16} on the eight scales: x-axis is stimulus number, y-axis is the scale value. The left term above each scale is associated with positive scale values.

Table 2. Results of eight repeated measures ANOVAs examining the effect of eight scales on the perception of a meaning change during the tonal continuum (alpha = 1%): F-values, adjusted degrees of freedom (d.f.), and significance levels.

Scale	F	d.f.	p
Scale 1	4.855	5.434	< 0.001
Scale 2	4.319	5.423	< 0.01
Scale 3	7.414	5.078	< 0.001
Scale 4	1.082	3.664	n.s.
Scale 5	7.389	5.706	< 0.001
Scale 6	1.179	6.219	n.s.
Scale 7	0.757	4.934	n.s.
Scale 8	6.544	5.346	< 0.001

here is not as marked as for scales 1–3 and 5. There were no category-dependent variations on scales 4, 6, or 7.

To test for the possibility of a *three*-way contrast, we ran the following additional analysis. We used paired-sample t-tests to test for significant differences within the first and within the last parts of the continuum which we assume to correspond respectively to the category distinctions between H* vs. H+!H* (st_{01}-st_{06} vs. st_{07}-st_{11}) on the one hand and H+!H* vs H+L* (st_{07}-st_{11} vs. st_{12}-st_{16}) on the other. These results are summarised in Table 3.

The results for scales 1, 3, 5 showed significant differences for H* vs H+!H*, but not for the H+!H* vs H+L* stimuli: that is, the subjects could reliably differentiate only between the stimuli from the beginning (H*) and the rest (H+!H*/L*) of the continuum. The results for scales 2 and 8 show that both parts (i.e., H* vs. H+!H* as well as H+!H* vs. H+L*) could be distinguished significantly. However, the mean differences on the scales are always greater for H* vs. H+!H* than for H+!H* vs. H+L* (Table III), suggesting that the locus of the main change in perceived meaning was between H* and H+!H* stimuli.

In summary, the first 5–6 points of the stimulus continuum seem to be compatible with the semantic interpretation of a H* pitch accent, the last 5 stimuli are compatible with the semantics of an early peak accent, while about five stimuli from the middle of the continuum (st_{07}-st_{11}) mark a transition region in which there is a greater variation in semantic judgements. There is no evidence for a separate category formation in the middle of the synthetic continuum which could be labelled H+!H*.

Table 3. Results of t-tests for paired samples (alpha = 1%): mean and t-values, degrees of freedom (d.f.), and significance levels.

Scale	Comparisions	mean	T	d.f.	p
Scale 1	H* vs. H+!H*	1.36	3.743	11	< 0.01
	H+!H* vs. H+L*	0.53	2.547	11	n.s.
Scale 2	H* vs. H+!H*	1.16	3.148	11	< 0.01
	H+!H* vs. H+L*	0.72	3.874	11	< 0.01
Scale 3	H* vs. H+!H*	1.66	7.456	11	< 0.001
	H+!H* vs. H+L*	0.41	2.244	11	n.s.
Scale 4	H* vs. H+!H*	0.48	0.982	11	n.s.
	H+!H* vs. H+L*	0.10	0.437	11	n.s.
Scale 5	H* vs. H+!H*	1.31	3.746	11	< 0.01
	H+!H* vs. H+L*	0.92	2.886	11	n.s.
Scale 6	H* vs. H+!H*	−0.30	−1.073	11	n.s.
	H+!H* vs. H+L*	−0.43	−1.753	11	n.s.
Scale 7	H* vs. H+!H*	−0.03	−0.103	11	n.s.
	H+!H* vs. H+L*	0.17	0.573	11	n.s.
Scale 8	H* vs. H+!H*	0.89	3.098	11	< 0.01
	H+!H* vs. H+L*	0.70	3.487	11	< 0.01

2.4.2. *Imitation*

To support the hypothesis that there is a three-way distinction, f0 is expected to fall across the sequence of weak-strong syllables in /baˈnaː/ but that the fall is expected to be great in H+L* and smaller in H+!H*. In H* by contrast, the pitch is likely to be either level, or rising. If subjects are able to produce three categories in the imitation test, then f0 across /baˈnaː/ should show evidence of a trimodal distribution: across these syllables it should be level or rising for H*; steeply falling for H+L*; and slightly falling for H+!H*.

