Late pitch accents in hat and dip intonation patterns

Niebuhr, Oliver; Zellers, Margaret

Published in:
Understanding Prosody - The role of context, function and communication

Publication date:
2012

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
1 Introduction

1.1 On the forms and meanings of German pitch accents

A number of acoustic, perceptual, and functional studies have demonstrated the existence of three pitch-accent categories in German (e.g., Kohler 1987, 2005; Dombrowski 2003; Niebuhr 2007a,b). The first category (a) is used by speakers to characterize the accented information as fixed and settled and hence as not worth further discussion. The second pitch-accent category (b) puts information up for discussion. This not only includes opening a new line of argument, and also taking up an item of information again in order to discuss or present it in a new light. The third pitch-accent category (c) builds on the meaning of the second category. However, it additionally indicates that a piece of information related to the accented word violates the speaker’s expectation.

Regarding these outlined meanings, it is tempting to associate the pitch-accent categories with the signaling of information structure (as defined by Halliday 1967) such that categories (a) and (b) co-occur with given and new information, while category (c) is used in connection with contrastive focus. Indeed, such correlations between pitch-accent category and information-structure condition have been found, but they were far from perfect (cf. Féry 1993; Röhr and Baumann 2010). The reason for these nondeterministic correlations is that the pitch-accent meanings go beyond mere information-structure concepts. For example, the sentence Bier ist ausverkauft (‘Beer is sold out’) from the ‘Kiesel’ corpus (cf. Niebuhr 2010) was produced with pitch-accent category (a) on ausverkauft (‘sold out’), even though this information was newly introduced into the discourse. By using category-(a), the speaker signals to the listener that the unavailability of beer is reliable information that will not change in the (near) future. Further discussion is not necessary. Conversely, the sentence Okay, wir treffen uns am Montag (‘Ok, we meet on Monday’) in the Kiel corpus of spontaneous speech (IPDS 1996) showed pitch-accent category (b)
on Montag (‘Monday’), although this date had already been agreed upon in the previous discussion and, as such, is information that could be taken as given. The application of pitch-accent category (b) reflects that the speaker is going to deal with the given date from a different angle, viz. travelling. In line with this intention the summarizing statement was followed by the question Und wie kommen wir da hin? (‘And how do we get there?’). Furthermore, the ‘Kiesel’ sentence Benzín kostet 2 Euro pro Liter (‘Gasoline costs 2 Euro per liter’) was typically realized by speakers with two prenuclear category-(c) pitch accents on 2 and Euro, despite the absence of any explicit contrast to previously introduced information. Moreover, unlike in the contrastive focus productions of the same sentence, the category-(c) pitch accents were followed by a clear nuclear (category-a) pitch accent on Liter, i.e. the entire sentence was put into focus. The reason for employing the category-(c) accent was to signal the listener that the price of the gasoline differed from the expectation of the speaker.

In order to account for their wider scope of application in speech communication, it has been claimed that the three German pitch-accent categories are not used for information-structure purposes, but as part of the speaker’s argumentation structure. On this basis, the pitch-accent meanings were labeled as ‘concluding’ (a), ‘opening’ (b), and ‘astonished’ (c) (cf. Kohler 2005; Niebuhr 2007a,b). The linguistic concept of astonishment must not be confused with the basic emotional category of surprise. From the point of view of phonetic form, linguistic astonishment is signaled by a local pitch accent contour, whereas emotional surprise is characterized by phrase-level exponents like strongly variable intonation at a high level, slow speaking rate, and breathy voice (cf. Hammerschmidt and Jürgens 2007). Functionally, the emotional category of surprise says something about the speaker and his/her mental state. Linguistic astonishment, on the other hand, is oriented towards the propositional content of the speaker’s message. It relates a local piece of information to the overall argumentation by denoting a factual mismatch between the speaker and the external world. Kohler (2005:89) referred to this mismatch as “realizing in contrast to one’s expectation”. Other researchers, such as Dombrowski (2003), have used similar paraphrases. Dombrowski explicitly demonstrated the difference between astonishment and surprise by applying a semantic-differential task to a stimulus continuum, showing that a change towards a category-(c) pitch accent does not make listeners perceive the corresponding stimulus utterances as more surprised (on a scale from not-surprised to surprised) or as more emotional (on a scale from neutral to emotional).
As linguistic astonishment is separate from emotion, it can, but need not co-occur with different emotional categories. For example, Pfitzinger and Kaernbach (2008) elicited the sentence *Er hat es wirklich getan* (‘he really did it’) with linguistic astonishment and the corresponding category-(c) pitch accent on *wirklich* (‘really’), which indicates that the speaker accepted a given fact, but expected a different behaviour from the person being referred to. In addition to these emotionally neutral elicitations, the actors who produced the sentence added different emotions like fear, joy, anger, sadness, and surprise, each with specific global phonetic changes, beyond applying the category-(c) pitch accent indicating linguistic astonishment.

Figure 1 shows a phonetic manifestation of the tripartite pitch-accent contrast (a)-(c) in the German utterance *weil du Rotlicht magst* (‘because you like red light’), which was produced naturally by the first author with an accent on the initial syllable *Rot* ([ʊʊ t ʰ], ‘red’) of the disyllabic word *Rotlicht*. The most obvious difference between the pitch accents lies in their F0-peak alignment. In (a) the F0 rise ends right after the accented-syllable onset, so F0 falls into the accented vowel. In contrast, (c) shows a rise that starts within the accented vowel and extends beyond the accented-syllable offset. The rise of (b) is located in between those of (a) and (c) and therefore reaches its maximum within the vowel boundaries.

![](image1.png)

Figure 1: F0 course (top), and spectrogram (bottom) of the German utterance *weil du Rotlicht magst* (‘because you like red light’), produced with a tripartite pitch-accent contrast (a-c) on *Rot* ([ʊʊ t ʰ]), indicated by continuous vertical lines. Dotted lines mark the accented vowel.

1.2 Predictions of phonological models

The form–meaning relations that are represented by (a), (b), and (c) in 1.1 are accounted for in two important theoretical frameworks for intonation research in German: in the Kiel Intonation Model, KIM (Kohler 2006),
and in GToBI, which is based on the autosegmental-metrical framework (Grice and Baumann 2002; cf. also Pierrehumbert 1980). In the following, the KIM categories 'early', 'medial', and 'late' will be used as theory-neutral labels to refer to the three pitch accents. Although they differ in the concrete phonological modelling of the pitch-accent categories, GToBI and the KIM agree in three central aspects and hence make three common predictions.

