Tonal association and derived nuclear accents—The case of downstepping contours in German

Martine Grice [*], Stefan Baumann, Nils Jagdfeld

IfL Phonetik, University of Cologne, Herbert-Lewin-Strasse 6, 50931 Köln, Germany

Abstract

We report on two production experiments which together provide additional support for treating downstep as orthogonal to the tonal structure of utterances, in the sense that certain intonational meanings are expressed by a downstep relation between two H tones, rather than by a particular accent type or accent sequence. The intonational meanings investigated were inferentially accessible information on the one hand and broad focus on the other. In the first experiment an increase in pitch span provided evidence for an analysis of the early peak accent under investigation as H+!H* rather than H+L*, since the starred tone was raised along with pitch span increase, as would be expected of a high tone. Thus we confirmed the presence of a downstepped tone in this accent. In the second experiment we showed that the distribution of accents favoured a downstep in broad focus utterances, as opposed to narrow or contrastive focus utterances.

Within each of these contexts alternation was found between downstep across accents (e.g. H* !H*) and downstep within an accent (H+!H*). This suggests first of all that downstep applies independently of how the tones are associated (e.g. whether they are the starred tones of two separate accents, or the leading and starred tone of a single accent), and secondly that H* !H* and H+!H* are related and should be analysed as such.
© 2008 Elsevier B.V. All rights reserved.

Keywords: Intonation; Tonal association; Downstep; Early peak; Pitch accent; Focus

1. Introduction

The aim of this paper is to establish a common underlying structure for a number of pitch accent sequences in German which occur in well-defined semantic–pragmatic contexts, as well as to provide further evidence for downstep as orthogonal to the tonal structure of utterances, in

M. Grice et al. / Lingua 119 (2009) 881–905

$$\sigma^*$$


```
                σ*
                |
    - - - - - - - - - - -  ∘*
         ∘             |
        / \            |
       /   \           |
      s     w          |
      H     L      H      (H+L)+H*
```
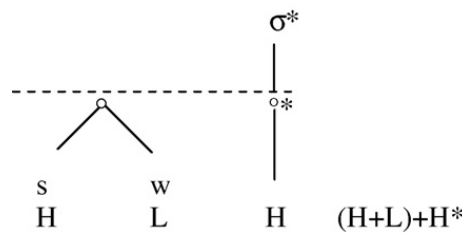
Fig. 1. Example of a downstepped pitch accent with a leading tone (from Grice, 1995:219).

the sense that certain intonational meanings are expressed by a downstep relation between two tones rather than a particular accent type or sequence of accent types.

For an H* tone in a nuclear accent to be downstepped, a previous H tone is required within the same phrase, so that the 'step down' can take place. In the early stages of development of autosegmental-metrical theory (Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986), this previous H tone had to be part of a prenuclear bitonal pitch accent, the whole of which acted as the downstep trigger, analogous to the HL alternations which trigger downdrift in African tone languages, where an underlying HLH sequence leads to a surface H !H* sequence. A later study on downstep in English extended this model to allow for the HLH sequence to occur in an underlying tritonal pitch accent which surfaced as a bitonal pitch accent with a downstep relation across its two tones (Grice, 1995). Thus an underlying (H+L)+H* accent surfaced as H+!H* (see Fig. 1).

However, in other autosegmental-metrical approaches, downstep does not contribute to the identity of accents. Instead, downstep is seen to be orthogonal to the tonal string (Ladd, 1983, 1996), and regarded as a feature or operation applying to H tones (Ladd, 1983; Gussenhoven, 2004). For downstep to take place, however, there have to be two H tones in the phrase. Ladd's later analysis of downstep (1993, 1996) appeals to a syntagmatic relation between the two H tones, represented as a metrical tree (as shown in Fig. 2). In what follows, we refer to this earlier H tone as H1.

Although H1 is typically understood to be in a separate accent, in more recent studies (e.g. Gussenhoven, 2004), a leading H tone can function as H1 too, in which case the accent H+!H* does not have to be derived from an HLH sequence. Instead, there is provision for a separate accent-internal downstep relation. Gussenhoven also discusses the possibility of H1 being an initial high boundary tone, %H (Gussenhoven, 2004), although he does not discuss this in detail.

All of these studies refer to the analysis of English, and have had repercussions for the English ToBI (Tones and Break Indices, see Beckman and Hirschberg, 1994; Beckman et al., 2005) annotation system, where downstep is labelled with a special diacritic and treated as an optional process (in essence, also a modification on a basic tone). Furthermore, ToBI includes an H+!H*


```
              /\
             /  \
            /    \
           h      l        Phrase level
          /\      /\
         /  \    /  \
        h    l  h    l      Accent level
        T*   T* T*   T*
```
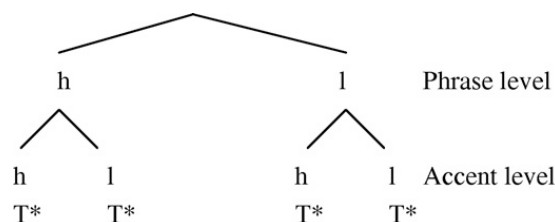
Fig. 2. Metrical analysis of downstep in higher level constituents (from Ladd, 1996:279), where T* stands for any accent.
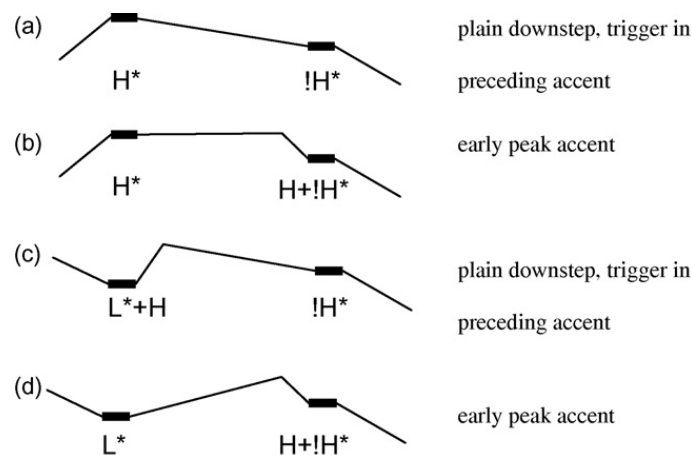
Fig. 3. Stylised downstepping contours.

pitch accent, although its phonological analysis is not what it appears to be on the surface. Instead it is taken to be an H+L* accent underlyingly, but since the L* tone is mid rather than low it is labelled in a phonetically more transparent way as H+!H* to facilitate learning.

A downstep diacritic is also used in the German ToBI system, GToBI (Grice et al., 2005). However, GToBI makes the distinction between H+L* and H+!H*. The pitch accent H+!H* is thus not taken to be underlyingly H+L* but is rather analysed as having a fully-fledged downstepped tone. In the GToBI training materials, the distinction is made between (a) a plain downstepped nuclear accent (!H*) preceded by an H from a prenuclear pitch accent, and (b) an early peak contour where the H1 is in the same accent as the downstepped tone (!H*) (as in Fig. 3(a) and (b) respectively). However, this distinction is not always easy to make, as reflected in the GToBI guidelines, where labellers are advised to place a monotonal accent label if in doubt as to the correct tonal analysis.

According to the GToBI guidelines (as for the English ToBI), H1 does not have to be starred. It can also be the trailing H tone of a bitonal pitch accent, L*+H, as in Fig. 3(c), which is distinguished from (d), where both H tones are in the same accent (H+!H*). Similar contours were also identified by Shattuck-Hufnagel et al. (2004) for English. Here too, they pointed out that it was not easy to distinguish whether an F0 peak on an unstressed syllable belongs to an accent before or after it, especially if there were few syllables between the two accents. In Shattuck-Hufnagel et al.'s experiment, listeners were asked to judge the first or second accent as more prominent as the time alignment of the H peak in (L+)H* !H* sequences was varied. If listeners heard the first accent as more prominent, they attributed the peak to the first accent of the sequence. If they heard the second accent as more prominent, they attributed the peak to the second accent, giving the sequence a different nuclear accent (H* H+!H*). They found that if there were three inter-accentual syllables, a peak on the syllable immediately after the first accent lent prominence to that accent, whereas a peak on the second or third lent prominence to the second accent, and was heard as a leading tone belonging to it. However, if there was only one inter-accentual syllable, the peak could not be unambiguously assigned, although there was a strong tendency to attribute the H peak to the second accent, in which case the second accent is analysed as an *early peak* accent. The term *early peak* refers to the F0 peak in the acoustic signal, which is early in relation to the onset of the accented syllable (i.e. it is *before* the accented syllable).
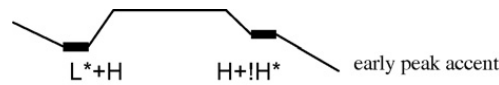
Fig. 4. Stylised L*+H H+!H* L contour.

In Gussenhoven's approach (1984, 1985, also summarised in 2004) tones can reassociate such that what appears on the surface as Fig. 3(c) and (d) are merely variants on the same pitch accent sequence L*+H !H*+L.[1] Note that the final L is part of the nuclear pitch accent, whereas in GToBI it is a boundary tone, a difference which has little bearing on the analysis here.

The shifting of the H peak from the prenuclear to the nuclear accent is achieved through a process of 'partial linking' (see Gussenhoven (1984) for the general theory and for related West Germanic languages, and Uhmann (1991), Féry (1993) and Grabe (1998) specifically for German). Grabe aptly terms this modification to the pitch accent 'displacement' (in the sense that it is the peak that is displaced).

It is not clear whether these models can capture the distinction between the contours in Figs. 3(c) and 4, i.e. whether they distinguish between the presence or absence of a plateau.

Similarly, Gussenhoven's H*+L !H*+L sequence could be taken to be underlying for the examples in (a) and (b) in Fig. 3. The gradual slope in (a) is achieved by deletion (Gussenhoven's term is 'complete linking') of the prenuclear L tone. Since the L is no longer in the surface representation, the pitch can fall gradually through interpolation between the prenuclear H* target and the target for !H* in (a). The plateau in (b) is harder to account for, but possibly by means of spreading of the H* tone to the right. In principle, many of the contours attested in our two studies can be accounted for with Gussenhoven's model, especially if we incorporate spreading as a further modification.