The production data from 12 subjects who participated in the previous listening experiment were digitised and labelled in EMU (Bombien, Cassidy, Harrington, John and Palethorpe 2006) after calculating f0, The data from two subjects were removed from the evaluation: one subject had creaky voice and the other produced all the stimuli on a monotone. So, the following results are based on the analysis of imitations from 10 subjects.

The f0-values in Hz, f_{Hz}, were converted into semitones, f_{st}, using the formula:

$$F_{st} = 12(\log_2 f_{Hz} - \log_2 k)$$

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 /baˈnaː/-tokens (160 tokens per speaker). The above formula sets each speaker's mean f0-value to 0 st and thereby acts as a speaker-normalization of the f0-data. In order to test for a trimodal distribution, we measured the f0 semitone difference between (1) the vowel midpoint of the first syllable /ba/ and (2) the midpoint of the following accented vowel in /ˈnaː/ for all imitations produced by all 10 speakers in the database.

The results of this semitone difference from all 10 subjects are shown in the histogram in Fig. 5. The histogram on the left pools their f0-differences between the midpoint of the first syllable /ba/ and the midpoint of following accented syllable /ˈnaː/. Negative values on the x-axis indicate that f0 in /ba/ was lower than in /ˈnaː/ (i.e., a rising pitch) while positive values indicate that f0 in /ba/ was higher than in the following /ˈnaː/ (i.e., the pitch was falling). This histogram on the left in Figure 5 provides some evidence for a bimodal, but not for a trimodal, distribution.

The histogram on the right shows mean normalised f0-differences per stimulus, pooled across subjects. The normalisation was arrived at by averaging the f0-differences over all stimuli for each subject, and subtracting this mean from each of the subject's f0-differences per stimulus. (So if for any token, the result of this subtraction is zero, then the f0-change for that token from /ba/ to /ˈnaː/ was equal to speaker's average f0-change across these syllables). In this way, we could assess which synthetic stimulus numbers (from st_{01} to st_{16}) were associated with a greater than average f0-drop from /ba/ to /ˈnaː/ and which with less. As the histogram on the right shows, the first four stimuli were all imitated with approximately the same f0-drop that was much *less* than the speaker's average f0-change across all stimuli. The figure also shows that the last 4–5 stimuli were imitated with about the same f0-drop that was much *greater* than the average f0-change.

Taken together, these results seem to be indicative of a two-way categorical distinction. Thus, there is clear evidence of differences in imitations to stimuli 1–4 in comparison with stimuli 12–16, but very little evidence of differences in imitations *within* stimuli 1–5 or *within* stimuli 12–16. Moreover, there is more or less an S-shaped progression in the height of the bars across stimuli 1–16 that is very reminiscent of the way in which listeners categorise a synthetic continuum into two categories in segmental perception experiments. Thus, it seems that listeners' imitations are consistent with their perception of two, but not three, tonal categories. We would suggest that one of these is likely to be H* in which the f0-drop from the first to the second syllables of /baˈnaː/ is not present or minimal, as it is in stimuli 1–4; and the other is likely to be some kind of early