(1) ‘Prenuclear accents’: The same pitch-accent categories can occur in nuclear as well as in prenuclear position. That is, unlike the intonational approach of the British School (cf. O'Connor and Arnold 1973), prenuclear and nuclear accents have the same formal status in the intonation phrase and share the same set of paradigmatic oppositions.

(2) ‘Phonological points of reference in the segmental string’: The only phonologically relevant segmental references for pitch accents are the accented syllable or its vowel. Thus, the signalling of a pitch-accent category requires a certain F0 pattern showing certain temporal relationships with the syllable and/or vowel boundaries. As long as these relationships remain constant, different accented-word boundaries do not affect the pitch-accent identity. For example, by replacing the initial lateral /l/ in the second syllable of *Rotlicht* with the nasal /n/, the utterance in Figure 1 can be changed into *weil du rot nicht magst* ('because you don’t like red'). That is, the former disyllabic accented word is split up into two monosyllabic words, an accented and an unaccented one. However, this shift of the final accented-word boundary should not affect the pitch accent identity triggered by the F0 patterns (a)-(c).

(3) ‘Relevance of an F0 fall’: The signaling of the late pitch-accent category is based on a rising pitch movement, which starts within the accented vowel and continues beyond the accented-syllable offset. The phonetic properties or the mere presence of a falling movement are irrelevant.

All the studies that have provided the empirical basis for the distinction of the early, medial, and late pitch accents and investigated their signalling in more detail were concerned with nuclear accents. So prediction (1), according to which the members of the tripartite pitch-accent contrast also occur in prenuclear position, has never been tested exhaustively. In addition, some recent studies, summarized in sections 1.4 and 1.5, raise doubts about the general validity of predictions (2) and (3). Against this backdrop, it is the aim of the present study to test predictions (1)-(3) by means of a perception experiment that investigates the late pitch accent in prenuclear position.

Any study examining the signaling of pitch-accent categories or pursuing the question of additional categories requires a criterion that allows
assigning a spectrum of phonetic intonation patterns to a specific pitch accent and vice versa. Such a criterion cannot come from the phonological forms or the phonetic patterns themselves without circular reasoning (cf. Xu 2004). The present work follows the methodological approaches of Kohler (1987, 2005) and Niebuhr (2007a,b) and uses meaning as the external point of reference. Since the meanings of German pitch accents are relatively well understood, as was exemplified in 1.1, they form a solid basis for our study. The identification of a specific pitch-accent category is defined as the identification of a specific meaning in the stimuli of the perception experiment. In the case of the late pitch accent, this meaning may be outlined by the term ‘astonished’ in the previously-explained linguistic sense. Starting from this meaning-based perspective, concrete experimental hypotheses are put forward in the following sections, one for each of the three theoretical predictions of KIM and GToBI.

1.3 Hypothesis (1)

It is clear that prenuclear pitch accents exist in German, and their phonetic manifestations have been intensively studied (e.g., Atterer and Ladd 2004; Kleber and Rathcke 2008). However, since these investigations were primarily acoustic, there is currently no experimental evidence that these prenuclear accents convey the same attitudinal meanings as early, medial, or late nuclear accents. It is only known that accented syllables in prenuclear position can show pitch patterns that are comparable to those of all three nuclear accents, i.e. early, medial, and late. (cf. Peters et al. 2005). On this basis, hypothesis (1) is put forward with reference to the medial-late contrast (cf. Fig. 1b-c) and its current phonological modelling reflected in the third prediction: A delay of a prenuclear accentual F0 rise that simultaneously shifts the rise onset into the accented vowel and the rise offset out of the accented syllable can cause a change in the interpretation of the stimulus utterances towards ‘astonished’.

1.4 Hypothesis (2)

Hypothesis (2) is as follows: For a constant prenuclear accentual rise that starts in the accented vowel and that continues beyond the accented syllable, the stimulus utterances sound the more astonished, the more of the F0 fall is completed before the end of the accented word. Some previous research gives substance to this hypothesis that contradicts prediction (2).
Niebuhr and Ambrazaitis (2006) investigated the F0 patterns of prenuclear pitch accents in German that show a late alignment and a dipped concatenation with the following accent. They found that the F0 fall of this late-aligned accent was steeper when there were fewer unaccented syllables in between the accented one and the end of the corresponding accented word. Based on this observation we performed a systematic acoustic analysis with XASSP (cf. Scheffers and Thon 1991), using two German speech corpora consisting of informal dialogues: the ‘Lindenstraße’ corpus (cf. IPDS 2006) and the ‘Kiesel’ corpus (cf. Niebuhr 2010). Since both corpora were labelled prosodically with the KIM-based PROLAB system (cf. Kohler 1997), the instances of late-aligned prenuclear pitch accents were easy to identify reliably. Only pitch accents that were based on regularly accented (i.e. neither partially de-accented, nor emphatically accented) syllables, and that were separated from the following, nuclear accent by a word boundary and at least three unaccented syllables, were included in the study. This threshold aimed to avoid potential word-boundary induced variation of the falling slope being masked by variation due to time pressure and tonal crowding (cf. Caspers and van Heuven 1993). The selection was further restricted to those accents that were concatenated with the following one by a low-falling indentation. The tokens were divided into three conditions: word-final accented syllable (0-syllable condition, 21 tokens), one unaccented syllable preceding word boundary (1-syllable condition, 29 tokens), and two or more unaccented syllables preceding word boundary (2+-syllables condition, 13 tokens). The 63 tokens came from 19 different speakers, 9 females and 10 males. The samples for each condition included female and male productions. Therefore the F0 differences that were initially measured in Hz were later recalculated in terms of semitones (st, relative to 100 Hz).

Figure 2: Schematically illustrated late-aligned prenuclear pitch patterns across the accented syllable (acc), the final accented-word boundary (WB) and the subsequent post-accented syllables (pacc1,2,3) that represent the unaccented-syllable conditions 0, 1, 2+. Dotted lines indicate that the F0 fall ended in a variable position after WB. Arrows (a,b,c) point to significant differences between the F0 patterns in the conditions 0, 1, 2+ (arrow lengths are not proportional).
A number of time and F0 measurements were taken for each pitch accent, and a series of one-way ANOVAs were calculated; each of them addressed the effects of the unaccented-syllable conditions (subsumed under a 3-level independent variable) on a different acoustic measure (dependent variable). The results of the acoustic analysis are outlined in Figure 2.