However, we cannot use this model to derive the contours where H+!H* is the only accent in the phrase. This would be treated in Gussenhoven's model as a separate contour for English, which has a specific meaning and is relatively rare. According to Gussenhoven (for English), there is an observable difference in the F0 trace up to the pretonic syllable, depending on whether there is underlyingly H+!H* or a partially linked (displaced) trailing H, and followed by !H*. In the former there is no effect of the scaling of the leading H tone until the pretonic syllable, whereas in the latter (trailing H of L*H), there is a gradual rise from the L* to the H peak just before the next accent (Gussenhoven, 2004:312) as in Fig. 3(d). We return to this issue at the end of section 3.2.

As mentioned above, GToBI has two early peak accents, H+L* and H+!H*. However, there has been some discussion as to whether these should be regarded as two distinct entities. Grabe (1998) argues that early peak accents can involve either partial or total downstep, accounting for the two GToBI accents, H+!H* and H+L* (both of which Grabe analyses as !H*+L, 1998:90). Thus in her view both accents have a downstep component, but one is downstepped to a greater extent than the other.[2] Furthermore, she provides evidence for a gradient distinction between the two types of downstep, indicating that it is not appropriate to analyse them as two different accent types. Note that her representation of GToBI H+!H* (followed by an L− edge tone) is the same as

---

[1] To avoid confusion we use the plus sign to join together tones belonging to one pitch accent; Gussenhoven transcribes the two basic pitch accents L*H and H*L without a plus sign.

[2] Although she explicitly discusses early peak accents at this point, her representation does not distinguish between early peak accents and plain downstep.

GToBI !H* where downstep occurs across accents (i.e. she does not distinguish between (a) and (b) of Fig. 3.

Rathcke and Harrington (to appear), on the other hand, provide arguments for an analysis of early peak accents as H+L*, i.e. with a low tone rather than with a downstepped one. We suspect that there is indeed more than one type of accent with an early peak. But rather than address this issue here, we concentrate on a pitch accent which we have found in two specific contexts: (a) on accessible referents and (b) in broad focus sentences and attempt to find out the appropriate way to analyse it.

Since we are concerned with the distribution of different *downstep* patterns, our first task is to establish whether the accent in question can be analysed as one containing downstep at all (rather than one with a starred L tone instead). We do this in experiment 1, in which we elicit early peak accents on accessible referents in declarative utterances.

In the second experiment we investigate the distribution of downstepped, upstepped and unmodified H* across different focus conditions. Within the downstepped category we then examine the alternation of plain downstep and early peaks, particularly in broad focus contexts. Our aim is to find out whether there is distributional support for the analysis of !H* and H+!H* as related in that they both contain a downstep, thus providing an additional argument for downstep as orthogonal to pitch accent type, rather than as a property of pitch accents themselves, and at the same time providing motivation for the application of downstep regardless of the association of individual H tones.

2. Experiment 1 — Downstep accents on inferentially accessible referents

The first study investigated the scaling of pitch accents in neutral and lively (read) speech.[3] One main aim in this experiment was to ascertain whether the accent under investigation should be analysed as having a downstepped tone or not (otherwise it would have an L* tone). Establishing the identity of the starred tone is crucial for the phonological interpretation of the alternating patterns in the same semantic–pragmatic context.

For that purpose, we collected data using a context which has already been shown to be appropriate for early peak accents, i.e. declarative utterances with nuclear accents on referents which are inferentially accessible (see Baumann, 2006; Baumann and Grice, 2006). For example, if a speaker has just mentioned a *convent*, the word *nun* in a sentence like *He was waiting for the nun* is not 'out of the blue' for the hearer; instead, the presence of a nun can be inferred from the scenario, which, in this case, is the convent. We thus expect an early peak accent on the word *nun*.[4] Similarly, if an *art gallery* is mentioned in which someone wants to buy a *painting*, the presence of the painter can be inferred from that scenario, thus leading to an early peak accent on the word *painter*.

An example F0 trace is given in Fig. 5. In the example, the early peak is clearly on the pretonic syllable, *den*, and is labelled H+. The final low pitch towards the end of the phrase is an edge tone, and is labelled L−% (a sequence of the intermediate phrase boundary tone L− and

---

[3] A preliminary version of this experiment was presented in Grice et al. (2007), although in that version we did not distinguish between downstep types. For the current version we treat H+!H* and plain downstep separately.

[4] Perception experiments on German revealed significant preferences for early peak accents on referents occurring in a given scenario. However, some variation as to the preferred prosodic encoding could be found, both in perception and in production data. Nevertheless, (rising) medial peak accents were as rarely preferred/used for marking referents inferred from a scenario as they were for marking deaccentuation. Plain downstepped tones (i.e. !H* accents) were not investigated (see Baumann, 2006).
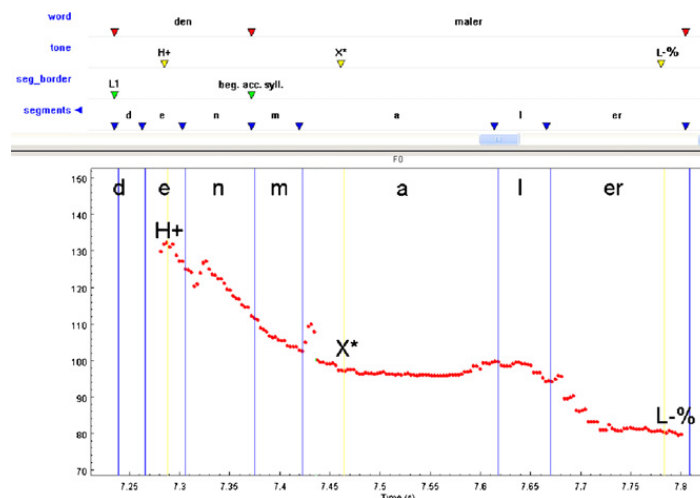
Fig. 5. Example of early peak accent on *den MAler* ('the painter') with F0 trace and labels; from: *Er wartete auf den Maler* ('He was waiting for the painter').

the Intonation Phrase boundary tone L%). The question we aim to address here is whether the apparent F0 target around the middle of the speaker's range on the tonic syllable *MA* should be classified as a downstepped H tone or a (somewhat raised) L tone. That is, is the accent best analysed as H+!H* or H+L*?[5]

As pointed out above, the German ToBI system distinguishes between H+L* and H+!H*, although this distinction has been disputed. Rathcke and Harrington (to appear) argue that the correct representation of the early peak accent is H+L*. Rather than exploring the issue as to whether there are two accent types, we restrict our study to a well-defined context in which one type of early peak accent is appropriate and investigate its scaling across different pitch range conditions.

It is known that pitch range and pitch span can be affected by a number of paralinguistic, attitudinal and emotive factors (see e.g. Ladd et al., 1985; Schröder, 2004). For example, Gussenhoven and Rietveld (2000) show in a perception experiment that as stimuli sound more surprised, H* tones are higher, whereas L* tones are lower. Translated into production, surprised utterances will use an expanded pitch range – also referred to as an increased *pitch span* (Ladd, 1996) – where the topline (affecting the scaling of the H tones) is raised, and the baseline (affecting the scaling of L tones) is lowered. Eliciting comparable speech material with different pitch spans can thus inform us about the identity of the starred tone, X* in early peak accents. As the pitch span increases, X* would be raised if it were a (downstepped) H and lowered if it were an L.

Our first concern was to find a paralinguistic dimension which was likely to manipulate pitch span without affecting the information structural interpretation. For example, if an utterance is spoken with surprise, one might assume that the content of that utterance has been incorrectly inferred from the context. Liveliness, on the other hand, reflects the level of emotional activation or general mood of the speaker, rather than the speaker's attitude towards what is said. It is therefore more compatible with an utterance containing accessible information.

---

[5] The transcription 'X*' in the figure means that although it is clear that there is a starred tone, its identity is left open—it could be H* or L*.

Our second concern was to make sure there was an increase in pitch span without a raising of the overall level. From the attributes which Gussenhoven and Rietveld suggest as promising for this purpose (expressiveness, liveliness and insistence), informal testing led to a preference for liveliness. Thus, we chose to manipulate this paralinguistic dimension, in that the speakers were asked to speak in a *neutral* and a *lively* manner.

2.1. Method

Seven native speakers of German (three female and four male) from undergraduate and graduate courses of Cologne University took part in the experiment. They were aged between 22 and 28, and were asked to read aloud two different sets of mini-dialogues in which the following types of exchange were embedded. The target words and tones which are relevant for the analysis are in bold face (see also Appendix A).

---

(1) A: Ein Bekannter von mir hat seinen Urlaub in einem Kloster verbracht. Jeden Tag betete er stundenlang in der Kapelle. Er wartete auf die **[Early Peak Accent H+X*] Nonne [L−%]**.
B: Auf die **[L*] Nonne**? Warum das denn?
(1A: Someone I know spent his holidays in a convent. Every day he prayed for hours in the chapel. He was waiting for the nun. 1B: For the nun? Why was he doing that?)
. . .
(2) A: Auf wen hat Klaus gewartet?
B: Auf die **[H*] Nonne**. Sie wollte ihm das Kloster zeigen.
(2A: Who was Klaus waiting for? 2B: For the nun. She wanted to show him the convent.)

In 1A the early peak accent is elicited on the word *nun* by means of the scenario, a convent and a chapel. 1B and 2B provided us with reference values for undisputed L* and H* accents, as part of an echo question and a narrow focus declarative respectively.[6]

Exchanges of type 1 and 2 were separated by at least one other, to avoid unintended givenness of the target noun in 2B. Each set of mini-dialogues (including 1A, 1B, 2A and 2B) was read 5 times by four of the speakers and four times by the remaining two.