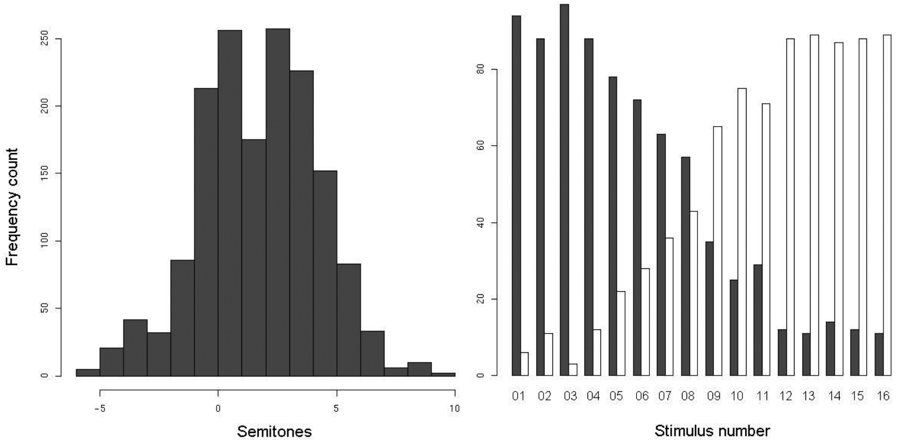

Figure 5. Left: Histogram of P – N where P is the f0 in semitones at the temporal midpoint of initial, prenuclear to [a] in [ba] and N is the f0 in semitones of nuclear [a] in [ˈnaː] of *Bananen* pooled across all speakers. Right: Histogram for all 10 speakers together of $P_k - N_k - m_k$ where P_k and N_k are the same as P and N as defined above but for speaker k, and m_k is the mean of $P_k - N_k$ ($k = 1, 2, \ldots$ 10 speakers).

peak, most likely H+L* in which the f0-drop is large, as it is in stimuli 12–16. But we do not find much evidence in Figure 5 in favour of a division of the synthetic continuum into three tonal categories.

3. Acoustic analysis of production data

Since the results of the perception and production tasks were mostly consistent with a two-category distinction between a mid and some form of early peak, we sought evidence from a large corpus of standard German (the Kiel Corpus of Read Speech, IPDS 1994) for whether phonetic factors might contribute to the perception (and annotation) of f0-contours with H+!H* as opposed to H+L*. The results of this investigation so far for two speakers show firstly that when some form of early pitch accent is in phrase-final position, there is a steep drop to a low f0-value (that is likely to be augmented by the L-phrase tone); and secondly that when early pitch accents are non-phrase final, i.e., when there are syllables following the nuclear accent in the tail, then the fall across the nuclear accented syllable tends not to be quite so steep, nor to fall to such a low value. Compatibly, there are no examples in the GToBI training materials where H+!H* is in an

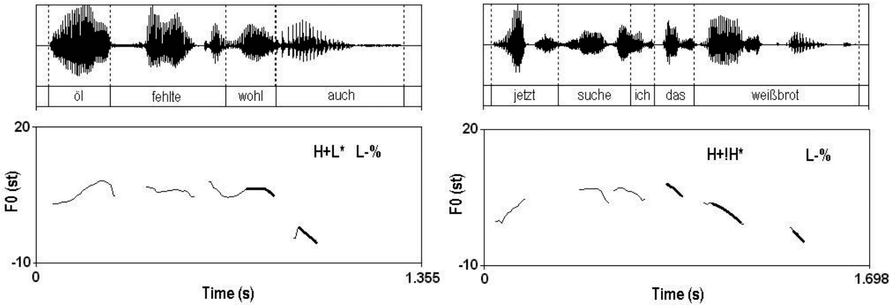

Figure 6. Synchronized time-waveform and f0-trajectories showing pitch accent labels of two read German utterances produced by a male speaker of standard North German (left) of *Öl fehlte wohl auch.* ('oil was missing too') and (right) *Jetzt suche ich das Weißbrot.* ('now I'm looking for white bread'). Tonal labels are in accordance with the current guidelines of GToBI (Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005).