The duration of the F0 fall of the late-aligned pitch accent increased significantly with a larger number of syllables between the accented syllable and the end of the corresponding accented word (cf. arrow (a) in Fig. 2). At the same time, the end of the fall was consistently found after the end of the word, in a variable interval that was neither affected systematically by the three unaccented-syllable conditions nor pointed to a stable alignment relative to any segment-based phonological unit. Closer inspections of individual cases suggested that the alignment of the fall offset after the accented word was mainly determined by the category of the following, nuclear pitch-accent. The fall offset occurred earlier when followed by an early than by a medial or late nuclear accent. The duration of the rise also increased slightly from the 0-syllable to the 2+-syllable condition, while its onset was always located closely after that of the accented vowel (cf. Atterer and Ladd 2004). The end of the rise was defined by the onset of the fall. Therefore, the significant increase in the duration, which is represented by arrow (b) in Figure 2, actually involved two separate phenomena. Firstly, the rise was successively extended (further) beyond the accented syllable. Secondly, the transition phase from the rising to the falling movement became successively longer, so the F0 peak maximum was more and more flattened. However, the range of the rise remained constant across the three unaccented-syllable conditions. The range of the fall up to the end of the first post-accented syllable decreased significantly for a larger number of unaccented syllables between the accented one and the final word-boundary, which fits in with the significant increase in fall duration (cf. arrow (c) in Fig. 2).

The significant differences in the acoustic measurements appear to reflect a single mechanism in the time domain which compresses the pitch-accent movements for shorter intervals between the accented syllable and the end of the accented word. This compression and the resulting change in the overall contour dynamics may be conceptualized as an adjustment of the falling movement. A certain amount of the falling movement must be completed within the word, while the end of the fall must be reached after the final word boundary. This adjustment between the temporal domain of the late-aligned prenuclear pitch pattern and the temporal domain of the corresponding accented word implies a phonological relationship between pitch accents and segmental references that goes beyond predic-
1.5 Hypothesis (3)

Hypothesis (3) constrains the general hypothesis (1): *The delay of a prenuclear accentual F0 rise changes the interpretation of the stimulus utterances towards ‘astonished’ solely in a dip pattern, in which the rise is followed by a fall. In a hat pattern, the same delay of the rise does not affect the astonishment conveyed by the stimulus.* This hypothesis contradicts prediction (3), on the basis of the following evidence. Figure 3 displays four different stylized intonation patterns. Each of them comprises two pitch accents. All four patterns are quite frequent in German (cf. Peters et al. 2005). Prenuclear pitch accents may be concatenated with following accents in two ways: with or without a dip. The latter pattern may be referred to as the ‘hat pattern’ (cf. ’t Hart et al. 1990; Adriaens 1991). Both GToBI and the KIM, despite conceptual differences as to the phonological status of the hat-dip contrast, project the delay of the initial, prenuclear pitch accent in the hat and the dip patterns of Figure 3 onto the same phonological difference: that between medial and late.

In the case of the dip pattern, we may well assume a medial-late contrast, since the two rising-falling F0 movements and their temporal relations to the accented syllables are comparable to those of the nuclear accent patterns (b) and (c) in Figure 1. The parallel phonological analysis of the two initial F0 movements in the hat pattern as medial and late pitch accents is crucially related to prediction (3). On the basis of his experiments, Niebuhr (2007b) claims that – unlike in the case of the medial pitch accent – the presence of an F0 fall is obligatory for the perception of the late category. Further evidence against prediction (3) comes from Peters et al. (2005). They observed that sequences of late and early pitch accents are rather rare in dip patterns, whereas they are the most frequent pitch-accent combination in hat patterns. In connection with this, Peters et al. noted that the meanings of late and early seemed to conflict in dip patterns, as may be expected from their meanings (cf. 1.1). This meaning conflict is absent in the hat pattern.
Late pitch accents in hat and dip intonation patterns

Figure 3: Frequent two-accent intonation patterns of German. The patterns (1) and (2) at the top are dip patterns, whereas (3) and (4) below represent hat patterns. The delays of the initial rising(-falling) F0 movements (dotted lines) are analyzed as changes in the prenuclear pitch accent category from medial to late in both the dip and the hat pattern. The ‘n’ values specify the frequencies of occurrence of the patterns within two-accent phrases of the Kiel corpus of spontaneous speech (e.g., IPDS 1996, 2006).

2 Method

2.1 Stimulus utterances

The experimental stimuli are based on three different utterances that were produced naturally by the first author, a trained phonetician.

- Der [Müll] | wur| de | bent’ | abgehol|t (‘The trash was picked up today’), 0-syllable condition;
- Der [Müll|er] | wur| de | abgehol|t (‘The miller was picked up’), 1-syllable condition;
- Die [Müll|er in] | wird | abgehol|t (‘The miller’s wife is picked up’), 2-syllable condition.
As is indicated by the underlined segments, each of the utterances was realized with two comparably prominent pitch accents, a medial prenuclear accent on Müll ([m*yl]) and an early nuclear accent on ab ([ab]). The F0 peaks of the two accents were separated by a moderate indentation, and the maximum of the second peak was realized about 1 st lower than the first one and followed by a terminal fall towards the end of the utterance. Each of the utterances had three unaccented syllables in between Müll and ab, cf. vertical bars in the utterance list above. But the crucial point is that these unaccented syllables are differently distributed across the words, cf. bracketing in the utterance list. In the case of the third utterance, two of the unaccented syllables and the preceding accented Müll together constituted the accented word Müllerin. This is referred to as the 2-syllable condition. The second utterance may be referred to as the 1-syllable condition. That is, one unaccented syllable follows the accented one within the same word Müller, whereas the remaining two unaccented syllables form a separate word. The first utterance represents the 0-syllable condition, in which the first accented syllable Müll was at the same time the complete accented word.

