Original Paper

Phonetica 2007;64:174–193
DOI: 10.1159/000107915

Received: March 20, 2007
Accepted: July 12, 2007

The Signalling of German Rising-Falling Intonation Categories – The Interplay of Synchronization, Shape, and Height

Oliver Niebuhr

Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Kiel, Germany

Abstract

Based on the phonology of the Kiel Intonation Model (KIM), a tripartite opposition of German intonation is investigated: early, medial, and late peaks. These intonation categories, which can be projected onto H + L*, H*, and L* + H in the AM framework, are described in the KIM as rising-falling F_0 peak patterns differentiated by their synchronization with the accented vowel onset. Perception experiments were carried out, showing that the function-based identification of the peak categories is not only influenced by peak synchronization, but also by peak shape and height. While the complete spectrum of findings is not covered by the current phonological modelling, the findings corroborate the existence of all three categories in German intonation and support the idea that the timing of the peak movements with regard to the accented vowel is important for their perceptual differentiation.

1. Introduction

1.1 Aim of the Study

It is common understanding that in intonation languages changes in pitch convey communicative meanings, e.g. the speaker's attitudes towards the listener or the information of the utterance [Bolinger, 1972; Ladd, 1996; Kohler, 2005]. Furthermore, it is assumed that languages use a repertoire of structurally complete units to code and carry these meanings. However, apart from this trivial statement that there are different meaningful intonation units within a language, we still know little about their specific meanings and phonological forms, i.e. about what is signalled and how it is done. With regard to these questions, rising-falling F_0 peak patterns which are linked to accented syllables belong to the most investigated intonation phenomena across languages.

Since the analysis of the Swedish word accents by Bruce [1977], rising-falling F_0 peak patterns have been – directly or indirectly – described and systematized with regard to their alignment, i.e. the relative synchronizations of local F_0 landmarks like

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/07/0643–0174
$23.50/0
Accessible online at:
www.karger.com/pho

Oliver Niebuhr
Freyastrasse 10
DE–24939 Flensburg (Germany)
Tel. +49 178 553 49 29, E-Mail on@ipds.uni-kiel.de

the peak maximum with local landmarks that are represented by spectral changes in the acoustic speech signal, such as segment or syllable boundaries. However, there are many observations and indications across languages that the signalling of meaningful peak patterns involves more than just the temporal relationships of single landmarks from the phonatory and articulatory string. Holistic peak characteristics also seem to play a role. This issue is addressed in the present study for (Northern Standard) German. Following a line of perception research begun by Niebuhr [2003a, b], a comprehensive experimental investigation is performed. It considers three independent experimental variables, which are frequently confounded in perception research [Andreeva and Barry, 1999; Redi, 2003; Chen, 2003]: F_0 peak synchronization, peak shape, and peak height. In this study, peak shape refers to the durations of the rising and falling movements, while peak height concerns their frequency range.

The three experimental variables are integrated into a uniform methodological framework. The substantial basis of this framework is provided by three phonologically different F_0 peak patterns of German: the early, medial, and late peaks. They were selected for two reasons. First, compared with others, the communicative meanings of these peak categories are relatively well understood [Kohler, 1991c; Dombrowski, 2003; Kohler, 2005] and can thus be used as reliable points of reference in their perceptual identification. Second, other languages like British and American English, Hungarian, Neapolitan Italian, and Swedish have intonational oppositions that are phonetically comparable, or at least quite similar, to early versus medial and/or medial versus late in German [Kleber, 2006; Pierrehumbert and Steele, 1989; Redi, 2003; Ambrazaitis and Bruce, 2006; Gósy and Terken, 1994; D'Imperio and House, 1997]. Hence, the findings of the present study may be of cross-linguistic interest.

1.2 Early, Medial, and Late Peaks

The alignment concept sketched in 1.1 can also be found in the KIM for German [Kohler, 1991a, b], which provides the phonological basis for the present investigation. The intonation categories of the KIM are inspired by the analyses of British English intonation by Halliday [1967] and O'Connor and Arnold [1970]. However, Kohler conceptualizes the categories as arising from uniform F_0 contour types being in different synchronizations with the vowel of the accented syllable.

Following Kohler [1991a, b], accent is a concrete phonetic property of the syllable. It means that the hearer experiences the corresponding syllable as standing out from the surrounding ones, i.e. it is marked by a higher perceptual prominence. Kohler further distinguishes accent from lexical (or word) stress. In languages like German and English, the latter refers to an abstract property of each word which specifies the syllable that may be accented by the speaker in the production of utterances. In this aspect, these languages differ from, for example, French, which lacks lexical stress and allows to select the accented syllable more freely. Moreover, by subdividing the continuum of perceptual prominence, the KIM postulates an accent phonology, which is basically independent from the intonation phonology and its categories. The coexistence of the two phonological systems has two important implications, which are explained in more detail in Niebuhr [2006]. First, intonation categories entail the presence of accented syllables, but not vice versa. Second, accents and the accent patterns of utterances form a separate prosodic meaning layer in addition to the intonational one. Accent-based meaning types comprise, for example, emphasis [Kohler and Niebuhr, 2007] and focus [Xu et al., 2004; Baumann and Hadelich, 2003]. The present study