intonationally phrase-final position and only one instance where it is final in an intermediate phrase. Conversely H+L* occurs very often phrase-finally. Thus we would suggest that the extent and rate of the f0-fall of early pitch accents is phonetically conditioned by their position in the phrase and that it is this phonetic factor that has led some to postulate two distinct pitch accents H+!H* and H+L*. In Figures 6 and 7, these potential phonetic differences are illustrated by the different f0-levels in the nuclear accented syllables. For example, in the

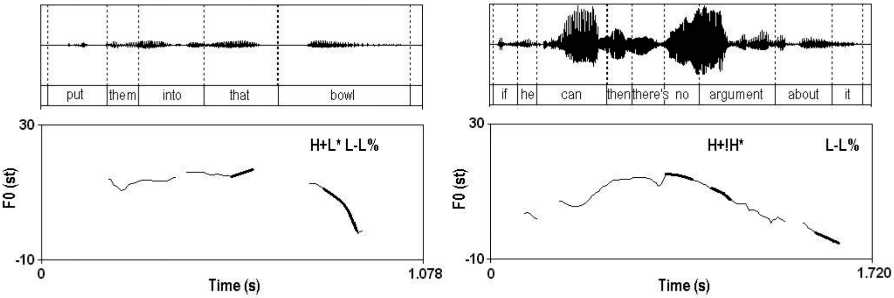

Figure 7. Synchronized time-waveform and f0-trajectories showing pitch accent labels of two English utterances produced by a female speaker (left) ... *put them into that bowl* and by a male speaker (right) ... *if he can then there's no argument about it.* Both examples are from American English ToBI training materials (Beckman and Hirschberg URL1).

left panel of Figure 6, there is an early peak on phrase-final nuclear-accented *auch* ('also'), which would be H+L* in GToBI and f0 falls to a low level in the speaker's range. Although phrase-*medial Weißbrot* ('white bread') in the right panel of the same figure is nuclear accented (on *Weiß*, 'white') and although it is evidently also produced with an early pitch accent, f0 does not fall to the base of the speaker's range as it does for *auch*. So we would argue that these two early peaks are phonetic variants of the same phonological pitch accent H+L*: that is, the early pitch accent is realised phonetically as H+L* in *auch*, but as H+!H* in phrase-medial *Weiß*.

The same kind of relationships holds in comparing American English examples *bowl* with *argument* in Figure 7 (the underlining indicates the rhythmically strongest syllable). Both are phonologically early pitch accents, but the fall is steeper and to a lower value in *bowl* in which the early pitch accent occurs in phrase-final position.

We carried out a pilot experiment to test the hypothesis that there was a link between the number of unaccented syllables following the nucleus and the f0-level placed within the nucleus suggested by the examples given above. We elicited productions of early peak-accents on the German target surnames *Lie, Liener* and *Lienerer* when these were nuclear accented in the context sentence *Das war Herr . . .* ('That was Mr . . . '). Since the target words all have primary lexical stress on the first syllable, then they differ in nuclear accented position in the number of syllables (0, 1, 2 respectively) in the tail, i.e., in the number of post-nuclear syllables to the right edge of the phrase. So far, we have analysed data from 10 speakers of standard North German (3 M and 7 F).

The f0-data of all speakers were normalised, smoothed, and averaged following the procedure in Rathcke and Harrington (2006). The resulting averaged contours for these three words in the left panel of Figure 8 show a somewhat steeper fall for *Lie* (zero syllables in the tail) than for *Liener* or *Lienerer* (one and two syllables in the tail). This is what we would expect if the realisation of an early pitch accent is phonetically dependent on the number of post-accentual syllables. That is, the f0-contour of *Lie* falls more rapidly to the phrase edge than in *Liener* or *Lienerer* because in the monosyllable *Lie*, there is less time in which the falls can place. We would also expect the f0-level to be a good deal lower in the nuclear accented /iː/ vowel in *Lie* for the same reason. This is because, if f0 drops more rapidly in *Lie*, then it is also more likely to have reached a lower value at the temporal midpoint of /iː/ than in either *Liener* or *Lienerer*: just this is supported by the data in the right panel of Figure 8 showing the f0-level at the midpoint of /iː/ for these three words. Taken together, these data in Figure 8 support the idea that an early pitch accent is phonetically variable depending on the number of post-accentual syllables: the more syllables in the tail, the