A tonal analysis was carried out according to the categories in GToBI with the exception of early peak accents, which were labelled H+X*. Alignment points for each of the target tones were placed by hand after inspection of the F0 display in the EMU labelling and database system (see Fig. 5 for a screen shot). In addition, in the early peak contours, the beginning of the tonic syllable and the pretonic syllable were labelled, so that the location of the peak could be determined in relation to these landmarks. (For the *den Maler* context, only cases where a clear segment boundary could be found were labelled, e.g. decrease in F2 in the nasal spectra from /n/ to /m/.) F0 values were transformed into semitones relative to a reference of 50 Hz (Nolan, 2003) to facilitate cross-speaker comparisons.

---

[6] We concede that echo questions may be more involved than neutral questions. Our main reason for chosing an echo question is that we were able to reliably elicit the final rising contour, whereas on other question types there is considerable variation. Furthermore, it is still possible to ask an echo question with differing amounts of liveliness. However, our study did not restrict itself to one type of L tone, since we also looked at the scaling of L−%.

Further categorisation of downstepped tones occurring in context 1A into early peak accents and plain downstep was carried out separately. This categorisation was carried out by two labellers independently (SB and NJ) so that labels could be compared for consistency.

2.2. Results

The recordings from three female and three male speakers were analysed, making a total of 112 mini-dialogues potentially containing one instance of each target tone (H*, L*, H+, X* and L−%). The values of the fourth male speaker were excluded because he frequently employed a non-downstepped (L+)H* pitch accent on the test word in 1A rather than the expected early peak contour, or any other kind of downstep.

Plain downstep was labelled instead of early peak in 18 instances in contexts of type 1A. The 18 utterances containing these cases were removed from the data set before statistical analysis, and are dealt with separately in section 2.2.2. Labels placed by the two transcribers were identical 83% of the time. In the remaining 17% of cases a consensus transcription was reached.

To support the analysis of transcribed early peak accents in contexts of type 1A, latencies were calculated between the beginning of the accented syllable and the labelled H+ peak of H+X*. These were then compared with latencies for the medial H* peak in contexts of type 2B.

Table 1 shows that the H+ tone occurs, on average, 52 ms *before* the beginning of the accented syllable, whereas the H* tone occurs, on average, 160 ms *after* it. A series of *t*-tests show that the difference in timing between H+ and H* is highly significant for all speakers ($p < 0.001$).

Note that these results differ from what Rathcke and Harrington (2007) found in an imitation experiment on German: in their data, H+ was consistently aligned with (the first half of) the *accented* syllable, not with the preceding syllable. This provides an indication that there may indeed be two different types or variants of early peak contours here.

All utterances in the 1A context had a prenuclear accent, despite the fact that the sentences were relatively short and contained a pronoun as subject. This is possibly the result of a rhythmically induced initial accent (Bolinger, 1986; Shattuck-Hufnagel et al., 1994).

2.2.1. Establishing the identity of X*

Fig. 6 shows the mean F0 values (normalised in semitones above 50 Hz) for H and L tones in neutral and lively speech. These are tones whose analysis as H or L is undisputed, both in the case of the starred tones H* (from the context of type 2B) and L* (1B), as well as the leading tone H+ and the boundary tone L−% (both from the early peak context 1A).

Table 1
Mean position of early peak H tone (H+) and medial peak tone (H*) relative to the beginning of the accented syllable in ms

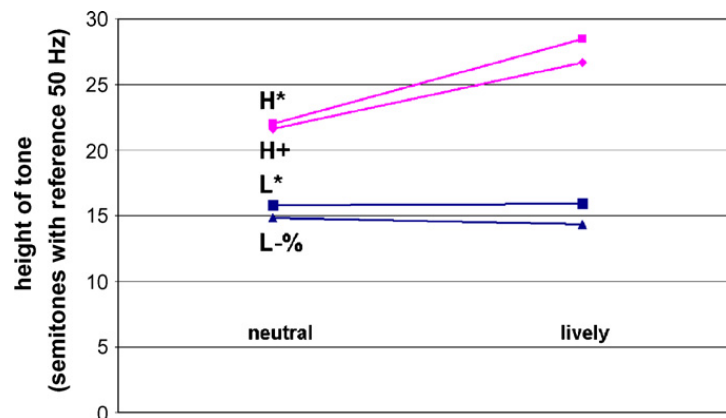| Speaker	(H+) − begin. acc. syll.	(H*) − begin. acc. syll.	*p*
F1	−59.5	170.4	***
F2	−40.9	166.4	***
F3	−56.4	183.3	***
M1	−51.6	187.1	***
M2	−65.2	82.8	***
M3	−38.2	171.4	***
Overall	−52	160.2	***

F = female, M = male.

Fig. 6. Mean F0 values in semitones for all H and L tones in both speaking styles (all speakers pooled).

Table 2
Difference in semitones between neutral and lively conditions for each speaker and each tone, positive values indicate an increase, negative values a decrease

Speaker	H+		X*		L−%		H*		L*	
	Δ (st)	Sig	Δ (st)	Sig	Δ (st)	Sig	Δ (st)	Sig	Δ (st)	Sig
F1	2.64	**	1.11	*	−0.19	*	3.58	***	0.43	0.342
F2	5.18	***	1.05	0.056	−0.62	**	9.98	***	−0.91	**
F3	4	***	3.66	***	0.98	*	8.69	***	0.83	0.055
M1	4.74	*	2.62	0.068	0.24	0.768	6.22	***	−0.71	0.071
M2	8.55	***	6.91	***	−1.1	0.056	6.26	***	0.85	*
M3	7.73	***	2	***	1.22	**	4.85	***	2.64	***

Two-tailed *t*-tests revealed that both H tones (H+ and H*) are significantly higher in lively than in neutral speech. By contrast, L−% tones were either significantly lower (speakers F1, F2), higher (F3, M3) or not significantly different (M1, M2). Similarly, L* tones were either significantly lower (F2), higher (M2, M3) or not significantly different (F1, F3, M1) (see Table 2).

The difference between neutral and lively values for H tones ranged from 2.6 to 10 semitones, whereas in the few cases where L tones were higher in lively speech, the differences were very small, ranging from 0.2 to 1.2 semitones (except for speaker M3, who had a difference as high as 2.6 semitones for L*).

Having presented the behaviour of H and L tones in neutral and lively speech, we now turn to the controversial tone, X*. This tone was found to be significantly higher in lively than in neutral speech for four speakers (F1, F3, M2, M3), the other two (F2, M1) showing clear tendencies in the same direction. The differences ranged from 1.1 to 6.9 semitones, with an overall mean of 2.9 semitones. Unlike the L tones, which were rather variable in their scaling, the values for X* were always higher in lively speech.

Fig. 7 shows how X* is scaled in neutral and lively speech in comparison to H+ and L−% of the same phrase. These two tones define the upper and lower bounds of the local pitch range (van den Berg et al., 1992).[7]

---

[7] Since Patterson and Ladd (1999) found that the final low value in a phrase is relevant for listener judgements as to certain types of affect, such as anxiety or confidence, we take this value into account.
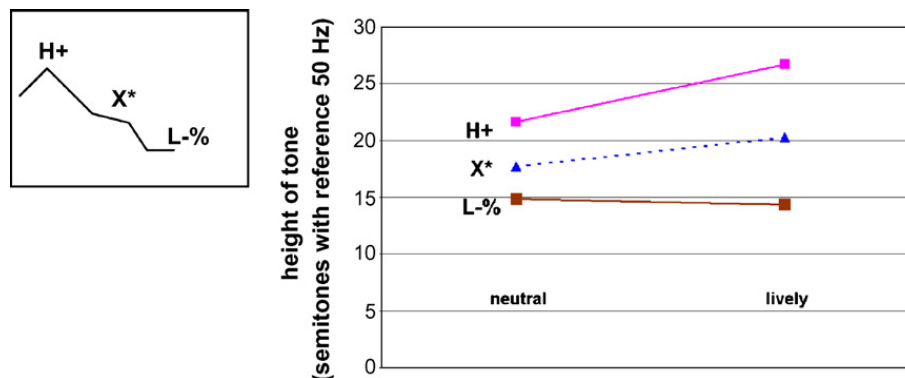
Fig. 7. Mean F0 values in semitones (reference 50 Hz) for H+, X* and L−% in neutral and lively speaking styles (all speakers pooled).

However, there are differences across speakers, as shown in Fig. 8. Speaker M3 not only increased the span but also raised his pitch level, as did F3, albeit by a small margin. The remaining four speakers increased their pitch span without raising the pitch level. These speakers are ideal subjects for investigating tonal identity. In their speech, X* is higher in lively speech, although this increase is only significant in two of the four cases. These results already point in the direction of X* being analysed as a high tone.

Stronger evidence for treating X* as a high tone is gained from looking at the difference between the value for X* and the following low boundary tone, L−%. Here two-tailed t-tests revealed that the pitch excursion in semitones from X* to L−% is significantly greater in lively as compared to neutral speech for all speakers except M3 (see Table 3). A comparison with pitch excursion differences between L* and L−% is not possible since this contour did not occur in our data. In fact, L* L−% is a rare combination in German.

We now turn to the location of X* within the local pitch range. For this we treated H+ and L−% as upper and lower bounds (100% and 0% respectively). Although for four speakers (F3, M1, M2, M3) the values for X* were towards the middle of the range (at 40%), as would be expected if the tones are downstepped. For the remaining two speakers (F1, F2), the scaling of X* is relatively low (28%).
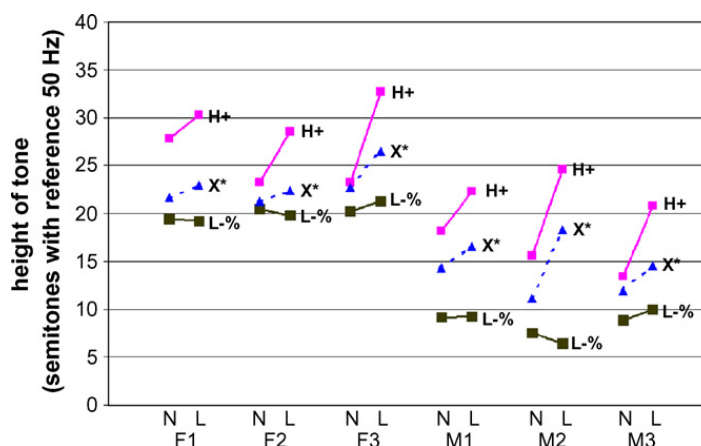

Fig. 8. Mean F0 values in semitones (reference 50 Hz) for H+, X* and L−% for each speaker separately in neutral (N) and lively (L) speaking styles.