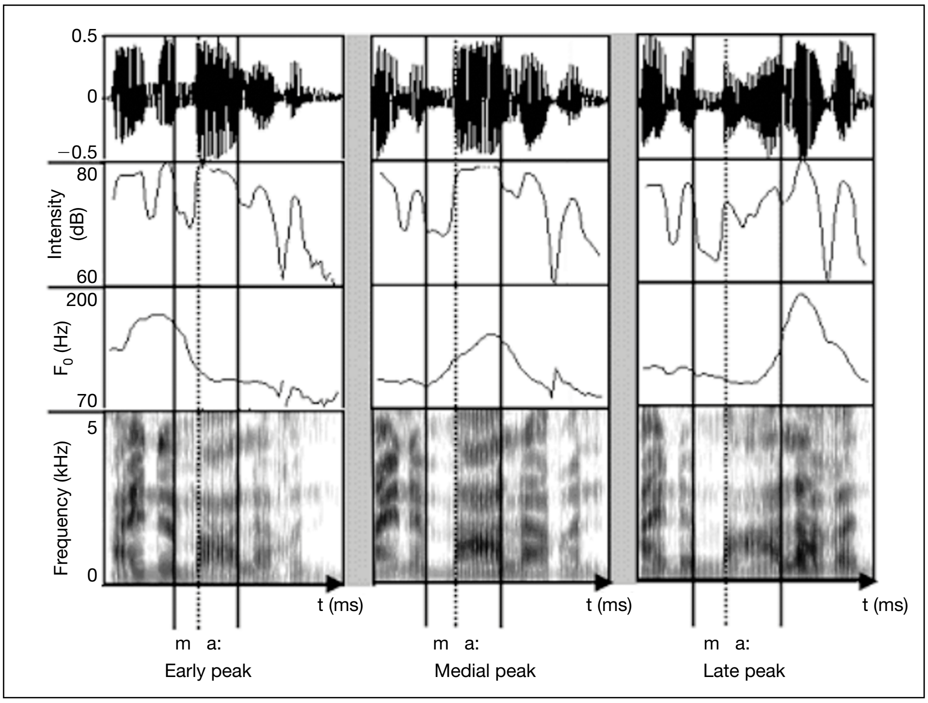

Fig. 1. Natural productions of the early, medial, and late peak category of the KIM in the utterance *Eine Malerin* ('a painter') with the accented syllable *Ma-*. The accented syllables and the boundaries between nasal [m] and accented vowel [a:] are marked by continuous or dotted lines, respectively. The four windows show (from the top) an oscillogram, intensity course, F_0 course, and a spectrogram.

addresses and systematically manipulates *intonational* categories against a (phonologically) *constant accentual* background.

The three rising-falling F_0 peak contours that are differentiated in the intonation phonology of the KIM and addressed in the present study are called early, medial, and late peaks. This terminology reflects the synchronization of the peak contour – represented by the F_0 maximum – relative to the boundaries of the accented vowel. As for the early and the medial peak, the accented vowel onset is defined as the boundary between them. That is, F_0 peak contours having their maximum before (or at) the accented vowel onset belong to the early peak category, whereas F_0 peak contours with a maximum clearly after the accented vowel onset constitute the medial peak category. In the case of unaccented syllables following the accented one (in the stimulus utterances of all the perception experiments presented here), late peaks are characterized by the F_0 peak maximum being aligned after the accented vowel offset, typically around the onset of the vowel of the following unaccented syllable. Figure 1 shows typical examples of F_0 peak patterns for the three KIM categories, produced naturally by the author in the German utterance *Eine Malerin* ('a painter') with an accent on the syllable *Ma-* ([ma:]).

The acoustic definitions of the three peak categories are founded on empirical evidence, mainly provided by the studies of Kohler [1987, 1991c]. In these studies, F_0

peak shift continua were judged by subjects with regard to changes in the perceptual properties and meanings of the melodic contours. As regards the latter, early, medial, and late peaks convey attitudinal meanings, which can be outlined by the terms *given (unchangeable)*, *new*, and *unexpected*.

It has to be emphasized that these terms are not used in the sense of information structure or packaging [Halliday, 1967; Chafe, 1976; Baumann and Hadelich, 2003], but in the (more general) sense of argumentation structure. That is, the peak categories indicate how the conveyed information is to be judged in the conversation context. So, for example, a guard making announcements in a delayed train uses early peaks not to indicate that the changed arrival and departure times have been mentioned before or are accessible in other ways, but to assure the passengers that these times are not going to change again. Likewise, medial peaks can be used to indicate that the corresponding information adds a new aspect to the argumentation, even if the information itself has been mentioned before. However, it has to be considered that information marked as given/new in the sense of argumentation structure can simultaneously be given/new in the sense of information structure. Furthermore, it is pointed out that *given (unchangeable)*, *new*, and *unexpected* only represent generic meaning characteristics of the three peak categories. The actual concrete interpretations can show additional features depending on the other (e.g. lexical and grammatical) meaning components involved in the utterance [Barry, 1981]. Correspondingly, by applying semantic differentials to F_0 peak shift continua created in different stimulus utterances, Dombrowski [2003] and Kohler [2005] have shown that changes in the peak category coincide with meaning changes on various scales.

1.3 Observations and Indications of Previous Studies Concerning Peak Shape and Height

The impetus to investigate the role of peak shape and height in the German peak categories came from observations made in Niebuhr [2005]. In this study, the author explored the possible concatenations of early, medial, and late peaks, based on systematic resyntheses of sequences of two pairs of peak categories, followed by their perceptual evaluation with regard to naturalness and peak category identification. In this, it turned out that F_0 peaks showing a maximum *before* the accented vowel onset could already be perceived as medial if the rise to the maximum was fast enough. On the other hand, shortening the duration of the falling movement was an effective means to create early peak perceptions, even if the peak maximum was located *after* the accented vowel onset. Moreover, it was observed that the adequate position and duration of the indentation between concatenated peaks depended on their phonological categories. So, for example, in a sequence of two early peaks, the F_0 minimum of the indentation had to be reached shortly after the first peak maximum around the end of the accented vowel, immediately followed by a gradual rise to the second peak maximum. A sequence of two medial peaks sounded most natural if the F_0 minimum of the indentation followed a gradual fall and was reached around the beginning of the second accented syllable. (A rise onset located around the beginning of the accented syllable is also in line with acoustic findings for medial peak productions in German [Gartenberg and Panzlaff-Reuter, 1991; Niebuhr and Ambrazaitis, 2006] as well as for similar peak categories in other languages [Ladd, 2003]. Finally, early-medial sequences were perceived to be most naturally concatenated by a low plateau between the end of the first accented vowel and the beginning of the second accented syllable.

As in the study of Niebuhr [2005], many observations concerning peak shape or height are made by chance or on an informal basis and thus often appear as marginal notes in the literature. Gósy and Terken [1994] performed the only known experimental study which systematically investigated the contribution of all three peak variables (synchronization, shape, and height) to the perception of F_0 peak categories signalling question versus statement in Hungarian. As regards the synchronization, Gósy and Terken created a peak shift continuum spanning both sides of the relevant vowel. The resulting stimuli showed that F_0 peaks having their maximum before the vowel onset signalled statements, whereas F_0 peaks with a maximum in the vowel signalled questions. Additionally, for F_0 peaks showing a postvocalic maximum, the question judgements decreased again. On this basis, Gósy and Terken found that a fast-rising/slow-falling peak supported the perception of questions as against slow-rising/fast-falling and fast rising-falling peaks. The latter also yielded more question responses than the slow-rising/fast-falling peak. Finally, the number of question judgements was increased by an extended peak height.