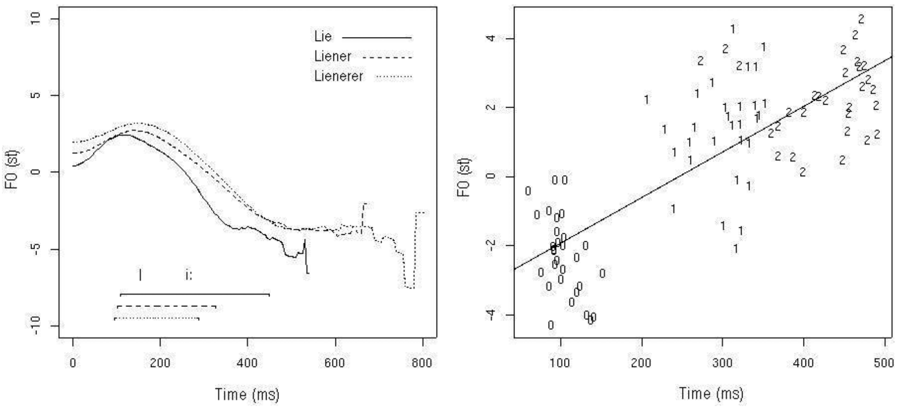

Figure 8. Left: Averaged and normalized f0-curves of the test words, synchronized at the onset of word *Herr* ('Mr.') in *Herr Lie/ Liener/ Lienerer* (three repetitions of each item for 10 subjects; n=30). The average durations of the nuclear accented syllables are given schematically by the length of the horizontal solid, dashed, and dotted lines in the same panel. Right: f0-values at the nuclear accented vowel's temporal midpoint as a function of the duration between this measurement point and the following phrase boundary, for different counts of postnuclear syllables (0, 1, 2).

flatter the fall (Figure 8, left panel) and therefore the higher f0 is in the nuclear accented vowel (Figure 8, right panel).

4. Discussion

In the preceding sections, we reported a number of experiments which were designed to clarify the question of whether there is a two-way distinction of early pitch accents in standard German as proposed by Grice *et al.* (Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005). The results suggest that, as reported elsewhere (e.g., Féry 1993, Kohler 1997), there is only a single *phonological* category distinction of German falling f0-contours spanning medial and early peaks. The subjects could neither produce the difference between H+!H* and H+L* nor label their different meanings. In the light of these results, it seems difficult to maintain the idea of a phonological contrast between H+!H* and H+L* in standard German while the evidence from the corpus analysis and from the pilot study point to a *phonetic* effect that depends on whether or not the nuclear accented syllable is phrase-final.

Two possible mechanisms might contribute to the lower f0 of early pitch accents when they are in (intonationally) phrase final syllables. Firstly, phrase-final lowering (Liberman and Pierrehumbert 1984) is likely to cause f0 to be even lower than would be predicted by declination when the nuclear accented syllable is phrase-final. Secondly, the observed f0-compression under time pressure as well as a tendency for the target to be aligned earlier (i.e., left-shifted) when there is an immediately following phrase boundary (e.g., Schepman, Lickely and Ladd 2006) affect the realisation of the f0-fall leading to lower f0-targets in the phrase-final nucleus.