Table 3
Excursion size, X* to L−% in semitones for each speaker (neutral and lively conditions)

X*-L−%

Speaker	Excursion size (st) neutral	Excursion size (st) lively	p
F1	2.25	3.55	**
F2	0.87	2.54	**
F3	2.36	5.04	***
M1	5.46	7.84	**
M2	3.88	11.89	***
M3	3.17	3.98	0.222

This variation in scaling is comparable to what Grabe (1998) found in German plain downstepped accents in nuclear position (recall that she transcribes them as !H*L, and does not distinguish them from early peaks in her transcription). She described German as having a 'gradiently varying continuum of excursion for final falls' (Grabe, 1998:198), comparing German contours to their English equivalent, in which the fall from a final downstepped peak to the final low was not as variable and did not include such small excursions. Fig. 9(a) is taken from Grabe (1998) to illustrate this continuum in terms of normalised F0 values for the downstepped tone and the final low, and Fig. 9(b) shows that the distribution for our male speakers is strikingly similar. Grabe normalized the F0 values as follows: the final low was set to the mean for all speakers for L−%, and the values for X* were calculated in relation to that value. We used this method of normalization, but treated male and female speakers separately, due to their marked pitch range differences.

Fig. 9(c) and (d) show that although there are low values for X* in the neutral speech of our male speakers, the values are generally higher and more diverse in lively speech. If X* had been a low tone, we would not have expected an increase in values accompanying liveliness.

We can observe the same type of distribution in the female speaker's data, Fig. 9(e) and (f). The scaling of X* in relation to the L tone at the end of the phrase varies in a way which has already been found to vary for plain downstep. That is, although there were some low values for two female speakers, the variation in our results in terms of height of X* in relation to the following boundary L are in line with those from Grabe's independent study on downstepped H tones in the same language. Furthermore, lively values involve a majority of greater excursions than neutral speech, also indicating that we are dealing here with an H tone. If X* had been an L tone we would have expected the relation between X* and L−% to be less varied across speaking styles. That is, we would expect the scaling of the starred tone and the boundary L−% to vary in a compatible way if they were both L tones.

2.2.2. Alternation between early peak accent and plain downstep

Having established the analysis of the early peak accent as H+!H*, we now examine the distribution of this accent in relation to plain downstep in the 1A context. We classified a contour as an 'early peak accent' if the nuclear syllable was immediately preceded by a higher pitch which could not be accounted for by interpolation from a previous high tone. Examples of an early peak contour and a plain downstep are given in Fig. 10.

As already mentioned, we labelled 18 cases of plain downstep. Since there were 112 cases of downstep in total, this constitutes 17% of all downsteps, the remaining 83% being early peak accents. Fig. 11 shows the distribution of both downstep types for neutral and lively speech. No significant differences were found across the two speaking styles ($\chi^2(1) = 0.977$; $p = 0.322$).

Furthermore, we observe considerable differences as to the preferred type of downstepping contour across single speakers (Table 4). Here, a chi-square test revealed a significant interaction ($\chi^2(5) = 15.897$; $p < 0.01$). Speaker F3, for example, does not use plain downstepped contours at all, whereas the other speakers do use them, albeit less frequently than early peak accents.

Thus, we can conclude that the choice to use early peak or plain downstep is to some extent speaker-specific, and that in this experiment early peak accents are generally preferred over plain downstep regardless of speaking style.

To summarise our results, we have provided evidence in favour of an analysis of the contour as H+!H* as opposed to H+L*. Our evidence is gained not from its location within the pitch range of a given speaker or phrase, but rather from the difference between its realisation across the two
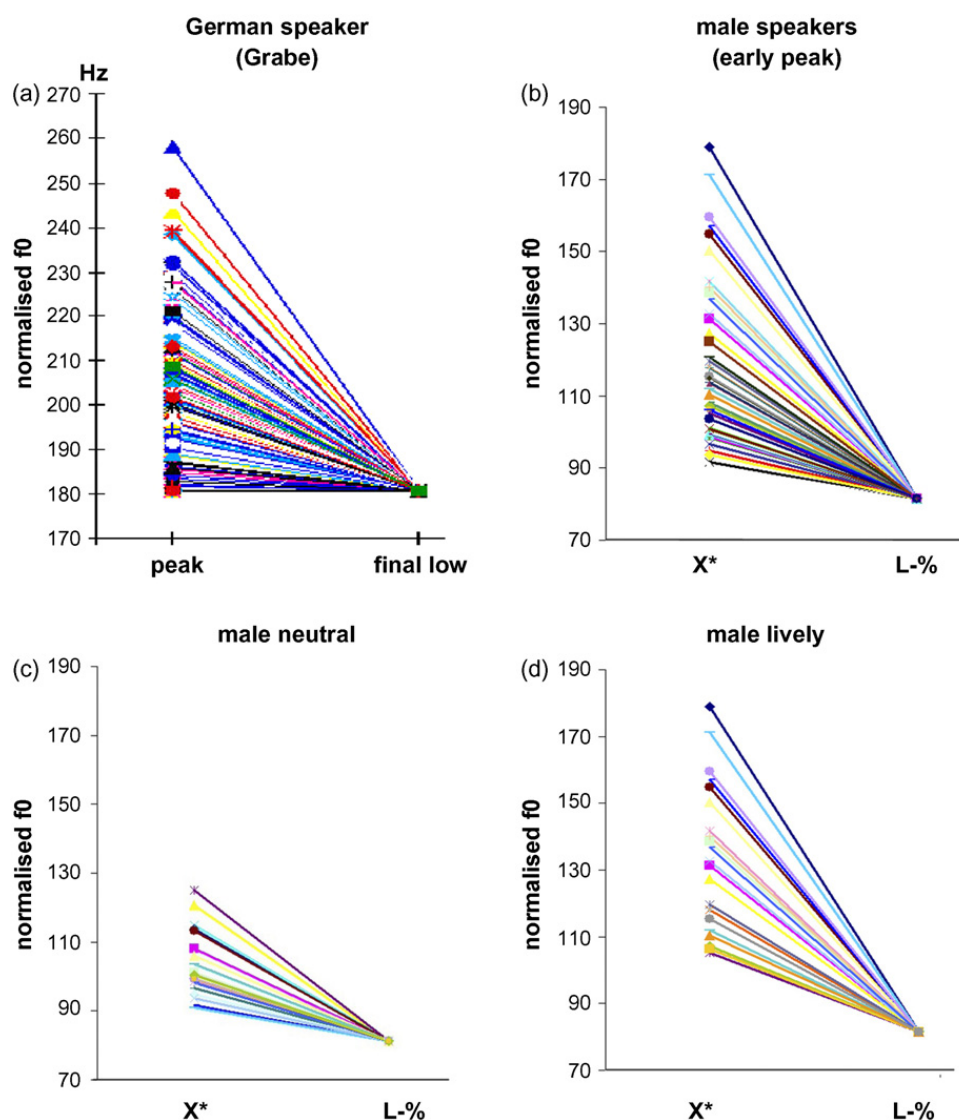

Fig. 9. Variation in the scaling of tones in relation to the final low tone: (a) a plain downstepped H* peak reproduced from Grabe (1998), (b) X* for male speakers, and (c, d, e, f) X* for male and female speakers in neutral and lively speech separately.
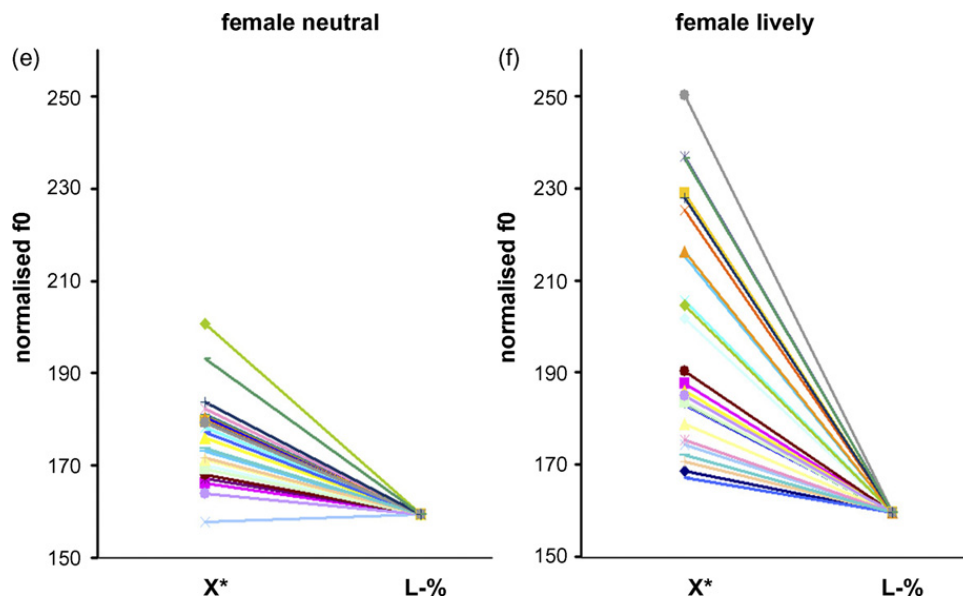
Fig. 9. (*Continued*)

Table 4
Percent of early peak accents for each speaker and speaking style (downstepping contours only)

Speaker	Neutral	Lively
F1	100	88
F2	90	100
F3	100	100
M1	60	70
M2	80	90
M3	60	70

speaking styles which involve differences in pitch span. More concretely, when the pitch span is increased, the target for the starred tone is generally raised, and the pitch excursion between the starred tone and the phrase final low pitch is increased. Moreover, the pitch excursion between the starred tone and the final low involves a similar variation to that found for plain downstepped tones in an independent study reported on in Grabe (1998) on the same language.