D'Imperio and House [1997] found that statements versus questions in Neapolitan Italian are, like in Hungarian, signalled by F_0 peaks showing maxima before or after the vowel onset, respectively. Moreover, D'Imperio and House found that the stimuli based on an original question utterance received more question judgements than the stimuli based on an original statement utterance. They consider that this may at least partially be due to an unintended peak shape difference. The series based on questions showed a slow rising-falling shape, whereas a fast-rising/slow-falling peak shape was used in the series based on statements.

Finally, Kohler [1991c] obtained different perceptual transitions from early to medial across two comparable F_0 peak shift continua. This deviation may be due to a clear but unintended difference in peak height between the two continua, indicating that medial peaks are supported by increased height.

1.4 Hypotheses

The results for the synchronization variable in the studies of Gósy and Terken [1994] as well as of D'Imperio and House [1997] reveal perceptual organizations of the F_0 peak shift continua that are comparable to the one found for German early, medial, and late peaks (see section 1.2). That is, a first abrupt perceptual change takes place if the maximum of the shifted peak contour enters the vowel. A second (and less dynamic) change follows if the F_0 peak contour is shifted with its maximum beyond the vowel offset. Assuming that this parallel indicates comparably signalled peak categories (which only show language-specific meanings) [Niebuhr and Kohler, 2004; Kohler, 2005], the findings for peak shape and height in Hungarian and Neapolitan Italian can be projected onto the German early, medial, and late peak categories. If this is complemented by the observations made directly for the German peak categories, a picture emerges from which the following hypotheses were derived:

(1) Peak shape influences the perception of German early, medial, and late peaks. In the contrast between early and medial peak,
  (1a) a fast-rising/slow-falling peak shape supports the perception of medial peaks compared with a slow-rising/fast-falling and a symmetrically fast rising-falling peak, and
  (1b) a symmetrically slow rising-falling peak supports the perception of medial peaks as against the fast-rising/slow-falling peak.

In the contrast between medial and late peak,
(1c) a slow-rising/fast-falling peak shape supports the perception of late peaks compared with a symmetrically fast rising-falling peak.
(2) Peak height influences the perception of German early, medial, and late peaks. Increasing the peak height supports the perception of the medial peak category.

2. Method

2.1 General Experimental Setup

The whole experimental series comprised four perception experiments. Two of them (the CV group) were based on peak shift continua spanning the CV boundary of the accented syllable, aiming at the contrast between the early and the medial peak. The remaining two experiments (the VC group) aimed at the contrast between the medial and the late peak and were thus marked by peak shift continua reaching from the center of the accented vowel to the end of the (voiced) consonant of the following unaccented syllable. In all experiments, the F_0 peak was shifted in equal-sized steps of 20 ms. Each of the experiments comprised either two or four stimulus series resulting from the different shape and height conditions integrated into the F_0 peak shift continua.

The perception experiments were uniformly performed with a function-based indirect (two-alternative forced-choice) identification task, which was also used in the studies of Kohler [1987, 1991c], Niebuhr [2003a, b], Niebuhr and Kohler [2004], and Kleber [2006], among others. It represents a further development of the task introduced by Nash and Mulac [1980]. The basic idea behind this experimental paradigm is that a change in the intonation category of the stimulus utterance also changes the meaning of this utterance (see section 1.2), and that this meaning change affects the possibility of combining the stimulus utterance with a constant (preceding) context utterance, produced by the same speaker. For this reason, the hearers are asked to judge whether context and stimulus utterances do or do not match. In this way, the matching judgements given by the hearer reflect the intonation category identified in the stimulus utterance. It has to be underlined in this connection that within the scope of this paradigm, meaning represents a point of reference for investigating the perceptual signalling of early, medial, and late peaks by means of synchronization, shape, and height. Using the indirect identification task was necessary, since the meanings of the intonational categories are not as straightforward as, for example, the ones in Hungarian and Neapolitan Italian (i.e., question vs. statement; see section 1.3). Therefore, they cannot directly be asked for.

2.2 Experiments 1 and 2: Synchronization with Peak Shape Variation

The stimulus series of experiment 1, i.e. the CV series aiming at the contrast between early and medial peak, were based on the utterance *Sie war mal Malerin* ('she was once a painter'). It was produced by a male speaker with *Ma-* ([ma:]) of *Malerin* ('painter') as the only accented syllable.

In the first step of the F_0 manipulation, the original F_0 contour was replaced by a contour stylized at five contour points. That is, based on a linear hertz scale, the contour points were connected by linear interpolation. As shown in figure 2, points 1 and 5 represent the F_0 values at the beginning and the end of (modal voicing in) the utterance. The final contour point 5 was set at a terminal F_0 level. The initial contour point 1 was given a higher F_0 value. By this, a moderate F_0 declination was created, which is frequently found in natural speech productions [Cohen and 't Hart, 1967; Ohala, 2004]. Within this declination, the remaining three contour points 2, 3, and 4 constituted the F_0 peak, represented by rise onset, maximum, and fall offset. All contour points reflected naturally produced F_0 values of the speaker.

As for the F_0 peak itself, 4 different peak shapes were created, 2 (approximately) symmetrical and 2 clearly asymmetrical ones, without changing the frequency values of the three contour points. The 4 peak shapes resulted from two durations, fast and slow, cross-combined with the rising and falling F_0 movements of the peak. Correspondingly, the peak shapes can be characterized as fast/fast (f/f), fast/slow (f/s), slow/slow (s/s), and slow/fast (s/f). The durations differed by factor two. According to the perception experiments of 't Hart et al. [1990] and Nabelek and Hirsh [1969], this was regarded sufficient for perceiving the differences between the fast and slow rises and falls and

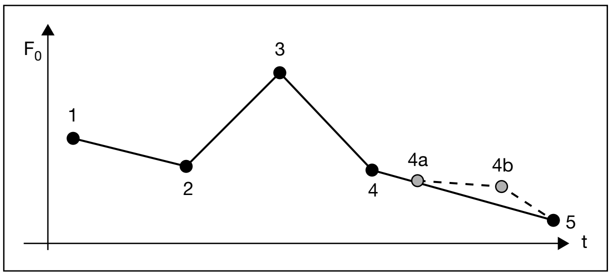

Fig. 2. Basic F_0 contour of the stimuli in the perception experiments 1–4. In experiment 1, the contour stylized with contour points 1–5 was used. In experiments 2–4, two further contour points, 4a and 4b, were introduced to delay the terminal fall.