Independently of these considerations, there are both theoretical as well as empirical reasons why we would propose not including H+!H* as a phonological tonal category in standard German. The transcription H+!H* implies some kind of downstepped tone on the one hand, but on the other it is being used in a way that is not consistent with the function of downstep in an AM model. The inconsistency comes about because there are quite tight phonotactic restrictions on the occurrence of downstepped tones like !H* and L+!H*: above all, they must always be preceded by a non-downstepped pitch accent and the only type of pitch accent that can follow a downstep is another downstepped pitch accent in the same intermediate or intonational phrase. In the present study, we only investigated phrases with single nuclear pitch accents, with the aim of clarifying whether there is a paradigmatic contrast in intonational phonology of German. But we also looked for some examples from the Kiel Corpus of Read Speech of phrases containing several accented words. Figure 9 gives two examples for a

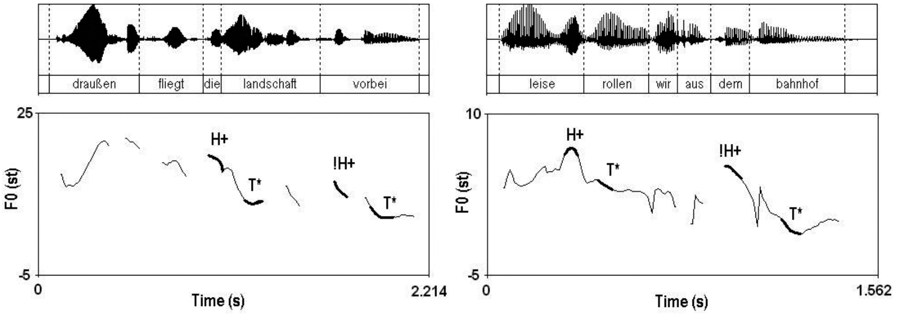

Figure 9. Synchronized time-waveform and f0-trajectories showing f0-peaks (H+) and starred tones (T*) in sequences of two early pitch accents in read German utterances produced by a female (left) and a male (right) speaker of standard North German: *Draußen fliegt die Landschaft vorbei.* ('the countryside is rushing past outside') and (right) *Leise rollen wir aus dem Bahnhof.* ('we're rolling out of the station quietly').

female (left panel) and a male (right panel) speaker of standard German (cf., the comparable example in Féry 1993: 103). It is clear from these examples that there is a sequence of early pitch accents which we will label for the moment as H+T*. If T* were a downstepped tone, then it would cause the H+ of the following tone to be downstepped. But this is not what we have found. Instead, we found that it is the H+ components of successive accented syllables that are downstepped relative to each other. This is entirely consistent with the examples of the terraced f0-contour that is characterised by Pierrehumbert (1980: 154) as a succession of H+L* pitch accents. If we want to account for the relatively high f0-level at the non-final accented syllable, then we should abandon the downstepped label in H+!H* and just transcribe it H+M*. However, introducing M* means that the system is founded on a tritonal contrast and this is a radical departure from AM's two-tone system plus downstep (Gussenhoven 2004: 104–105).

Furthermore, including bitonal downsteps begs the question of why the inventory should not also be expanded to include bitonal *up*steps like H+ˆH* or H*+ˆH (an upstepped boundary tone H–ˆH% is available in GToBI, see Grice and Baumann 2002; Grice, Baumann and Benzmüller 2005). In German, nuclear upsteps are common in phrases with several accents (Truckenbrodt 2002). Indeed, if we argue that the domain of downstep or upstep is both syntagmatic and paradigmatic, then this would be tantamount to expanding the system of contrasts from a two-tone, to a multi-tonal, paradigmatic system like that of Pike (1945) or Liberman (1975); yet just this type of multi-tonal paradigmatic system was heavily criticised by Bolinger (1952) and Liberman and Pierrehumbert (1984) in arguing for the primacy of the two-tone autosegmental-metrical model of intonation.