As can be seen in Figure 4, the three utterances were produced as comparable as possible with regard to the pronunciation of the individual sound segments as well as with regard to the overall voice quality and duration structure. The accented syllables Müll differed by at most 25 ms between the utterances (272 ms – 297 ms). The intervals between Müll and the subsequently accented ab differed by even less than 20 ms, i.e. they were between 569 ms – 586 ms in all utterances. The durations of each of the three unaccented syllables in between Müll and ab deviated by not more than 50 ms across the utterances. The same holds for the overall duration of the utterances, which varied between 1619 ms – 1748 ms. Some of the temporal differences between the utterances are likely due to the long known cross-linguistic phenomenon that the durations of syllables and segments (particularly of vowels) decrease as word length increases, as in Müll, Müller, and Müllerin (cf. Menzerath 1954; Klatt 1979). For example, it is obvious from Figure 4 that the syllable er is shorter in Müllerin than in Müller. The accented syllable Müll also becomes successively shorter across the three utterances, although this shortening is less pronounced than in er. Other duration differences may be ascribed to differences in the syllable structure. However, these differences did not interfere with the experiment; rather, they contributed to the naturalness of the utterances and counteracted differences in the perceived speaking rate.
Late pitch accents in hat and dip intonation patterns

2.2 F0 manipulation

By means of F0 manipulations on the basis of the PSOLA resynthesis in PRAAT (cf. Boersma 2001), two different stimulus series were generated for each of the three utterances. The ‘hat’ series, which addressed hypothesis (3), will be dealt with first. The original dip pattern of the utterance *Die Müllerin wird abgeholt* was stylized at 9 contour points. Four of them were used to create a hat pattern that took into account the naturally produced alignments of the medial F0 rise and the early F0 fall, as well as the ranges of the two pitch accents. As in the original production, the early F0 fall started around 1st below the maximum of the preceding medial F0 rise. Two further important characteristics of the original dip pattern were kept in the created hat pattern by means of the remaining five
contour points. First, the moderate declination was imitated by placing the two contour points at the beginning and at the end of the utterance at F0 values of 100 Hz and 80 Hz. The other three contour points formed a small rising-falling F0 peak on the second syllable *boll* of *abgeholt*. This small peak is due to a partial accentuation, a typical feature of past participle verb forms in German (cf. Kohler 2006; ‘phrase accent’ in GToBI, Grice and Baumann 2002).

On the basis of this first hat pattern, a second hat pattern was created in which the initial, prenuclear rise was delayed by 110 ms. The delayed rise no longer reached from the accented-syllable onset to the end of the accented vowel; instead, it started in the accented vowel and continued beyond the syllable offset into the vowel of the first post-accentual syllable. The rise duration of 120 ms and its range of 11 st remained constant. The delay of the prenuclear rise would be analyzed phonologically as a change from the medial to the late pitch-accent category (cf. Fig. 3). Figure 5a illustrates the two created hat patterns in the stimulus utterance *Die Müllerin wird abgeholt*; ‘hat1’ represents the medial-early pattern, whereas ‘hat2’ represents the late-early one. The same hat1, hat2 patterns were analogously created in the other two utterances by placing contour points with the same frequency values at the same temporal distances from the vowel onsets of the corresponding syllables. The cross-combination of hat patterns and utterances yielded a total set of 6 stimuli.

Figure 5: F0 contours (in st) created in each of the three utterances for the ‘hat’ (a, left) and the ‘dip’ (b, right) stimulus series of the perception experiment. The illustration is based on the utterance *Die Müllerin wird abgeholt*. The ‘hat’ series contains two different patterns that are analyzed as medial-early (‘hat1’) and late-early (‘hat2’). The ‘dip’ series results from 7 equal-sized shifts of the F0 fall in the context of the late pitch accent of the first accented syllable *Müll.*

A second set of stimuli, the ‘dip’ series, addressed hypotheses (1) and (2), cf. Figure 5b. It was based on the same F0 contour frame as the hat series. That is, the F0 courses before the first and after the second pitch accent were identical. Inside the contour frame the dip series was characterized
Late pitch accents in hat and dip intonation patterns

by two major differences from the hat series. Firstly, the F0 rise of the prenuclear pitch accent, which again comprised 11 st, was not shifted, but showed a constant alignment relative to the accented syllable and its vowel. The alignment corresponds to the late pitch accent category. While the rise was fixed, the subsequent, falling F0 movement was shifted towards the rise, in this way successively reducing the high plateau in between rise and fall. This leftward shift of the fall comprised 7 steps of 55 ms. As they only involve a single pitch accent, neither of the plateau patterns represents a hat pattern. The last shift 7 created a pointed, late-aligned F0 peak.

As for the following nuclear accent on ab, a separate rising and terminal falling F0 peak was created. The rising peak movement set in right before the onset of ab, and went over 8 st into the second half of the accented vowel. Alignment, shape, and height properties of the nuclear F0 peak on ab were known to be a clear signal for the medial pitch-accent category (Kohler 1987; Niebuhr 2007a,b). Moreover, the peak properties ensured that the prominence of the nuclear accent was comparable with that of the prenuclear accent. Finally, it was ensured during the whole F0 manipulation that the slopes of the F0 movements remained below the upper limit estimated by Xu and Sun (2002).

There are several reasons why a plateau-to-peak continuum was created for the prenuclear pitch accent in the dip series. First, with regard to hypothesis (2), the aim of the present study was to investigate whether the timing of the F0 fall relative to the final accented-word boundary influences the perceptual identification of the initial (prenuclear) the pitch accent as 'late'. Therefore, it was important to avoid confounding the variations in the timing with variation in the shape (i.e. in the slope) of the fall, which is known to affect pitch-accent identification in German (cf. Niebuhr 2007a), and which would have been created if the fall offset was shifted. Secondly, it was observed in the acoustic data (cf. 1.3) that the F0 peak maxima of the late pitch accents became more and more flattened as the falling movement was delayed and expanded to the right. The plateau-like F0 courses in the dip stimuli account for this observation, although in a stylized and hence probably exaggerated way.

Although the hat and dip stimulus series vary in the nuclear accent category – early in the hat and medial in the dip stimuli – this should not affect the subjects’ judgments, since neither category is related to astonishment (cf. 1.1). Care was also taken to exclude differences in perceived phrasing between the hat and dip series. Prior to the experiment, a group of 5 prosodically experienced phoneticians confirmed that none of the
created hat and dip patterns led to a phrase boundary in between the prenuclear and nuclear pitch accents.