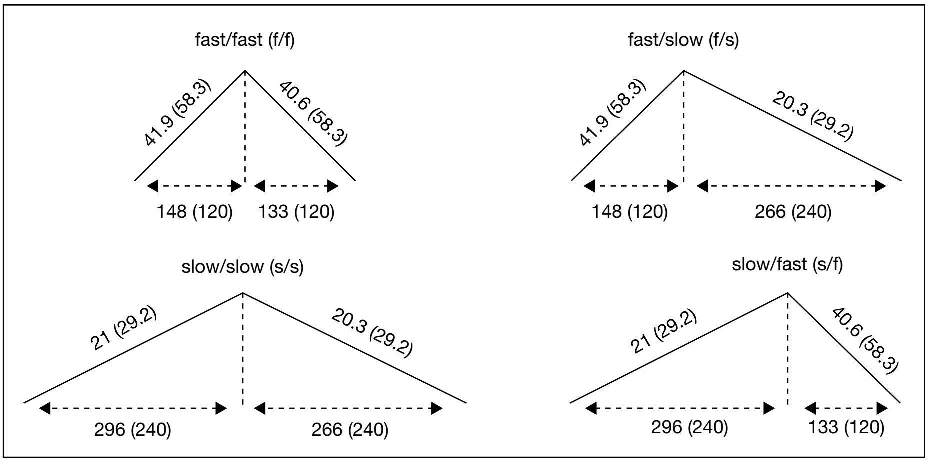

Fig. 3. The four different peak shapes used in the CV and VC series of experiments 1 and 2. Durations (in ms) and slopes (in semitones/s) of the rising and falling movements are shown. The values of the VC series of experiment 2 are given in parentheses.

hence the peak shape differences. Informal listening confirmed this assumption. Furthermore, the durations of the fast F_0 movements were chosen with regard to the physiological limits of pitch change. That is, the slopes resulting from the durations of the rising and falling movements (in connection with the constant peak height) did not exceed the limits estimated by Xu and Sun [2002] for similar F_0 ranges. The 4 peak shapes as well as the corresponding durations and slopes of the rising and falling F_0 movements are shown in figure 3. Additionally, the frequency values of the 5 contour points and the calculated peak height are given table 1.

Each of the 4 peak shapes was applied to an F_0 peak shift continuum, reaching in 11 steps from the onset of [m] to the center of the [a:] in the accented syllable *Ma-*. After each shift, a stimulus was resynthesized. The resyntheses as well as the preceding F_0 manipulations were done in *praat* [Boersma and Weenik, *www.praat.org*]. In this way, 4 stimulus series were created, one for each peak shape. Each series consisted of 11 stimuli, and in all 4 series, the F_0 peak maximum in stimulus 6 coincided with the accented vowel onset.

It should be noted that, while the slopes within the shifted peak remained constant, the peak shift strategy changed the declination slopes before and after the peak (i.e. P1–P2 and P4–P5; fig. 2). An alternative strategy would have been to shift the peak *on* the declination line so that a shift to the left/right increases/decreases the F_0 values of P2–P4 by a constant factor. However, such a strategy is not supported by acoustic analyses of the peak categories in German [Gartenberg and Panzlaff-Reuter,

Table 1. F_0 values (in Hz) for the 5 or 7 contour points P1–P5 in the stylized F_0 contours of the stimuli of experiments 1–4 and the heights of the rises and falls within the peak (in semitones, st)

	P1	P2	P3	P4	P4a	P4b	P5	Height rise, st	Height fall, st
Exp. 1	112	105	150	110			90	6.2	5.4
Exp. 2	100	89	134	89	78	76	66	7	7
Exp. 3	98	95	125 (108)	95	80	82	74	5 (2.5)	5 (2.5)
Exp. 4	98	89	134 (111)	89	80	82	74	7 (3.5)	7 (3.5)

Values in parentheses are used in the low peak condition.

1991] and may thus affect the identification of the peak categories. On the other hand, since there is no known evidence that a change in the declination slope affects the identification of peak categories, this peak shift strategy was preferred.

An analogous procedure was used to create the VC series of experiment 2, which were supposed to cover the contrast between medial and late peak. That is, 4 stimulus series were resynthesized using the peak shapes f/f, f/s, s/s, and s/f. There were only small deviations concerning the F_0 values of the contour points (table 1) and the durations of the fast and slow movements within the F_0 peak (fig. 3). These deviations were partly due to the empirical findings for the peak categories involved. So, for example, the peak height was slightly increased with regard to the late peak category (see section 1.4 and fig. 1).

However, as against the CV series of experiment 1, a few important modifications have to be mentioned, which were also responsible for the remaining deviations in F_0 and duration values. First, it turned out that the utterance *Sie war mal Malerin* was not suitable for creating the necessary peak shifts. The two unaccented syllables following the accented one were too short for the slow-falling peak shapes to be fully realized before the end of voicing. Therefore, the utterance was modified to *Sie's mal Malerin gewesen*. That is, the tense was changed from past tense to perfect. In this way, the (past) participle *gewesen* was added at the end of the utterance, while the relevant sequence of the accented syllable *Ma-* and the adjacent syllables *mal* ('once') and *-le-* was kept constant.

In connection with *gewesen*, a further modification was necessary. The F_0 declination between contour point 4 and 5 became too long and sounded artificial. This problem was solved by introducing two further contour points (4a, b) in between 4 and 5 (fig. 2). They were placed at an approximately equal F_0 level in the vowels of *ge-* and *-we-* in *gewesen*. This additional F_0 plateau delayed the terminal F_0 fall until the last two syllables, as in the stimulus utterance *Sie war mal Malerin*. However, the plateau resulted in the perception of another accent on *-we-*. In KIM, this accent is phonologically analyzed as a partial deaccentuation with an early peak (in GToBI, it would probably be analyzed as a low phrase accent [Grice and Baumann, 2000]).

Finally, the steps of the F_0 peak shift continuum and hence of the number of stimuli were reduced to 6. Starting from an F_0 peak with the maximum at the accented vowel offset, 5 peak shifts were performed, 2 towards the center of the vowel and 3 into the following unaccented vowel. In this connection, figure 4 illustrates that the slow-rising peaks always started before the accented vowel onset, whereas the opposite holds for most of the stimuli with the fast-rising peak shapes.

2.3 Experiments 3 and 4: Synchronization with Peak Height Variation

The stimulus series created for experiments 3 and 4 were part of a larger experimental framework, also addressing effects of the duration and intensity relations between the accented syllable and its adjacent syllables on the identification of the German peak categories. In what follows, however, only those aspects of the stimulus generation relevant to the peak height variable are dealt with. Further details about the stimuli and the method are described in Niebuhr [2006].