3. Experiment 2 — Focus domains and focus types

In the second study, we investigate in which context downsteps commonly occur, and how the two kinds of downstepping contour – early peak accents (which, on the basis of the results in experiment 1, we take to be H+!H*) and accents with plain downstep (!H*) – are distributed across three focus structures, namely broad, narrow and contrastive (corrective) focus.[8] In addition to !H* and H+!H*, we also examined other nuclear pitch accent labels, such as

---

[8] A subset of the results from this corpus has been described in Baumann et al. (2007), which placed particular emphasis on the articulatory domain. In the present paper, we investigate accent types and modifications to accents, neither of which was dealt with in any detail in that study.
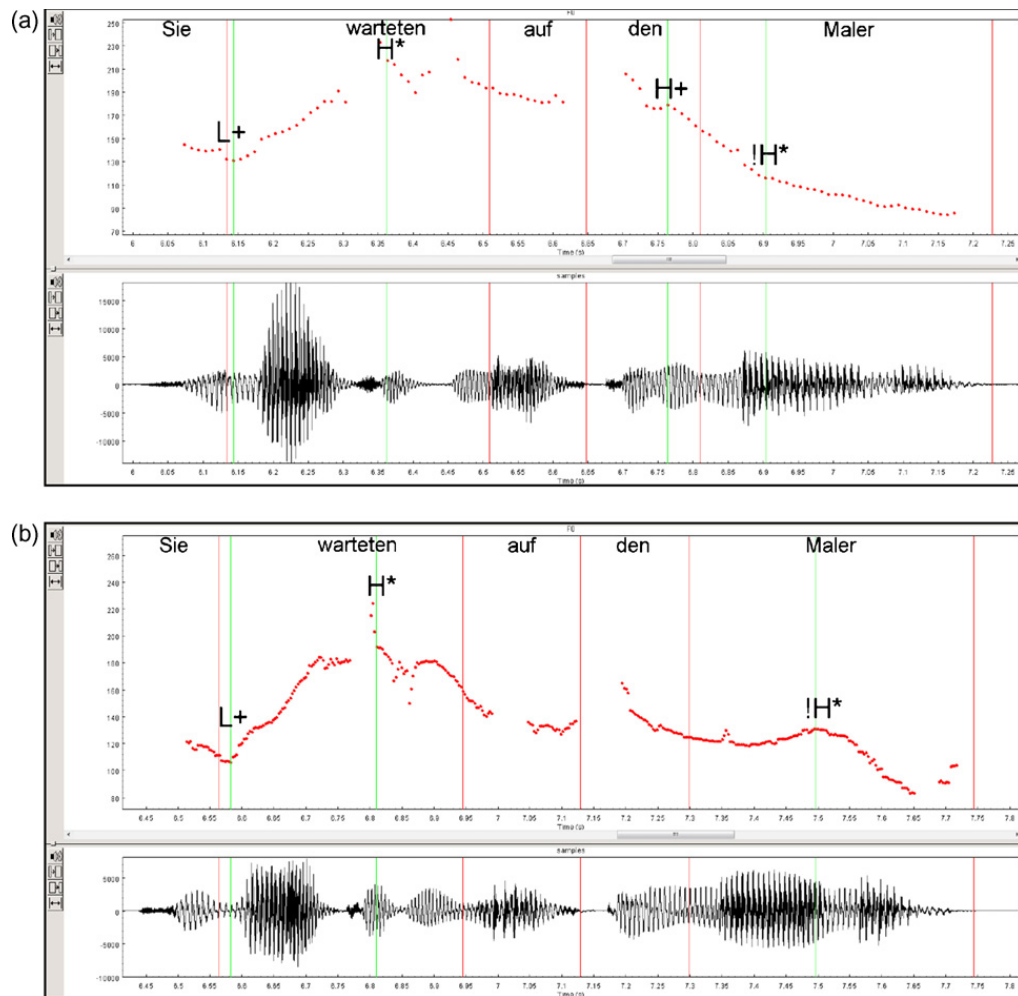
Fig. 10. Typical examples of an early peak accent (a) and a plain downstep (b) on *den MAler* ('the painter') with F0 trace and labels.
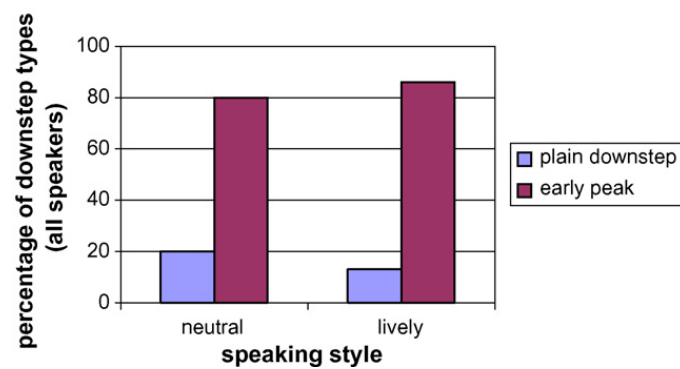

Fig. 11. Downstepping contours (all speakers pooled).

^H* (upstep, i.e. another modification nuclear H tones can undergo) and H* (unmodified medial peak accent).

3.1. Method

Six native speakers (three female, three male, originating from near Düsseldorf and aged between 20 and 28, none of whom had participated in experiment 1) read aloud target sentences displaying differences between broad and narrow focus on the one hand, and non-contrastive and contrastive (corrective) focus on the other. The sentences were the answers in question-answer pairs, eliciting the different focus structures, as in (3) below:

Questions:

1. Was gibt's Neues?   *What's new?*

2. Was will Marlene schälen?   *What does Marlene want to peel?*

3. Will Marlene eine Kartoffel schälen?   *Does Marlene want to peel a potato?*

Answers:

    Marlene will eine Banane schälen.

1.   [———————————————] focus   broad

2.   [————] focus   narrow

3.   [———] focus   contrastive

     lit.: *Marlene wants a banana to-peel*

The questions were presented to the subjects auditorily (pre-recorded sound files read by the instructor) and at the same time in written form, and speakers were asked to read out the answer in a contextually appropriate manner. After a training block of five question–answer pairs, recordings were made of each of the eight sentences in three focus conditions (see material in Appendix A) in three pseudo-randomized blocks.

As in experiment 1, segmental and tonal labels were assigned using the speech database system EMU. Labels included the target word, phoneme-sized segments, and the L and H tones associated with pitch accents. An example screen shot is given in Fig. 12.

One major concern in this experiment is to ascertain the height of the nuclear accent peak relative to a prenuclear peak as the focus structure is varied. The three categories *upstep* (^H*), *downstep* (!H*) and *unmodified* (H*) were distinguished perceptually (later acoustic analysis of the labelled tones revealed de facto threshold values of approximately 1.5 semitones between the prenuclear peak, labelled H0, and the nuclear peak, labelled H2 (slightly varying between speakers). That is, if the nuclear peak was 1.5 st (or more) higher than the prenuclear one, it was classified as ^H*, and if it was 1.5 st (or more) lower, it was classified as !H*). Values differing by
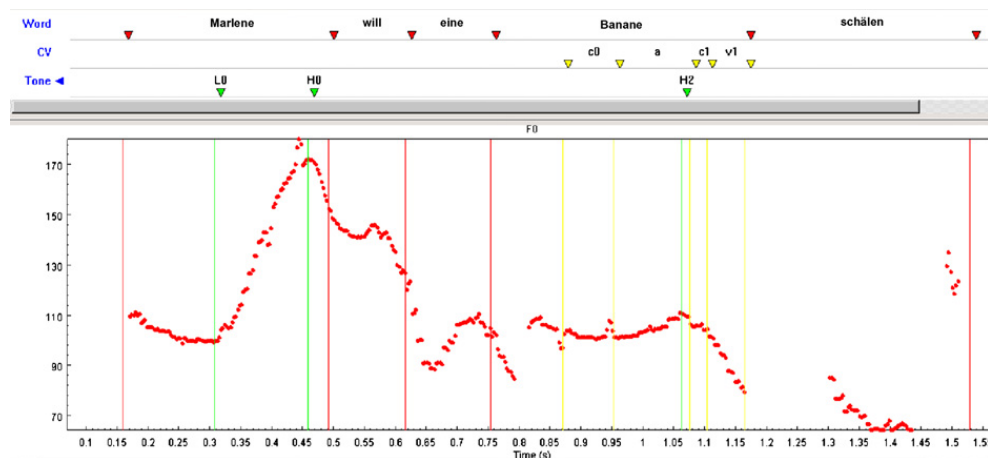
Fig. 12. Target sentence *Marlene will eine Banane schälen* (broad focus, speaker B) with labels and F0 trace in EMU.

less than 1.5 st in either direction led to a classification of the tone as H*. A consistency test was carried out on half of the data (speakers A, B, and C) in which two independent transcribers (SB and NJ) assigned one of these three categories to each token, and, in the case of downstep, whether the speaker produced an early peak or a plain downstep, following the same criteria as in experiment 1.

3.2. Results

All data from all six speakers were analysed, comprising three repetitions of eight sentences in three focus conditions. Thus, the corpus included 72 tokens per speaker, i.e. 432 tokens in total.

3.2.1. Modifications to H* and focus

A total of 37.5% of all utterances included downstep on the nuclear accent, and about the same (37%) included unmodified H*, i.e. neither upstepped nor downstepped. Somewhat less frequent was upstep, which occurred in 25.5% of all utterances.[9]

Looking at the focus structures separately, we found more downsteps in broad focus and more upsteps in contrastive focus, as shown in Fig. 13 (there were highly significant differences in the choice of accent type: $\chi^2(4) = 85.651; p < 0.001$). This result is in line with our assumption that a higher accent (relative to a preceding one) sounds more prominent than a lower accent and is thus more appropriate for marking a semantic contrast (Brown et al., 1980; Ladd and Morton, 1997). However, a comprehensive discussion of upstep, also analysed as H raising, is beyond the scope of this paper (see Truckenbrodt, 2002; Féry and Kügler, 2008), since our main focus of interest here is the distribution of downstepped tones in relation to H tones which are not downstepped.