As in the hat series, the 7 dip patterns shown in Figure 5b were first generated for the utterance *Die Müllerin wird abgeholt* and then transferred to the other two utterances, keeping the frequency values of the contour points and their alignments relative to the vowel onsets constant. This led to a total number of 7x3=21 stimuli for the dip series. The decisive result of the contour transfer was that the position of the F0 fall relative of the end of the accented word was – for each of the 7 intervals between the rise and the fall (dip1-dip7) – earlier in the *Müllerin* than in the *Müller* stimuli (2-syllable vs. 1-syllable condition), and again earlier in the *Müller* than in the *Müll* stimuli (1-syllable vs. 0-syllable condition).

### 2.3 Perception experiment

The perception experiment contained each of the 27 (6+21) stimuli 5 times in a randomized order. The resulting 135 stimuli were grouped into 9 blocks of 15 stimuli. Each stimulus was followed by a pause of four seconds in which it was judged by the subjects. 27 subjects took part in the experiment, 18 females and 9 males. They were all native speakers of German, naive with regard to the aim of the experiment, and with no reported hearing disorders. The subjects were between 19-33 years old, and most were undergraduate students of phonetics, linguistics or psychology at the University of Kiel.

The experiment was conducted in sessions with subsets of 5-8 subjects. Each session started with a constant set of written instructions that was read to the subjects. These instructions explained that German speech melodies convey meanings in addition to those of the string of words. These melodic meanings contribute to the argumentation of the speaker by signalling his/her expectations regarding the semantic content of the utterance, the dialog partner, or the dialogue itself. One of these meanings may be summarized by the term ‘astonished’ (*erstaunt*). Astonishment was explicitly and clearly defined in this context as: *The lexical information of the stimulus was not expected by the speaker.*

The subjects were further informed that their task in the experiment was to judge, if and to what extent this kind of astonishment is expressed by the male speaker in the stimulus utterances. In order to do this, they received a number of scales on a sheet of paper, one scale for each stimulus. The scales reached from 0 to 4, with 0 meaning ‘The speaker is not astonished at all; he is just stating a fact’, and 4 meaning ‘The speaker is
very astonished; he is not just stating a fact’. After listening to each stimulus, the subjects had to mark the appropriate number on the scale. Since the task required absolute judgments, it was expected that scalar judgments would be easier for subjects to apply to the stimuli than binary judgments, as in a 2AFC task. Moreover, a judgment scale should be a more sensitive instrument to detect changes in the argumentation structure of the stimuli. The scale started at 0, since the perception experiment also comprised stimuli whose intonation contours were suitable to signal only medial and early, but no late pitch accents (e.g., hat1). Therefore, the subjects were given the possibility to indicate the complete absence of astonishment.

In order to facilitate the task, the subjects received the following additional information that was considered helpful in the light of previous experiences with investigating the late pitch-accent category in German. Firstly, it was pointed out to the subjects that an ‘astonished’ speaker can be ‘überrascht’ in the sense that a given fact comes as unexpected while remaining readily acceptable. Alternatively, an ‘astonished’ speaker can be ‘ungläubig, entrüstet’ in the sense that a given fact comes as unexpected and is only reluctantly accepted. Due to the minimal semantic-pragmatic context provided by the stimuli, astonishment might be ‘überrascht’ for some subjects, and for others it might be ‘ungläubig, entrüstet’. Therefore both astonishment readings were to be equally judged as ‘astonished’. Secondly, the subjects were told that astonishment in the sense of unexpectedness is frequently accompanied in German by modal particles like ‘doch’, ‘aber’, ‘ja’, ‘tatsächlich’ and ‘wirklich’ (cf. Ickler 1994). Those particles would not be contained in the stimulus utterances. However, in deciding whether or not a stimulus conveyed astonishment, the subjects could ask themselves if the speaker could have used such a modal particle.

After the instructions, but before the actual experiment, the subjects heard and judged 15 practice items that were randomly selected for each session. Then, they had the chance to ask final questions. The 15 and 135 randomized practice and experimental items were contained in separate ‘wav’ files and played from a laptop via Hi-Fi loudspeakers with a constant, pre-adjusted loudness. The experimental session took place in a sound-treated room of the (former) Institute of Phonetics and Digital Speech Processing at the University of Kiel.
3 Results

The results of the descriptive analysis are summarized in Figure 6. It displays for each stimulus the mean value on the ‘astonished’ scale that was yielded across all 27 subjects. Error bars represent standard deviations. The values for the two hat-pattern stimuli hat1 and hat2 that differed in the alignment of the prenuclear F0 rise are given on the right-hand side of the Figure. The dip1-dip7 stimuli are shown on the left-hand side; dip1 refers to the longest high plateau between the late-aligned prenuclear rise and the fall, dip7 contains the pointed rising-falling peak with the same late alignment as dip1. Each value summarizes 135 judgments.

Figure 6: Means and standard deviations of the ‘astonished’ judgments yielded across all 27 subjects for the 2 stimuli of the hat-pattern series (right) and for the 7 stimuli of the dip series (left). The results are shown separately for each utterance or unaccented-syllable condition (0,1,2+), represented by the words Müll, Müller, and Müllerin. Each data point and error bar represents 135 judgments.
As for the inferential statistics, two different analyses were performed. The dip1-dip7 stimuli were analyzed by means of a univariate multinomial logistic regression (mlogit) analysis. It aimed at hypothesis (2) and was based on two independent variables (predictors). The variable ‘stimulus number’ represented the continuum from pointed peak (dip7) to the longest high plateau (dip1) and included 6 levels, with dip1 as reference level. The variable ‘stimulus utterance’ corresponded to the three unaccented-syllable conditions. It was divided into two binary (i.e. 2-level) variables, each with the 1-syllable condition as reference level. The scale values of the stimuli summed up across all subjects were used as the dependent variable.

In order to analyze the hat-pattern results, two t-tests for paired samples were performed. Following hypotheses (1) and (3), they compared the ‘astonished’ judgments of hat1 and hat2, as well as of hat2 and dip1. The two t-tests were based on the scale sums received for the 27 subjects across all three utterances.

The results of the inferential statistics are given in Table 1.