The generation of the CV and VC stimulus series for experiments 3 and 4 was done by analogy to experiment 2. So, for example, the (same) utterance *Sie's mal Malerin gewesen*, with *Ma-* as the relevant accented syllable, served as the basis for the stimulus series, and the symmetrical fast rising-falling shape f/f of experiment 2 was used in all stimulus series. As can be seen in figure 3, this

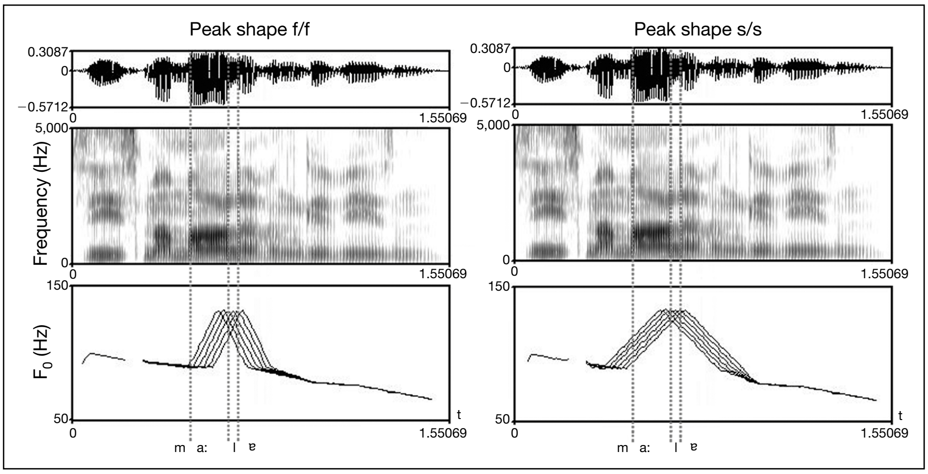

Fig. 4. The peak shapes f/f and s/s in the peak shift continuum used in the VC series of experiment 2. In contrast to the fast-rising f/f peak, the onset of the slow-rising s/s peak never enters the accented vowel onset [a:] in *Malerin* ('painter'). The same holds for f/s versus s/f.

corresponds to peak movements of 120 ms. Table 1 provides the F_0 values of the 7 contour points. The peak shift continuum for the VC series of experiment 4 is identical to the one of experiment 2 (fig. 4). In the CV continuum of experiment 3, however, the number of peak shifts and hence the number of stimuli was reduced from 11 to 7, due to the modified stimulus conditions and in order to reduce the length of the experimental session. Starting from an F_0 peak synchronized with the accented vowel onset, 6 peak shifts were performed, 3 into the accented vowel and 3 into the preceding nasal.

Experiments 3 and 4 comprised two stimulus series each, resulting from a high peak condition (>h) and a low peak condition (<h). They differed by a factor of two, starting from high peak conditions of 5 semitones in the CV series of experiment 3 (aiming at early vs. medial) and of 7 semitones in the VC series of experiment 4 (aiming at medial vs. late). Hence, in the low peak conditions, the peak heights amount to 2.5 or 3.5 semitones, respectively (see table 1 for the absolute frequency values). With regard to the findings of 't Hart [1981], the peak height difference of around 3 semitones was assumed to be perceptually detectable. Again, this was confirmed by informal listening before the actual experiments were performed.

2.4 Construction of the Perception Experiments

In total, 12 stimulus series were created; 6 belong to the CV group, comprising 4 series concerning peak shape and 2 series concerning peak height (i.e. experiments 1 and 3). The remaining 6 series belong to the VC group, again with 4 series addressing peak shape and another 2 series addressing peak height variations (i.e. experiments 2 and 4). For each of the 12 stimulus series, a separate identification test was constructed. In this, the stimuli were preceded by a context precursor, being constant for each test.

In the CV group, the context utterance *Jetzt versteh' ich das erst* ('Now, I understand') was used, realized with a single medial peak accent on *-steh* of *versteh* ('understand'). The lexical and intonational meanings of this context utterance signal that further information is going to follow and that this information will be characterized as new. This is not compatible with an early peak on the stimulus utterance, which indicates that the corresponding information is judged as generally known or at least as generally expected (see section 1.2). Context and stimulus utterance are only compatible if the latter contains a medial peak. Therefore, it was expected that subjects would judge the context stimulus pairs as matching if they perceive a medial peak in the stimuli. Otherwise, they judge the pairs as not matching. Hence, across the peak shift continuum from left to right, the judgements should change from 'not matching' to 'matching'.

In the VC group, the stimuli concerning peak shape and peak height were preceded by different contexts. In the peak shape series, the context utterance was *Ganz bestimmt* ('definitely'), produced with a medial peak on *-stimmt*. The context signals that the hearer is to acknowledge the following information as new and reliable. The latter rules out that the information is characterized as unexpected by the speaker, as is the case if the stimulus utterance is combined with a late peak (see section 1.2). Therefore, context stimulus pairs only match if a medial peak is identified in the stimuli. In the peak height series, the utterance *Ist doch logisch* ('but that's logical') served as context preceding the stimuli. It was produced with a medial peak on the only accented syllable *lo-* of *logisch* ('logical'). This context conveys a similar meaning as *Ganz bestimmt*. However, according to discussions on the context *Ganz bestimmt*, it should be brought out more clearly. Nevertheless, it is expected for both contexts that a perceptual change from medial to late is accompanied by a change from 'matching' to 'not matching' across the peak shift continuum from left to right.

In all series, context and stimulus were separated by a short pause of 100 ms. Considering the findings of Koiso et al. [1998] and Peters [2006], this was regarded appropriate to perceive a break between the utterances. On the other hand, this pause should be short enough for the utterances to be interpreted as elements forming a single holistic unit, which is decisive for the judgement task.

Each identification test was made up by the context stimulus pairs repeated several times and arranged in a randomized order. For the 4 tests of experiment 1, the 11 pairs were repeated 10 times, yielding 110 pairs to be judged in each test. Since it was observed that the concentration of the subjects decreased in the course of the tests, the remaining 8 identification tests of experiments 2–4 only comprised 7 repetitions of the context stimulus pairs, yielding 49 or 42 pairs, respectively. By this, the durations of the single tests were reduced from 16 min to about 7 min. Between the context stimulus pairs, a pause of 3 s was inserted, in which the subjects had to give their judgements. Each pair was introduced by a short beep; blocks of 10 stimuli were separated by 2 beeps.