Speakers varied as to their preferred choice of contours. One speaker (O) had a strong preference for unmodified accents in all conditions while another (B) favoured downstep in all three focus structures. Yet another speaker (C) restricted his use of downstep to broad focus. Nevertheless, the interaction between focus structure and modification type remains significant in the data of five out of six speakers (see Table 5).

---

[9] Féry and Kügler (2008) also found considerable variation as to the choice of accent pattern in broad focus structures in read German. Our results are thus compatible with theirs in this respect.
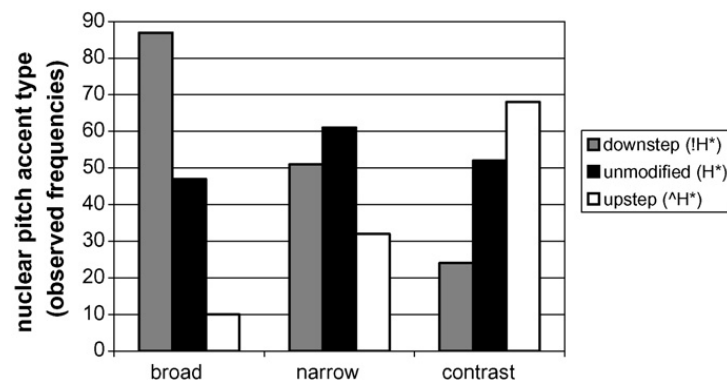
Fig. 13. Observed frequencies of modifications of H* in relation to focus structure across all speakers (total number of utterances: $N = 432$).

The consistency test showed a high agreement for the three categories across the two labellers (74% identical transcription).

3.2.2. Alternation between early peak accent and plain downstep

Of all the transcribed downsteps, 51% were early peaks and 49% plain downsteps. The consistency test for the categorisation of accents into plain downstep and early peak showed an

Table 5
Observed frequencies of modifications of H* in relation to focus structure for each speaker separately

| Speaker	Focus	Modification of nuclear accent			Chi-square (d.f. = 4)	*p*
		Downstep	Unmodified	Upstep		
A	Broad	8	12	4	13.739	<0.01**
	Narrow	14	7	3		
	Contrast	2	15	7		
B	Broad	20	4	0	9.348	=0.053n.s.
	Narrow	12	10	2		
	Contrast	15	5	4		
C	Broad	16	7	1	48.183	<0.001***
	Narrow	0	13	11		
	Contrast	0	7	17		
D	Broad	15	6	3	20.321	<0.001***
	Narrow	9	11	4		
	Contrast	3	7	14		
J	Broad	24	0	0	46.633	<0.001***
	Narrow	7	8	9		
	Contrast	3	4	17		
O	Broad	4	18	2	14.415	<0.01**
	Narrow	9	12	3		
	Contrast	1	14	9		
All speakers	Broad	87	47	10	85.651	<0.001***
	Narrow	51	61	32		
	Contrast	24	52	68		
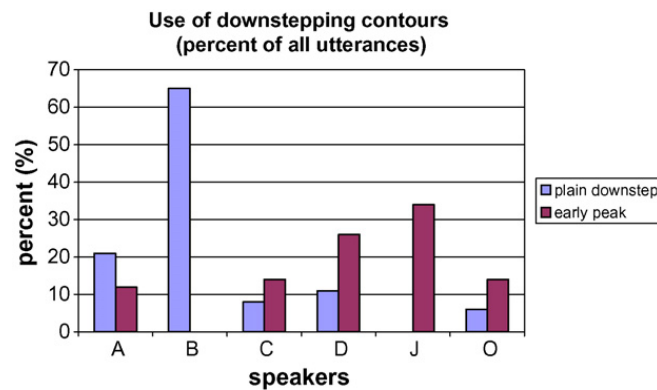
Fig. 14. Uses of downstep variants across different speakers (in percent of all utterances per speaker).
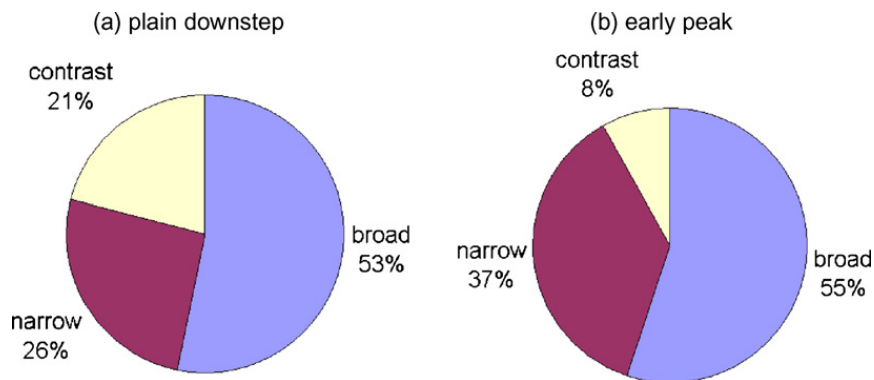

Fig. 15. Distribution of plain downstep (a) and early peak accents (b) across three focus conditions, numbers expressed as a percentage of total number of cases for each downstep type.

agreement across labellers of 95%. There were considerable inter-speaker differences as to the preferred type of downstep: speaker J, for example, exclusively uses early peaks (47% of all utterances), whereas speaker B does not use a single early peak contour, although 65% of his utterances include a downstep. Fig. 14 visualises the distribution of downstepping contours across all speakers.

Looking at the results for all speakers pooled, we can observe in Fig. 15 that broad focus has an equal distribution of both plain downstep and early peaks. The differences in narrow and constrastive focus are the result of speaker-specific preferences for one or other type of contour across the board (in particular influenced by the data of speakers B and J; see Fig. 14).
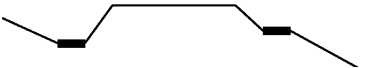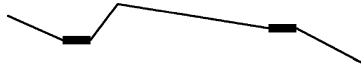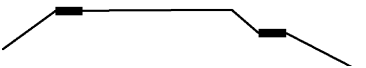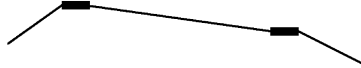
If we concentrate on broad focus, the context which has most of both types of downstep, and analyse the prenuclear accent type in each phrase, we find that early peak accents can be preceded by no pitch accent at all, by an L* (as a monotonal pitch accent or as part of L*+H), and by a peak accent (H* or L+H*). Table 6 summarizes the results.[10]

Returning to the contours schematized in Figs. 3 and 4 in section 1, we can see that in this context, broad focus, all five contours are attested, as well as an early peak contour without a prenuclear accent. Table 7 gives an overview, while Fig. 16 shows actual examples of the last four

---

[10] Plain downstep cannot occur if there is no prenuclear H tone to act as trigger, hence the two empty cells.

Table 6
Distribution of prenuclear accent type and type of downstepping contour in broad focus contexts (absolute numbers, all speakers pooled, with stylised contours)

Prenuclear accent type	Type of downstep on nuclear accent	
	Early peak accents	Plain downsteps
No accent	7	–
L*	2	–
L*+H	32	17
(L+)H*	12	25

contour types (example screen shots of the first two contour types have already been given in Fig. 10 in section 2.2.2).

The contour L*+H H+!H* (Figs. 4 and 16(c)) is possibly derivable through spreading of the trailing H tone of the prenuclear accent. H+!H* without a prenuclear accent (Fig. 16(d)) is considered to be a separate accent (not derivable), with a distinct meaning. Furthermore, the rise up to the leading H of H+!H* is gradual (see Fig. 16(d)) unlike the contour described by Gussenhoven for English H+!H* and more like the contour of a derived accent (from L*+H !H*+L, discussed in section 1).

In sum, early peaks appear to be distributed in a similar way to plain downstep and can be preceded by a number of different prenuclear accents, or by no accent at all. The high degree of variation within the same context and the rather subtle nature of some of the distinctions point to the relatedness of the contour in question.

Furthermore, the rise up to the leading H of H+!H* is gradual (see Fig. 16(d)) unlike the contour described by Gussenhoven for English H+!H* (where the rise doesn't start until the pretonic syllable) and more like the contour of a derived accent (where the leading H tone for the nuclear accent is a partially linked trailing tone from L*+H, as discussed in section 1).

Table 7
Attested variants of downstepping contours in German and their frequency of occurrence in the broad focus contexts of experiment 2, and their analysis following Gussenhoven's approach

	Contour	Frequency of occurrence	Analysis along the lines of Gussenhoven
Fig. 3(a)	(L+)H* !H*	25	Derivable by linking H*+L !H*+L
Fig. 3(b)	(L+)H* H+!H*	12	Not derivable (but analysed as distinct H+!H* accent)
Fig. 3(c)	L*+H !H*	17	L*+H !H*+L
Fig. 3(d)	L* H+!H*	2	Derivable by partial linking of L*+H !H*+L
Fig. 4	L*+H H+!H*	32	Not distinguishable from 3 d
	Ø H+!H*	7	Not derivable (but analysed as distinct H+!H* accent)

4. Discussion

We have shown that not one accent or accent combination but rather a number of different ones are found to occur within and across the two contexts we investigated (accents on accessible referents and on the focus exponent of broad focus structures). Moreover, what these intonation contours have in common is that they all have a downstepped tone on the nuclear syllable. This commonality provides additional confirmation for downstep as having a functional unity, as discussed by Ladd (1983), who uses this as an argument against downstep being the result of local phonological rules operating on specific sequences of tones.