Figure 6 illustrates that the mean values of the hat-pattern stimuli hat1 and hat2 are all within a narrow range of about 0.5 scale points in the lower half of the ‘astonished’ scale, i.e. between 1.3 and 1.7. Correspondingly, the t test showed that the delay of the initial rise from hat1 to hat2 did not significantly increase the ‘astonished’ judgments (cf. Tab.1). The opposite holds for the change from hat2 to dip1. The introduction of an F0 fall after the delayed F0 rise raised the astonishment that was perceived in the stimuli significantly.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>t</th>
<th>p</th>
<th>Mlogit analysis for hypothesis (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Var.</td>
</tr>
<tr>
<td>hat1 vs hat2</td>
<td>0.122</td>
<td>&gt;0.050</td>
<td></td>
</tr>
<tr>
<td>hat2 vs dip1</td>
<td>-3.652</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>intercept</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>stim. num.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 vs 0 syll.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 vs 2 syll.</td>
</tr>
</tbody>
</table>

Table 1: Left: paired-sample t tests comparing hat1 and hat2 as well as hat2 and dip1, based on the respective scale values, summed up for each subject (hence \( df = 26 \)). Right: mlogit analysis based on the two independent variables ‘stimulus number’ and ‘stimulus utterance; the sums of scale values received for all 27 subjects served as dependent variable.
Within the dip series the astonishment that was conveyed by the stimuli increased successively from dip1 to dip7, i.e. with the stepwise leftward shift of the falling movement towards the prenuclear rise. In line with this overall pattern, it was found in the mlogit analysis that the variable ‘stimulus number’ contributed significantly to the general linear-regression model (cf. Tab.1). In addition to this general increase in astonishment it may be seen in Figure 6 that the unaccented-syllable conditions, i.e. the three accented words Müll, Müller, and Müllerin, yielded different overall levels of astonishment. The stimuli of the 2-syllable condition (Müllerin) were generally perceived to be more astonished than the stimuli of the 1-syllable condition (Müller), which in turn yielded higher astonishment judgments than the stimuli of the 0-syllable condition (Müll). Consequently, the mean ‘astonished’ values of the Müllerin stimuli all exceeded the centre scale value of 2.0 and some dip7 stimuli received a 4 on the scale, whereas the mean values of the Müll stimuli were all located in the lower half of the scale and only increased to approximately 2.0 towards the ‘pointed’ end of the series (i.e. dip7), in which the fall came closest to the rise and the final word boundary.

4 Discussion

4.1 Assessing the judgment behavior

The analyses of the results yielded clearly significant differences and changes in judgment behavior between the hat and dip stimuli as well as between the stimuli of the dip series. However, it can also be seen in Figure 6 that almost all of the subjects’ judgments fall in a range of scale values from 1 to 3. A reviewer has pointed out that this limited use of the scale could suggest that the subjects were highly uncertain about the task, and in particular, about the interpretation of ‘astonishment’. We disagree with this conclusion for several reasons.

Firstly, as for the stimulus judgments, it is true that ‘erstaunt’ and ‘astonished’ may both cause ad-hoc associations with emotion. However, the English translation that we provided may be misleading in this respect insofar as its emotional connotation should be stronger than that of the critical German word ‘erstaunt’. For instance, the English Merriam-Webster’s Dictionary (2009) describes ‘astonished’ by “to strike with sudden and usually great wonder or surprise”, whereas the German Duden (Deutsches Universalwörterbuch, 2011) paraphrases ‘erstaunt’ merely by “Verwunderung ausdrückend” (‘wondering’). The stronger emotional connotation of
Late pitch accents in hat and dip intonation patterns

the English word ‘astonished’ is supported by text-based collocation analyses showing that ‘astonished’ in more frequently preceded by adverbs like ‘absolutely’ and ‘quite’ than the German word ‘erstaunt’ (cf. Cosmas II for German, http://www.ids-mannheim.de/kl/projekte/korpora, and CoCA for English (Davies 2008). Likewise, as for the stimulus intonations (cf. 1.1), Dombrowski (2003) found in a semantic-differential task that the intonational change from the medial to the late pitch-accent category in German does not increase the levels of emotion in general and of surprise in particular that are evoked by stimulus utterances. Other studies like that of Pfützinger and Kaernbach (2008) showed additionally that the late pitch accent category may co-occur with different types of emotions and is hence not bound to a specific emotion (like surprise). In summary, neither the experimental task nor the stimuli themselves required the subjects (implicitly) to assess the emotional arousal of the speaker and to translate this assessment subsequently into an astonishment level. On the contrary, the key instruction that the subjects explicitly received at the beginning of the experiment was simply to rate, if and to what extent the facts stated in the stimuli were unexpected by the speaker. This rating was not directly related to emotion and only in a following step linked to the ‘astonished’ label. We have no indications from the subjects’ questions in between the practice session and the actual experiment that they misinterpreted the instruction in terms of emotional judgments.

Secondly, high uncertainty would manifest itself in inconsistent judgments across subjects and stimulus repetitions, which would lead to large standard deviations. However, as Figure 6 shows, our standard deviations were fairly small, i.e. between 0.5 and 1 scale point. Relative to the range of values covered by the scale, standard deviations of this magnitude are similar or even smaller than those found, for instance, in many previous experiments based on the semantic-differential paradigm (e.g. Dombrowski 2003; Kohler 2005; Ambrazaitis 2005). Making reference to these studies, it seems that our stimuli were judged quite consistently on the ‘astonished’ scale.

Thirdly, it is a well-known phenomenon that subjects avoid using the endpoints of a given scale (cf. Stevens and Greenberg 1966; Friedman and Amoo 1999), particularly when the endpoints are linked with semantically strong, categorical labels, as was the case in our experiment. A possible strategy to compensate for the fact that “subjects tend to avoid extreme responses” (Kahneman and Tversky 1973:248) is to explicitly familiarize them with the full stimulus spectrum prior to the experiment, so that they can learn which stimuli are to represent the left and right endpoints of the given scale (i.e. 0 and 4 in our experiment). However, we refrained from
applying this kind of explicit category learning to our experiment, since such a priori established form-function relations would have seriously undermined the information value of our results. Against this backdrop, the fact that the judgments we elicited covered the full range of non-extreme scale values 1-3 (cf. means) and in some cases also the endpoints 0 and 4 (cf. standard deviations) shows that the subjects were able to apply the scale to the stimuli and make differential judgments despite there not having been any explicitly learned associations between stimulus intonations and scale values.