2.5 Experimental Sessions

The identification tests of the experiments 1–4 were judged by different groups, either comprising 28 subjects (experiment 1) or 20 subjects (experiments 2–4). According to their own information, all subjects were native speakers of German and showed no hearing disorders. Due to technical restrictions, the experimental groups were split up into smaller groups of at most 8 subjects. The identification tests of experiments 1 and 2 were permuted across these subgroups to reduce order effects. Since the identification tests of experiments 3 and 4 were part of a larger experimental framework (see section 2.3), the tests were not permuted. The identification tests of experiments 3 and 4 concerning the low peak condition $<$h were always the final ones, i.e. the tests concerning the high peak condition $>$h were judged first. This strict order was intended, since it allowed the subjects to apply the form-meaning relationships they acquired and consolidated during the preceding tests to the low peak series, which were marked by different perceptual qualities than the other series.

Subjects were told that they were going to participate in a perception experiment dealing with German intonation. The attitudinal meanings of the German peak categories were explained to them by using examples. Further, they were told that they would hear pairs of utterances. In these, the first utterance would remain constant, whereas the second changed. The task would be to judge the utterance pairs as either 'matching' or 'not matching' with regard to their meanings. Subjects received the instructions in oral and written form. They gave their judgements by pressing buttons. The judgements, together with the corresponding reaction times, were collected and stored in a central measuring device.

3. Results

Figure 5a–d summarizes the results received for the 12 identification tests of experiments 1–4. In figure 5a and c on the left-hand side, the identification courses of the 6 CV series of experiments 1 and 3 are given, concerning peak shape (fig. 5a) or height (fig. 5c), respectively. For the ascending stimulus numbers, representing the chronological F_0 peak shift from left to right (i.e., into the accented vowel), the identification courses show in how many of the 280 (28 subjects $\times$ 10 repetitions; fig. 5a) or

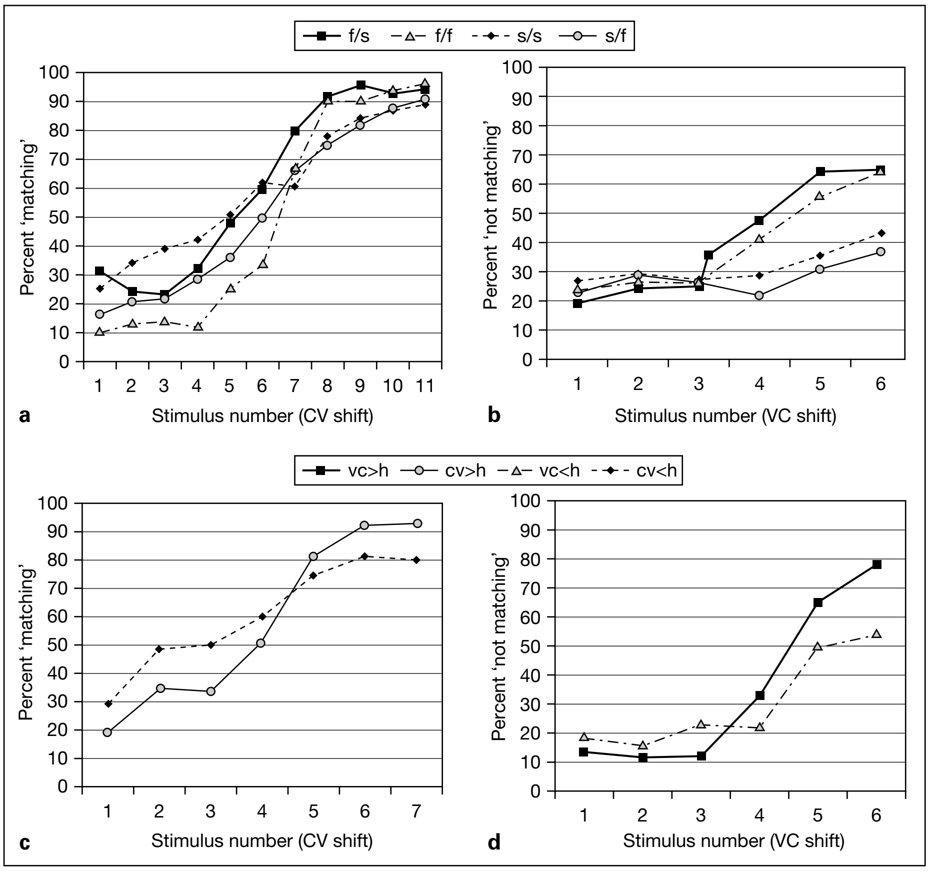

Fig. 5. **a, c** Percentages of 'matching' judgements for the context stimulus pairs in experiments 1 and 3 of the CVgroup, concerning either peak shape (**a**) or peak height conditions (**c**). **b, d** Percentages of 'not matching' judgements for the context stimulus pairs of the corresponding experiments 2 and 4 of the VC group. Judgements are arranged by ascending stimulus numbers, reflecting peak shifts from left to right, i.e. into or out of the accented vowel. Each data point represents 280 (**a**) or 140 (**b–d**) judgements.

140 (20 subjects $\times$ 7 repetitions; fig. 5c) cases each stimulus was judged as 'matching' with the preceding context utterance. Values are given in percentages.

Correspondingly, figure 5b and d on the right-hand side shows the identification courses for the 6 VC series of experiments 2 and 4. For these series, the ascending stimulus numbers represent a peak shift beyond the accented vowel offset. Each of the percentages comprises 140 judgements. However, unlike in figure 5a and c, the percentages of figure 5b and d refer to 'not matching' judgements. In this way, all identification courses ascend, which facilitates comparisons between the results for the CV and VC series.