Ladd (1993, 1996) later provides a metrical representation of downstep relations across accents. The main problem with this representation is that it does not allow for downstep within *one* accent. Our results point to the need for a representation in which both downstep across
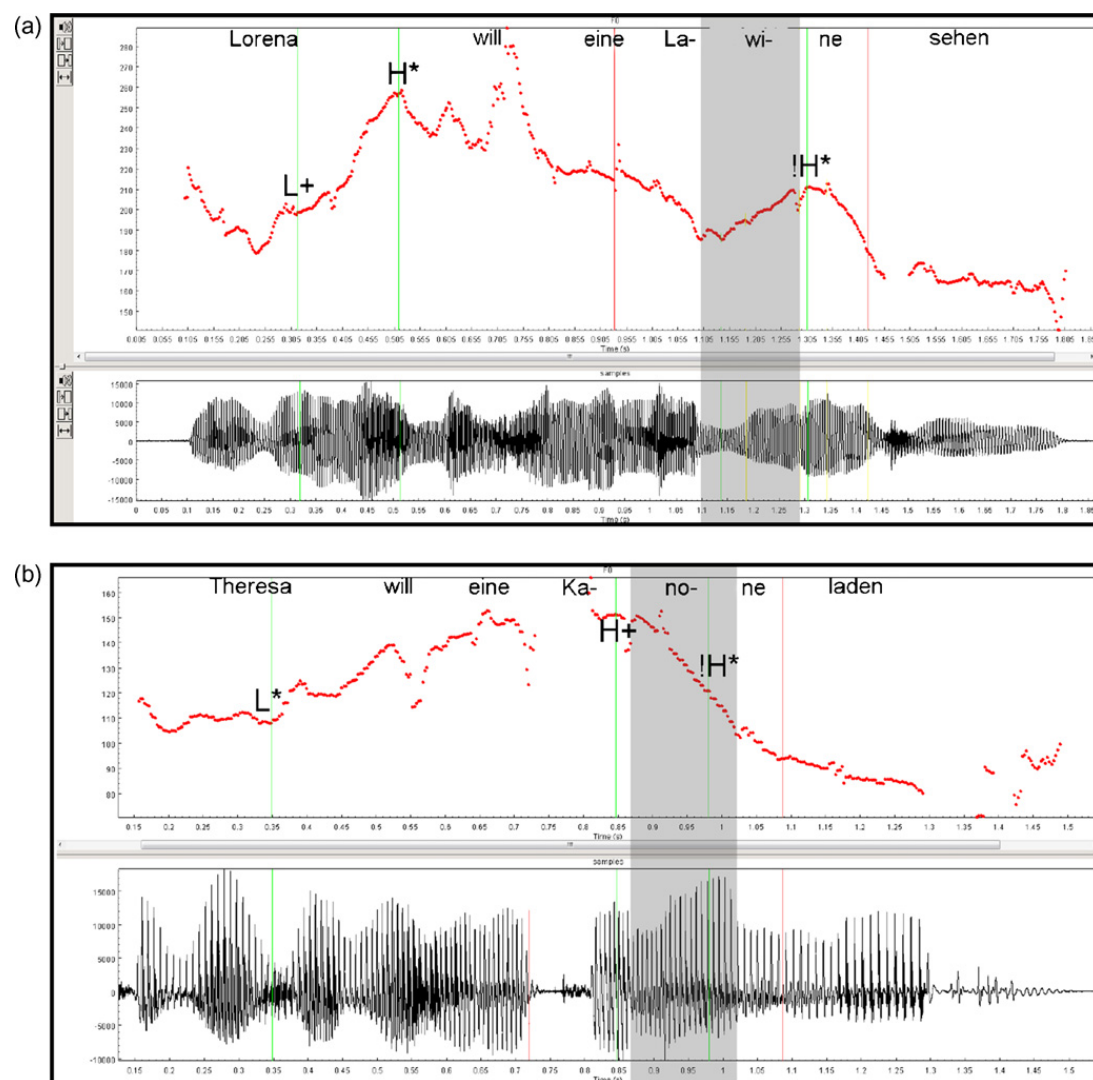

Fig. 16. Example screen shots of (a) L*+H !H*, (b) L* H+!H*, (c) L*+H H+!H*, and (d) Ø H+!H*, nuclear syllables shaded.
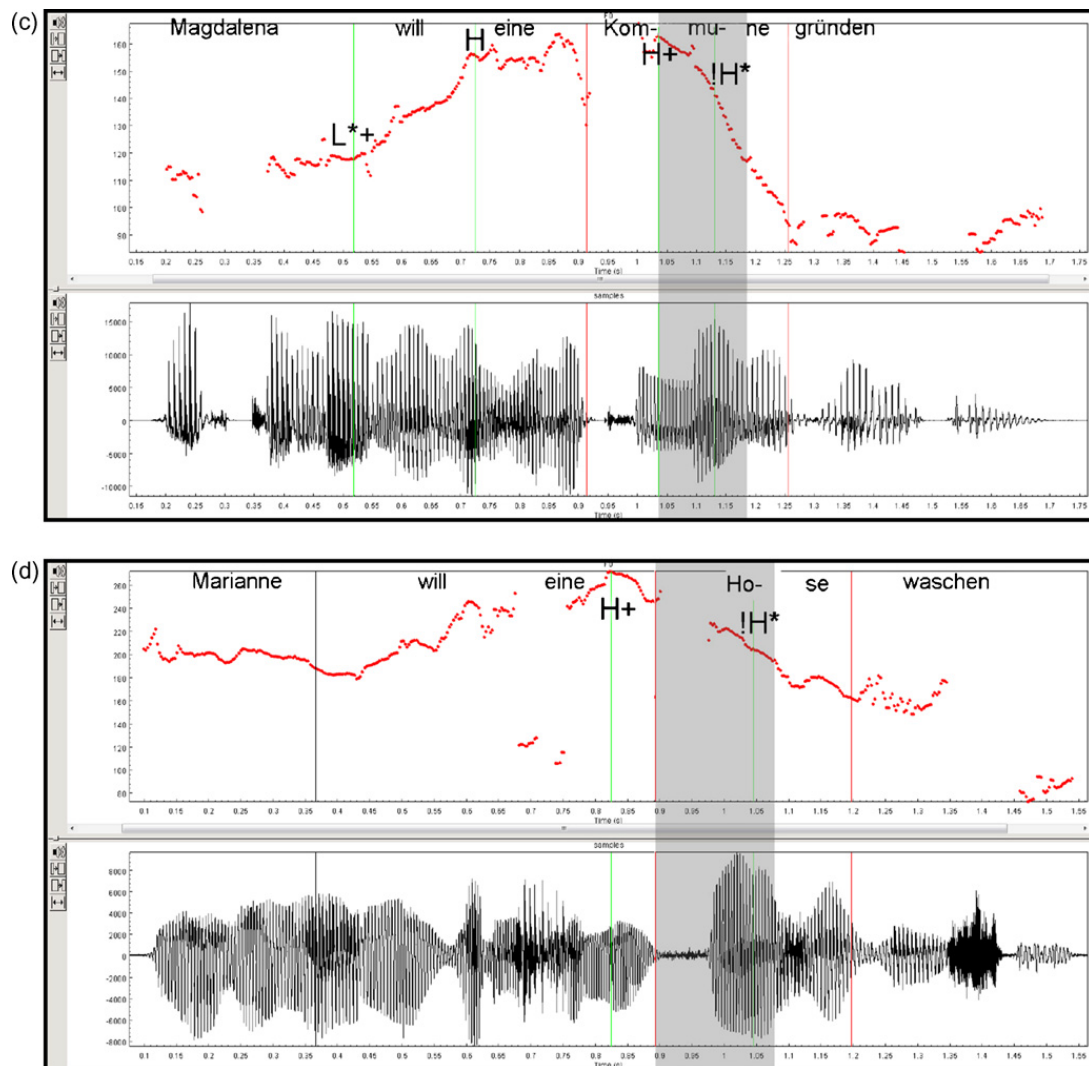
Fig. 16. (*Continued*)

accents and downstep within an accent should be captured using a common phonological representation.

The relatedness of most of the intonation contours can be accounted for in Gussenhoven's model, downstep being an operation which is applied to underlying contours (Gussenhoven, 1984, 1985, 2004). For example, it is possible to relate (i) and (ii)

(i)   L*+H !H* and
(ii)   L* H+!H*

in that they both have a common underlying contour, which is the one in (i). The contour in (ii) is derived through partial linking (displacement to the right) of the middle H tone (see Table 7). However, the differences in association of the H tone before the nuclear !H* can only be partially accounted for. Gussenhoven's model is not able to derive a surface form with a prenuclear H* or

L+H* and a nuclear H+!H*, nor a single accented phrase with H+!H*, except with a distinct H+!H* accent.

Our proposal is to allow for certain tones to be unspecified as to their association properties. So that, instead of deriving (ii) from (i) as above, both would be derived from a common form such as (iii).

(iii)   L* H !H*

That is, the association of the H in the middle is left open (it is, in a sense, floating) in the underlying representation. A unified downstep mechanism could then apply on this underlying representation. It would be blind to whether it is operating across accents or within an accent.

Using this underspecified underlying representation, all of the attested contours can be related to each other. A representation would look like (4), once downstep had been applied but before association has been finalised for the first H.

(4)   ... H ... !H* L L%

If H+!H* is indeed derived, then some of the arguments which could be levelled against it – especially those relating to the implications for the number of combinations of tones into new pitch accent types, leading to a large inventory of pitch accents – are less compelling. Our analysis differs from that of Kügler (2006), who also has derived pitch accents, but through affixation, i.e. the addition of tones to the basic representation, although the affixation is constrained, and is therefore also free of the implications relating to an expansion of the inventory.

Our analysis would also support a different look at the phrase accent, which in Grice et al. (2000) was analysed as a boundary tone with a secondary association to a tone bearing unit (if available, and according to dialect or language-specific characteristics). The phrase accent may also be a tone which is unspecified for association. Downstepped phrase accents are part of the American English ToBI inventory (Beckman et al., 2005; Beckman and Hirschberg, 1994) as well as the German one (Grice et al., 2005) and are transcribed as !H−. Since downstep operates at the phrasal level before association has taken place, this type of downstep might also be accounted for in the same way.

In sum, we propose an underlying specification of intonational tunes which does not specify directly for all tones whether they are part of a pitch accent or not. Operations such as downstep (and possibly upstep) are applied to this underlying representation, so that such operations are not affected by the association properties of the H tones they apply to. The surface association of the tones may contain an optional element, and will undoubtedly be affected by various factors, such as speed of delivery and the number of metrically strong syllables available as potential docking sites for pitch accent tones. However, we leave the investigation of the influence of these factors for future experiments.