The problem discussed here is also related to the defintion of L and H tones. Obviously, there are at least two possible ways of defining high and low tones: (1) paradigmatically, as related to speaker's range and his/her baseline or (2) syntagmatically, as local minima and maxima with their internal reference to each other as f0-turning points during a given phrase. Whereas the first view is represented by Pierrehumbert (1980), the latter is advocated by Bruce (1977). However, the views are not necessarily mutually exclusive. For example, Pierrehumbert (1980) allows some syntagmatic effects in her paradigmatic definition, for example: "*In H+L H, the L's are related to the H in the same accent by the same factor, k, which control downstep*" (1980: 50). At the same time, Bruce (1977) also aimed to set the relative local minima and maxima in relation to the range of a speaker's voice by dividing it into four f0-levels (Bruce 1977: 137). Actually, neither approach precludes the possibility that a low tone (e.g., at the beginning of a sentence) can have the same actual f0-value as a high tone (e.g., at the end of a sequency of downstepped tones). So, since the f0-value of a low tone deviated from the baseline in a predictable way, there no difficulty in principal with marking an f0-target as a low tone even if it is scaled higher than the baseline. But this implies that in German H+L*, the scaling of the L* is not independent of its temporal alignment relative to the phrase boundary.

The aim of the additional pilot study that we have reported on in this paper was to investigate a predominantly linguistic factor, the number of syllables after the nucleus, on f0-realisation. Factors that are more paralinguistic, such as varying the degree of prominence or liveliness also influence tonal realisation. As noticed by Pierrehumbert for English pitch accents (1980: 68), the effect of increasing prominence is to lower and raise f0 of L* and H* pitch accents, respectively. Compatibly, Gussenhoven and Rietveld (2000) have found that lower L* tones as well as higher H* sound more surprised. However, recent results by Grice, Baumann and Jagdfeld (2007) have shown that increased liveliness does not produce a lowering of the starred tone in an early peaked accent – as a result of which they favour an analysis of H+!H* (as opposed to our proposed phonological analysis of early peaks as H+L*). Thus, further analysis is required on this issue. However, the main issue under investigation here is not disputed in Grice, Baumann and Jägerfeld (2007): that there is little evidence that the distinction between H+!H* and H+L* are associated with differences in meaning.

5. Conclusions

Presented experiments suggest that there is a distinction between a mid and an early pitch accent in German, but our results suggest that this distinction is phonetic, not phonological. Preliminary analysis of production data shows that scaling of the starred tone is depending on phonetic factors like right-hand segmental context. At this stage of current research on this issue, we propose that H+L* is the most probable phonological tonal representation of the early peak category. Further empirical research should shed more light on the phonetic properties and their theoretical implication for low tones.

Acknowledgements

This research was supported by German Research Council grant HA 3512/2-1 *'The structure of tonal representations: evidence from tune-text synchronization (STRETTS)'* to Martine Grice and Jonathan Harrington. The authors thank Ernst Dombrowski for methodical and statistical advice, Pilar Prieto and one anonymous reviewer for valuable comments on the first draft of this paper.

References

Ambrazaitis, G.
2005 Between fall and fall-rise: substance – function relations in German phrase-final intonation contours. *Phonetica 62,* 196–214.

Beckman, M.E., and G.M. Ayers
1994 Guidelines for ToBI Labelling, version 2.0. Ohio State University.

Beckman, M. and J. Hirschberg
no year The ToBI annotation conventions. URL1: *http://www. ling.ohio-state.edu /∼tobi/ame_tobi/annotation_conventions.html*

Beckman, M. and J.B. Pierrehumbert
1986 Intonational structure in Japanese and English. *Phonology Yearbook3,* 255–310.

Benzmüller R., M. Grice and S. Baumann
no year Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. Version 2 (überarbeitete Fassung). URL2: *http://www.uni-koeln.de/phil-fak/phonetik/gtobi/guidelines-version2c.html*

Bolinger, D.
1951 Intonation: levels versus configurations. *Word 7,* 199–210.

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

Brosius, F.
2002 SPSS 11. Bonn: mit Verlag.

Dombrowski, E.
2003 Semantic features of accent contours: effects of f0 peak position and f0 time shape. *Proceedings of the XVth International Congress of Phonetic Siences*. Barcelona, 1217–1220.