Additionally, a statistical analysis was carried out. Each stimulus series was divided into two subsamples. The divisions were done with regard to the acoustic definitions of the three peak categories in the KIM (see section 1.2). That is, for the CV peak shift series, all stimuli with peak positions before or at the accented vowel onset

Table 2. Values t and R of the test statistics of either t tests or Wilcoxon matched-pairs signed-rank tests, as well as the corresponding degrees of freedom (*df*) and the levels of α errors (p) for significant outcomes

Exp. var.	Peak shape CV series (fig. 5a; n = 28)				Peak shape VC series (fig. 5b; n = 20)			
	☒ stim.	t ∣ R	*d.f.*	p	☒ stim.	t ∣ R	*d.f.*	P
f/f vs. f/s	1–6	−4.586 ∣ −	27	0.001	1–3	0.661 ∣ −	19	n.s.
	7–11	− ∣ 114	25	n.s.	4–6	−1.322 ∣ −	19	n.s.
s/s vs. s/f	1–6	2.522 ∣ −	27	0.05	1–3	0.987 ∣ −	19	n.s.
	7–11	− ∣ 148	25	n.s.	4–6	− ∣ 43.5	16	n.s.
f/f vs. s/f	1–6	2.575 ∣ −	27	0.05	1–3	0.043 ∣ −	19	n.s.
	7–11	−4.038 ∣ −	27	0.001	4–6	−3.955 ∣ −	19	0.001
s/s vs. f/s	1–6	− ∣ 117	26	n.s.	1–3	− ∣ 33.5	16	n.s.
	7–11	− ∣ 55	27	0.01	4–6	−3.366 ∣ −	19	0.01
s/f vs. f/s	1–6	−1.710 ∣ −	27	n.s.	1–3	− ∣ 36	15	n.s.
	7–11	−5.025 ∣ −	27	0.001	4–6	−4.642 ∣ −	19	0.001
f/f vs. s/s	1–6	6.189 ∣ −	27	0.001	1–3	0.507 ∣ −	19	n.s.
	7–11	−3.652 ∣ −	27	0.01	4–6	−2.601 ∣ −	19	0.05
	Peak height CV series (fig. 5c; n = 20)				Peak height VC series (fig. 5d; n = 20)			
>h vs. <h	1–4	− ∣ 45	20	0.05	1–3	1.467 ∣ −	19	n.s.
	5–7	−2.296 ∣ −	19	0.05	4–6	−3.192 ∣ −	19	0.01

The 28 tests compare the (distributions of) ‘matching’ judgements for the experimental variables (exp. var.), summed up for each subject (i.e., n = 28; 20) and the stimulus numbers given in ☒ stim., integrating peak positions to both sides of the vowel.

constituted one sample and the remaining stimuli with peak positions after the accented vowel onset constituted another sample. Correspondingly, for the VC series, all stimuli with peak positions before or at the accented vowel offset were allocated to one sample, whereas the stimuli with postvocalic F_0 peaks represented another sample. The divisions are indicated by the dotted lines in figure 5a–d. The samples comprising the same range of peak positions were then compared across the different peak shape or height conditions by means of the number of ‘matching’ judgements summed up for each subject. In this way, influences of peak shape and height could be analyzed separately within the peak synchronization ranges known to be the perceptual domains of early, medial, and late peaks.

Prior to the statistical analysis, the method developed by David et al. [1954] was used to reveal whether the sums of ‘matching’ judgements received for the 28 or 20 subjects could (approximately) be regarded as normally distributed. If this was the case, parametrical t tests for paired samples were used to compare the sums of ‘matching’ judgements between the corresponding conditions. Otherwise, the nonparametrical counterpart of the t test was used, i.e. the Wilcoxon matched-pairs signed-rank test. In total, the statistical analysis comprised 28 tests. Results are summarized in table 2.

As can be seen in figure 5a-d, almost all peak shifts led to a change in the majority of judgements from ‘not matching’ to ‘matching’ (in the CV series) or from ‘matching’ to ‘not matching’ (in the VC series), respectively. However, there are a few exceptions in the VC series. For the slow-rising shapes s/s and s/f (fig. 5b), no change in the majority of judgements occurred across the peak shift continuum. Furthermore, the identification in the low peak condition vc<h crosses the 50% boundary only marginally with a clearly asymptotic course (fig. 5d), indicating a saturation in the increase of 'not matching' judgements for further peak shifts. These findings led to significant differences in the number of 'matching' judgements compared with the fast-rising shape or the high peak conditions, respectively (table 2).

The further influences of peak shape and height variations can be summarized as follows. As for the CV series, shape effects concentrate on the F_0 movement which dominates in the accented vowel. That is, for F_0 peaks which reach the maximum before the vowel onset and thus fall into the vowel (stimuli 1–6; fig. 5a), the 'matching' judgements were primarily influenced by the falling slope. Stimuli with slow-falling F_0 movements were more frequently judged as 'matching' than the ones with fast-falling F_0 movements (table 2). On the other hand, for the F_0 peak rising into the vowel (i.e. stimuli 7–11; fig. 5a), the frequency of 'matching' judgements was predominantly influenced by the rising movements. In this connection, a faster rise significantly increased the number of 'matching' judgements as against a slower rise (table 2).

The results for the peak height variable also revealed significant effects. These effects were different for F_0 peaks synchronized before and after the accented vowel onset. Compared with the high peaks of the cv>h condition, the number of 'matching' judgements in the low peak condition cv<h (fig. 5c) significantly increased before and significantly decreased after the accented vowel onset.

Finally, since identical peak shift continua were used in all VC series, it has to be pointed out that the final stimulus 6 in the VC series of experiments 2 and 4 refers to the same peak position. Furthermore, the high F_0 peak (vc>h) of experiment 4 is identical with the s/s peak used in experiment 2 (see sections 2.2 and 2.3). Hence, the 2 stimuli were identical in almost all respects (except for small differences in duration and intensity due to the larger experimental framework experiments 3 and 4 belong to). However, the 2 stimuli yielded clearly different percentages of 'not matching' judgements. While the sixth stimulus of the vc>h condition was judged as 'not matching' with the context utterance in almost 80% of the cases, the sixth stimulus of the s/s condition received only slightly more than 60% of the possible 'not matching' judgements. This deviation will be the subject of a more general discussion in section 4.3.

4. Discussion

4.1 Interpretation of the Findings

The context utterances *Jetzt versteh' ich das erst* ('Now, I understand'), *Ganz bestimmt* ('definitely'), and *Ist doch logisch* ('but that's logical') were selected to match with the following stimulus utterance only, if the latter has a medial peak on the accented syllable *Ma-* of *Malerin* ('painter'). Otherwise, the meanings of the two utterances do not go together in a sensible way. Provided that this meaning relation indeed guides judgements, 'matching' can be equated with the identification of the medial peak category in the stimuli, whereas 'not matching' can be interpreted as identification of the early peak (in case of the CV series) or the late peak category (in the case of the VC series), respectively. Based on this assumption, the results of the present study can be summarized as follows (see fig. 3 for peak shape illustrations and the corresponding abbreviations).

First, the results clearly show that the perception of German early, medial, and late peaks is not determined solely by the F_0 peak synchronization, represented by the position of the peak maximum relative to the boundaries of the accented vowel. Additionally, the overall shape of the peak as well as the peak height have a considerable influence on the perception of the three peak categories.