Fourthly, unlike in many experiments with a semantic-differential paradigm, we did not use a Likert scale that leads from –X via 0 to +X. For such a scale, judgments around the scale centre may indeed be interpreted in terms of uncertainty, since 0 indicates ‘not applicable’ or ‘neither/nor’. In contrast, the centre of our scale represented a specific level of astonishment – a midpoint between ‘not astonished’ and ‘very astonished’.

However, it is possible that our experimental method created artefacts that may have affected the judgment of the stimuli. For example, a considerable part of the instruction was concerned with explaining the concept of astonishment. It is possible that this focus in the instruction hypersensitized the subjects so that they determined the degree of astonishment in each stimulus before asking themselves first whether or not a stimulus actually conveyed astonishment. In this way the instruction may have created a bias towards astonishment and hence towards scale values higher than 0. Moreover, the F0-based stimulus generation started from a medial pitch accent in prenuclear position. Even though empirical evidence leaves no doubt that F0-peak alignment is the most powerful acoustic cue to pitch-accent category in German (cf. Niebuhr 2007b), it may be assumed that our stimuli lacked other concomitant features of the late pitch accent in the vicinity of the accented syllable, such as a specific intensity pattern, a higher F0 peak and, most importantly, a specific duration pattern, including a lengthened accented vowel (cf. 4.4). Therefore, the fact that the subjects hesitated to use the highest astonishment value 4 of the scale may reflect that the F0 manipulation alone was sufficient to identify the late pitch accent, but that a complete perceptual transition from medial to late had required a multiparametric manipulation of the stimuli (in line with Niebuhr 2007a). Finally, it cannot be ruled out (although there is no supporting evidence for it) that scalar (‘astonished’) judgments are harder for subjects to make than binary judgments or scalar judgments with other types of scales (e.g., with more or less numbers or even no numbers at all). It is also possible that there are labels other than ‘aston-
ished’ that our subjects might have found easier to apply to our stimuli. The label ‘unexpected’ is one obvious candidate, which has been used previously in perception experiments on the German medial-late pitch-accent contrast (cf. Niebuhr 2007a). We chose ‘astonished’ since it is currently an established functional label for the German late pitch accent. Moreover, we assumed that astonishment is semantically a less complex concept than unexpectedness. This is assumption would be worth testing in future studies.

In view of these potential shortcomings of our experimental design, it is crucial to note that all of them would only be able to create overall biases in the judgments and/or to mask systematic effects of the stimulus conditions. The potential shortcomings cannot be responsible for the observed effects of the hat vs dip contrast, of the F0-plateau reduction within the dip series, and of the unaccented-syllable conditions. In other words, the observed effects are not experimental artefacts. On this basis, we continue discussion of our hypotheses.

4.2 Hypotheses (1)-(3)

The present study started from hypotheses (1)-(3): (1) The German late pitch accent occurs in prenuclear position of the intonation phrase. It is not only signalled by an accentual rise, which starts in the accented vowel and continues beyond the accented syllable. Rather, the identification of a late pitch accent also requires (3) a falling movement, which (2) must begin before the end of the corresponding accented word. The hypotheses were tested by means of a perception experiment with two-accent stimulus utterances, in which the identification of the late category at the initial accent position was determined indirectly on the basis of its meaning that was summarized by the term ‘astonished’ in the linguistic sense of unexpectedness. In other words, given a scale on which the astonishment conveyed by the stimulus utterances is judged from 0-4, a change towards late pitch-accent identification manifests itself in a change of the judgments towards 4.

The experiment showed that a delay of the prenuclear accentual rise that shifted the rise onset into the vowel and the rise offset beyond the end of the accented syllable caused the subjects’ scalar judgments to change in the direction of 4 (i.e. ‘astonished’), however, only if the rise was followed by a fall, as in the dip pattern. If the same delayed accentual rise is concatenated with the subsequent accent by a high plateau, as in the case of the hat pattern, the astonishment ratings for the corresponding
stimuli remain at a constant, low level. Moreover, for each interval that separated the rise and the fall in the dip pattern, the stimuli sounded more astonished the earlier the fall was located in relation to the end of the accented word. If the fall started before the end of the accented word, the mean astonishment ratings exceeded the centre value 2.0 of the ‘astonished’ scale. The ratings increased up to 3.0 when almost the entire fall was completed before the end of the accented word.

Although the experimental methods for the creation and the identification of the late pitch accent leave room for improvement and hence for a clearer pattern of results (cf. 4.1), translating the above judgment pattern into pitch-accent identifications allows us to conclude that the late category can occur at prenuclear accent positions in German, and that it contrasts in this position with pitch-accent categories that show a vowel-internal peak alignment and are not related to astonishment, which applies, for example, to the medial peak. However, identifying late prenuclear accents requires not only a rising, but also a subsequent falling pitch movement. Hence, the late category is restricted prenuclearly to dip patterns, in which the pitch accent is separated by an indentation from the following nuclear one. Moreover, the earlier the fall begins relative to the end of the corresponding accented word, the more clearly the late category is perceived.

We assume that a stimulus in which the late category is clearly identified will be judged as predominantly ‘astonished’, i.e. it will yield a scale value which is higher than the centre value of 2.0. Based on this threshold, it may be concluded that a clear signalling of the German late pitch accent requires a falling movement which is partly realized before the final accented-word boundary. Such an within-word alignment of the prenuclear fall was given for all dip stimuli based on Müllerin; and accordingly all those stimuli were perceived with a late pitch accent. In contrast, in the case of Müll, in which the fall always began after the end of the accented word, none of the dip stimuli were able to clearly signal the late pitch-accent category.

In summary, the results confirm hypotheses (1)-(3). This confirmation is independent of possible biases that may have been introduced by the stimulus generation and judgment. However, the threshold-related conclusions that concern the perceptual boundary for the identification of late pitch accents on the one hand and the coordination of the rising and falling F0 movements on the other should be cross-validated by further experiments that use different stimulus continua and judgment tasks.
4.2 Prediction (1): ‘Prenuclear accents’

By confirming hypothesis (1), our findings are in line with the common prediction (1) of GToBI and the KIM that accent categories may be defined, on the basis of meaning, independently from accent positions (i.e. nuclear vs. prenuclear). Until now, evidence in favour of this prediction was restricted to the fact that sets of phonetically comparable F0 patterns may be found nuclearly and prenuclearly. The present study shows for the first time that phonetically comparable nuclear and prenuclear accent patterns convey the same meaning, which supports regarding them as instances of “the same thing”.