Féry, C.
1993 *German Intonational Patterns*. Tübingen: Niemeyer.

Grice, M.
1995 Leading tones and downstep in English. *Phonology 12,* 183–233.

Grice, M. and S. Baumann
2002 Deutsche Intonation und GToBI. *Linguistische Berichte 191.* Helmut Buske Verlag, 267–298.

Grice, M., Baumann, S. and R. Benzmüller
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.

Grice, M., Baumann, S. and Jagdfeld
2007 Evidence for tonal identity from peak scaling under pitch span variation. *Proceedings of the XVIth International Congress of Phonetic Siences*. Saarbrücken.

Gussenhoven, C.
2002 Intonation and Interpretation: Phonetics and Phonology. In: *Proceedings Speech Prosody 2002*. Aix-en-Provence. URL2 : *http://www.lpl.univ-aix.fr/sp2002/pdf/gussenhoven.pdf*

Gussenhoven, C.
2004 The Phonology of Tone and Intonation. Cambridge University Press.

't Hart, J.
1981 Differential sensitivity to pitch distance, particularly in speech. In: *Journal of Acoustical Society of America 69 (3),* 811–821.

Heise, D.R.
1970 The semantic differential and attitude research. In G.F. Summers (ed.) *Attitude Measurement*. Chicago: Rand McNally, 235–253.

IPDS
1994 *The Kiel Corpus of Read Speech.* Volume 1. CD-ROM No. 1. IPDS, Kiel.

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

Kohler, K.J.
1997 Modelling prosody in spontaneous speech. In Y. Sagisaka, N. Cambell, and N. Higuchi (eds.) *Computing prosody. Computational models for processing spontaneous speech.* N.Y.: Springer, 187–210.

Kohler, K.J.
2005 Timing and communicative functions of pitch contours. *Phonetica 62,* 88–105.

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

Ladd, D.R.
1996 *Intonational Phonology.* Cambrigde University Press.

Leonhart, R.
2004 *Lehrbuch Statistik: Einstieg und Vertiefung.* Göttingen, Hans Huber.

Liberman, M.Y.
1975 *The Intonation System of English.* Ph.D. dissertation, MIT. [Published by Garland Press, New York].

Liberman, M.Y. and J.B. Pierrehumbert
1984 Intonational invariance under changes in pitch range and length. In M. Aronoff and R.T. Oehrle (eds.) *Language, Sound, Structure: Studies in Phonology Presented to Moris Halle by His Teacher and Students.* MIT Press, Cambridge, 157–223.

Ohala, J.J.
1984 An ethnological perspective on common cross-language utilization of f0 of voice. *Phonetica 41,* 1–16.

Osgood, C.E., Suci, G.J., and P.H. Tannenbaum
1957 *The measurement of meaning.* University of Illinois Press, Urbana.

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

Pierrehumbert, J.B. and S. A Steele
1989 Categories of tonal alignment in English. *Phonetica 46,* 181–196.

Rathcke, T. and J. Harrington
2006 Is there a distinction between H+!H* and H+L* in standard German? Evidence from an acoustic and auditory analysis. *Proceedings Speech Prosody 2006,* Dresden, 783–786.

Schepman, A., Lickely, R. and D.R. Ladd
2006 Effects of vowel length and “right context” on the alignment of Dutch nuclear accents. *Journal of Phonetics 34,* 1–28.

Silverman, K. and J.B. Pierrehumbert
1990 The timing of prenuclear high accents in English. *Laboratory Phonology 1,* 72–106.

Truckenbrodt, H.
2002 Upstep and embedded register levels. *Phonology 19,* 77–120.

Uldall, E.T.
1964 Dimensions of meaning in intonation. In Abercrombie, Fry, MacCarthy, Scott, Trim (eds.) *In honour of Daniel Jones.* Longman, London, 271–279.