Acknowledgements

Appendix A. Test material

Experiment 1

Mini-dialogue 1:

A: Am Sonntag haben Anna und Klaus eine Ausstellung besucht. Die beiden hatten vor, sich ein Bild zu kaufen. Sie warteten auf **den Maler**.

B: Auf **den Maler**? Warum das denn?

(*A: On Sunday Anna and Klaus went to an exhibition. They planned to buy a picture. They were waiting for the painter. B: For the painter? Why were they doing that?*)

A: Auf wen haben die Gäste gewartet?

B: Auf **den Maler**. Sie halten ihn für einen großen Künstler.

(*A: Who were the guests waiting for? B: For the painter. They consider him to be a great artist.*)

Mini-dialogue 2:

A: Ein Bekannter von mir hat seinen Urlaub in einem Kloster verbracht. Jeden Tag betete er stundenlang in der Kapelle. Er wartete auf **die Nonne**.

B: Auf die **Nonne**? Warum das denn?

(*A: Someone I know spent his holidays in a convent. Every day he prayed for hours in the chapel. He was waiting for the nun. B: For the nun? Why was he doing that?*)

A: Auf wen hat Klaus gewartet?

B: Auf **die Nonne**. Sie wollte ihm das Kloster zeigen.

(*A: Who was Klaus waiting for? B: For the nun. She wanted to show him the convent*)

Experiment 2

broad focus

Was gibt′s Neues? Milena will eine Mine suchen.

(*What's new? Milena wants to look for a mine.*)

Was gibt′s Neues? Elvira will eine Muse finden.

(*What's new? Elvira wants to find a muse.*)

Was gibt′s Neues? Manuela will eine Nase malen.

(*What's new? Manuela wants to paint a nose.*)

Was gibt′s Neues? Marianne will eine Hose waschen.

(*What's new? Marianne wants to wash some trousers.*)

Was gibt′s Neues? Lorena will eine Lawine sehen.

(*What's new? Lorena wants to see an avelanche.*)

Was gibt′s Neues? Magdalena will eine Kommune gründen.

(*What's new? Magdalena wants to build up a commune.*)

Was gibt′s Neues? Marlene will eine Banane schälen.

(*What's new? Marlene wants to peel a banana.*)

Was gibt′s Neues? Theresa will eine Kanone laden.

(*What's new? Theresa wants to load a cannon.*)

narrow focus

Was will Milena suchen? Milena will eine Mine suchen.

(*What does Milena want to look for? Milena wants to look for a mine.*)

Was will Elvira finden? Elvira will eine Muse finden.

(*What does Elvira want to find? Elvira wants to find a muse.*)

Was will Manuela malen? Manuela will eine Nase malen.
(*What does Manuela want to paint? Manuela wants to paint a nose.*)
Was will Marianne waschen? Marianne will eine Hose waschen.
(*What does Marianne want to wash? Marianne wants to wash some trousers.*)
Was will Lorena sehen? Lorena will eine Lawine sehen.
(*What does Lorena want to see? Lorena wants to see an avelanche.*)
Was will Magdalena gründen? Magdalena will eine Kommune gründen.
(*What does Magdalena want to build up? Magdalena wants to build up a commune.*)
Was will Marlene schälen? Marlene will eine Banane schälen.
(*What does Marlene want to peel? Marlene wants to peel a banana.*)
Was will Theresa laden? Theresa will eine Kanone laden.
(*What does Theresa want to load? Theresa wants to load a cannon.*)

contrastive focus
Will Milena einen Schatz suchen? Milena will eine Mine suchen.
(*Does Milena want to look for a treasure? Milena wants to look for a mine.*)
Will Elvira einen Freund finden? Elvira will eine Muse finden.
(*Does Elvira want to find a friend? Elvira wants to find a muse.*)
Will Manuela eine Blume malen? Manuela will eine Nase malen.
(*Does Manuela want to paint a flower? Manuela wants to paint a nose.*)
Will Marianne einen Pullover waschen? Marianne will eine Hose waschen.
(*Does Marianne want to wash a sweater? Marianne wants to wash some trousers.*)
Will Lorena einen Baum sehen? Lorena will eine Lawine sehen.
(*Does Lorena want to see a tree? Lorena wants to see an avelanche.*)
Will Magdalena eine Partei gründen? Magdalena will eine Kommune gründen.
(*Does Magdalena want to build up a party? Magdalena wants to build up a commune.*)
Will Marlene eine Kartoffel schälen? Marlene will eine Banane schälen.
(*Does Marlene want to peel a potato? Marlene wants to peel a banana.*)
Will Theresa ein Gewehr laden? Theresa will eine Kanone laden.
(*Does Theresa want to load a gun? Theresa wants to load a cannon.*)

References

Baumann, S., 2006. The Intonation of Givenness—Evidence from German. Niemeyer, Tübingen.
Baumann, S., Becker, J., Grice, M., Mücke, D., 2007. Tonal and articulatory marking of focus in German. In: Proceedings 16th ICPhS, Saarbrücken, pp. 1029–1032.
Baumann, S., Grice, M., 2006. The intonation of accessibility. Journal of Pragmatics 38 (10), 1636–1657.
Beckman, M., Hirschberg, J., 1994. The ToBI Annotation Conventions. Ohio State University, http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.
Beckman, M.E., Hirschberg, J., Shattuck-Hufnagel, S., 2005. The Original ToBI System and the evolution of the ToBI framework. In: Jun, S. (Ed.), Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press, Oxford, pp. 9–54.
Beckman, M.E., Pierrehumbert, J., 1986. Intonational Structure in English and Japanese, Phonology Yearbook 3. , pp. 255–309.
Bolinger, D., 1986. Intonation and Its Parts. Stanford University Press, Palo Alto.
Brown, G., Currie, K.L., Kenworthy, J., 1980. Questions of Intonation. Croom Helm, London.
Féry, C., 1993. German Intonational Patterns. Niemeyer, Tübingen.
Féry, C., Kügler, F., 2008. Pitch accent scaling on given, new and focused constituents, in German. Journal of Phonetics 36 (4), 680–703.

Grabe, E., 1998. Comparative Intonational Phonology: English and German. MPI Series in Psycholinguistics 7. Ponsen en Looien, Wageningen.

Grice, M., 1995. Leading tones and downstep in English. Phonology 12.2, 183–233.

Grice, M., Baumann, S., Benzmüller, R., 2005. German intonation in autosegmental-metrical phonology. In: Jun, S.-A. (Ed.), Prosodic Typology. The Phonology of Intonation and Phrasing. Oxford University Press, Oxford, pp. 55–83.

Grice, M., Baumann, S., Jagdfeld, N., 2007. Evidence for tonal identity from peak scaling under pitch span variation. In: Proceedings 16th ICPhS, Saarbrücken, pp. 977–980.

Grice, M., Ladd, D.R., Arvaniti, A., 2000. On the place of phrase accents in intonational phonology. Phonology 17 (2), 143–185.

GToBI Guidelines and Training Materials, http://www.uni-koeln.de/phil-fak/phonetik/gtobi/training/index_training.html.

Gussenhoven, C., 1984. On the Grammar and Semantics of Sentence Accents. Foris, Dordrecht.

Gussenhoven, C., 1985. Intonation: a whole autosegmental language. In: van der Hulst, H.,Smith, N. (Eds.), Advances in Nonlinear Phonology. Foris, Dordrecht.

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

Gussenhoven, C., Rietveld, T., 2000. The behavior of H* and L* under variations in pitch range in Dutch rising contours. Language and Speech 43, 183–203.

Kügler, F., 2006. L-tone affixation: evidence from German dialects. In: Proceedings Speech Prosody, Dresden, pp. 201–204.

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

Ladd, D.R., 1993. Constraints on the gradient variability of pitch range (or) Pitch Level 4 lives! In: Keating, P. (Ed.), Papers in Laboratory Phonology III. Cambridge University Press, Cambridge, pp. 43–63.

Ladd, D.R., 1996. Intonational Phonology. Cambridge University Press, Cambridge.

Ladd, D.R., Morton, R., 1997. The perception of intonational emphasis: continuous or categorical? Journal of Phonetics 25, 313–342.

Ladd, D.R., Silverman, K., Tolkmitt, F., Bergman, G., Scherer, K., 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signalling speaker affect. Journal of the Acoustical Society of America 78, 435–444.

Nolan, F., 2003. Intonational equivalence: an experimental evaluation of pitch scales. In: Proceedings 15th ICPhS, Barcelona, pp. 771–774.

Patterson, D., Ladd, D.R., 1999. Pitch range modelling: linguistic dimensions of variation. In: Proceedings 14th ICPhS, San Francisco, pp. 1169–1172.

Pierrehumbert, J., 1980. The Phonetics and Phonology of English Intonation. Ph.D. Thesis, MIT.

Rathcke, T., Harrington, J., 2007. The phonetics and phonology of high and low tones in two falling f0-contours in Standard German. Interspeech 2007, Antwerp.

Rathcke, T., Harrington, J. The variability of early accent peaks in Standard German. In: Fougeron, C. (Ed.), Laboratory Phonology 10 Mouton, to appear.

Schröder, M., 2004. Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis. Ph.D. Thesis, PHONUS 7, Research Report of the Institute of Phonetics, Saarland University.

Shattuck-Hufnagel, S., Dilley, L., Veilleux, N., Brugos, A., Speer, R., 2004. F0 peaks and valleys aligned with non-prominent syllables can influence perceived prominence in adjacent syllables. In: Proceedings Speech Prosody 2004. pp. 705–708.

Shattuck-Hufnagel, S., Ostendorf, M., Ross, K., 1994. Stress shift and early pitch accent placement in lexical items in American English. Journal of Phonetics 22, 357–388.

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

Uhmann, S., 1991. Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie. Niemeyer, Tübingen.

van den Berg, R., Gussenhoven, C., Rietveld, T., 1992. Downstep in Dutch: implications for a model. In: Docherty, G.J., Ladd, D.R. (Eds.), Papers in Laboratory Phonology II. Cambridge University Press, Cambridge, pp. 335–359.