4.3 Prediction (2): ‘Phonological points of reference in the segmental string’

In combination with the previous findings of Niebuhr (2007a), the present study has shown that the coding of the late pitch accent involves two different timings between intonational and segmental events. Firstly, the late pitch accent requires an F0 rise which starts within the vowel of the accented syllable and reaches the F0 maximum after the syllable offset. Secondly, the following fall must be partly located within the accented word. GToBI and the KIM are both in accord with these findings to the extent that they regard the accented syllable (GToBI) or its vowel (KIM) as crucial segmental references in pitch-accent production and perception. However, neither model predicts that the larger unit of the accented word or its final boundary can play a role in the coding of pitch accents in German. That is, provided that the accentual F0 pattern, its timing with the accented syllable, and the distance to the subsequent pitch accent (in terms of the number of intermediate unaccented syllables) remain constant, it should not matter whether the accented word is Müll, Müller, or Müllerin. However, our data show the opposite pattern, contradicting prediction (2). With regard to monosyllabic words like Müll, our results even suggest in line with acoustic findings on nuclear late-peak alignment (cf. Ambrazaitis and Niebuhr 2006) that the presence of a word-internal fall gets priority over the extension of the fall beyond the accented syllable.
Both the KIM and GToBI have conceptualized the late pitch accent as a rising movement that begins in and goes across the accented vowel and syllable and should therefore be perceived equally at the prenuclear accent positions in the dip and the hat-pattern stimuli. However, this was not supported by our findings, in which the late category was only identified for prenuclear accents in dip-pattern stimuli. In this way, the findings indicate, contrary to prediction (3), that the late category should be represented phonologically by both rise plus fall. This conclusion is in line with perceptual data of Niebuhr (2007b). Additional support comes from Prieto and Ortega-Llebaria (2009). They found for Catalan and Spanish that the realization of a complex rising-falling (i.e. tritonal) pitch accent induces a lengthening of the accent syllable. In fact, Niebuhr (2007b) described a similar lengthening for the late pitch accent of German.

By showing that the German late category consists of a rising-falling and not just of a rising pitch pattern, the present findings also offer an explanation for the results of Peters et al. (2005), who found that sequences of late and early pitch accents are almost absent in dip patterns, whereas they are the most frequent pairing in hat patterns (cf. Fig. 3). This asymmetry may be ascribed to the fact that the pitch accents which preceded the early ones in the dip and the hat patterns, and which were homogeneously labeled as ‘late’, are actually instances of two different categories with different meanings. The pitch accent that preceded the early one in the dip pattern represents the astonishment-conveying late category. In contrast, the early pitch accent is used to express that something is well known or settled (cf. 1.1). These two meanings are contradictory, and hence the corresponding late and early pitch accents are incompatible if they are related to the same information. The initial pitch accent in the hat pattern that shows a comparable delayed rise across the accented syllable to the late pitch accent in the dip pattern is not a variant of the same late category and does not signal astonishment. On the contrary, the hat-pattern initial delayed rise was found to be indistinguishable on the astonishment scale from the hat-pattern initial rise that represents the medial pitch accent. The hat patterns that were labeled by Peters et al. (2005) as sequences of late-early pitch accents therefore become compatible from a semantic-pragmatic point of view.

The lack of a difference between the two hat-pattern initial accentual rises on the astonishment scale does not rule out the possibility that they still have different meanings. It is possible that the astonishment scale was not able to cover their meaning difference. Several intonational analyses of
Late pitch accents in hat and dip intonation patterns

German suggest that the delay of the hat-pattern initial accentual rise is linked with a meaningful, phonological change (cf. Féry 1993; Wunderlich 1998). The meaning differences linked with the delay of the hat-pattern initial rise are frequently related to the creation of a contrast between the two pitch-accented items that are spanned by the hat pattern (cf. also Peters et al. 2005). Niebuhr's (2005) meaning-oriented hat-pattern analyses on the basis of German spontaneous speech corpora support the notion of two hat-pattern initial rising pitch accent categories. It is likely that the delay of the hat-pattern initial accentual rise indicates that the corresponding information is not just opening, but that it will be of particular relevance in the further discourse (cf. Steube 2001 on ‘I-Topik’). In any case, it is reasonable to assume on this basis that there is a further late pitch accent in addition to the one that was addressed in the present study. This further late pitch accent is likely to consist of a single rise.

5 Outlook

The present study sought to determine whether form-meaning relationships that have been established at nuclear accent positions also hold prenuclearly, whether the accented vowel or syllable are really the only phonologically decisive points of reference for pitch accents in the segmental string, and whether the late pitch accent consists only of a rising movement. Since the results contradict the validity of both the second and third predictions, as well as revealing the possibility of two late pitch accents in German, the most obvious task of follow-up studies is to extend the three research questions to other pitch accents.

In addition, the above research questions should be extended to other languages. In particular, it is worth performing parallel studies for British and American English, since the relationships between phonetic substance, phonological form, and meaning in the sets of pitch-accent categories are very similar to those of German (cf. Pierrehumbert and Steele 1989; Redi 2003; Kleber 2006). Irrespective of the linguistic background of follow-up studies, the alignment of prenuclear accents cannot simply be determined or compared across languages (cf. Atterer and Ladd 2004) on the assumption that there is just a single pitch-accent category. In order to avoid “comparing apples and oranges” the semantic-pragmatic context must be carefully controlled.

Finally, the critical discussion of our results demonstrated that using stimulus meanings and functions as points of reference in determining intonational contrasts and their coding entails new methodological chal-
challenges as to the generation of the stimuli and, most importantly, the task of the subjects. We have shown that scalar functional judgments are a feasible option. However, it is unclear whether this option is superior to traditional 2AFC tasks, whether other scalar methods can be applied more easily and reliably by subjects, and whether the applicability changes with different types of intonational categories. In any event, modern intonational studies need to refer to intonational functions and meanings, as well as making stronger efforts to pin-point these functions and meanings.

6 References

Late pitch accents in hat and dip intonation patterns