For F_0 peaks with maxima before the accented vowel onset, the slow falls in the s/s and f/s peaks support the perception of medial peaks. The same is true for the low peak height (cv<h) within this range of peak synchronizations. For F_0 peaks with maxima after the accented vowel onset, on the other hand, the perception of medial peaks is supported by the fast-rising peak shapes f/f and f/s as well as by the high F_0 peak (cv>h). In the case of F_0 peaks with maxima after the accented vowel offset, medial peaks can still be perceived if the corresponding F_0 peaks show a slow rise as in s/f and s/s or a low peak height as in vc<h. Fast-rising peak shapes and/or high F_0 peaks support the perception of late peaks. Hence, peak height is not only involved in the signalling of (different degrees of) emphasis [Kohler, 1991a, b; Ladd and Morton, 1997].

The findings are largely in line with the hypotheses put forward in the introduction (see section 1.4). Only hypothesis 1c has to be rejected. The slow-rising/fast-falling peak shape s/f did not support the perception of late peaks compared with the symmetrically fast rising-falling shape f/f. The results reveal an opposite shape effect. Hypotheses 1 and 1b, on the other hand, can be accepted without reservation. Peak shape had an influence, and the symmetrically slow rising-falling peak shape s/s supported the perception of medial peaks for peak synchronizations inside and outside the accented vowel. For the remaining hypotheses 1a and 2, the results show the expected effects of shape and height, but not as generally as postulated. For example, it was found that increasing the peak height supported the perception of medial peaks, as expected by hypothesis 2. However, this was only true for F_0 peaks showing the maximum within the vowel. For F_0 peaks before or after the accented vowel, an increased peak height supported the perception of early or late peaks, respectively. So, increasing the peak height generally raised the identification rate of all peak categories within their characteristic synchronization range. Furthermore, it was assumed in hypothesis 1a that the fast-rising/slow-falling peak shape f/s supports the perception of medial peaks compared with (a) the slow-rising/fast-falling shape s/f as well as with (b) the symmetrical fast rising-falling shape f/f. Again, this was only found for restricted to limited synchronization ranges. In comparison with s/f, the support of medial peak perception by f/s only holds for F_0 peaks with prevocalic maxima, while in comparison with f/f, it only holds for peaks with a maximum in the vowel.

Finally, the findings show that the influences of peak shape and height also affect the location of the perceptual boundary between the early and medial as well as between the medial and late peak categories. It may be assumed that a perceptual boundary between two categories causes a maximal uncertainty, which, in turn, results in judgements given by chance. Therefore, the perceptual boundary was defined as the ambiguous identification level of 50%. If this definition is applied to the F_0 results for the peak shifts from left to right, i.e. to the identification courses of figure 5a–d, a picture emerges which is displayed in figure 6. It shows for each of the peak shape and height conditions the corresponding peak synchronization which yields around 50% identification of early and medial or medial and late peaks, respectively.

Figure 6 shows that the perceptual change from early to medial can already take place before the maximum of the shifted F_0 peak enters the accented vowel. This

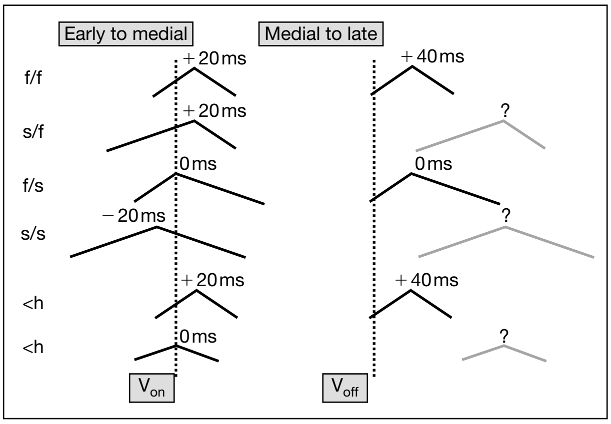

Fig. 6. F_0 peak positions (in ms) relative to the onset and offset of the accented vowel (V_{on} and V_{off}) at the perceptual transitions from early to medial peak and from medial to late peak for the different peak shape and height conditions. Perceptual transitions are defined by a change in the majority of judgements within the F_0 peak shift continua from left to right as shown in the identification courses of figure 5a–d. The grey peaks indicate that a perceptual transition (from medial to late) did not occur within the peak shift continua used in the present experiment, but may occur for further peak shifts to the right. Thus, the synchronizations of the grey peaks are arbitrary.

concerns the F_0 peaks showing either a low peak height or a slow-falling peak shape. For the peak shift from left to right, the slow rising-falling shape causes the earliest perceptual change. On the other hand, F_0 peaks with a fast-falling shape or a large peak height need to have their maximum shifted into the vowel, before they bring about a perceptual change from early to medial. For the peak shift beyond the vowel offset, covering the perceptual change from medial to late, the influences of shape and height are even stronger. That is, F_0 peaks with a low peak height or a slow rise did not cause a perceptual change from medial to late at all. Consequently, late peaks were only signalled by high or fast-rising F_0 peaks (with postvocalic maxima).

It has to be pointed out that the final stimulus 6 of the VC series with the low peak (i.e. vc<h; fig. 5d) received a late peak identification of about 52%. So, strictly speaking, the identification course does exceed the threshold for the change from medial to late. However, this can be ignored for two reasons. First, the identification course shows a clear saturation from stimulus 5 to 6. Second, informal listening by the author shows that a further shift of the F_0 peak to the right causes a disturbance in the rhythmic pattern of the utterance (due to an uncertainty in the accent position) and cannot contribute much to enhance the perception of late peaks. The latter is also true for further shifts of the slow-rising F_0 peaks s/s and s/f. So, it is unlikely that the corresponding identification courses would ever reach or go clearly beyond the 50% threshold (fig. 5b). To indicate, however, that this question cannot be settled, grey F_0 peaks with question marks are given in figure 6 at an arbitrary later point in time.

4.2 Phonological Modelling

The results of the present study clearly corroborate the existence of three meaningful (peak) categories in German intonation and shed light on their functions and signalling. The three intonation categories are considered in different phonological frameworks. In the KIM, they are represented by the early, medial, and late peaks. Moreover, in view of the examples in figure 1 and the functions outlined in section 1.2, they are presumably analyzed as H + L*, H*, and L* + H in the AM framework for German, GToBI [Grice and Baumann, 2000]. The phonological representations of the intonation categories in the KIM and in GToBI make (different) predictions about the phonetic manifestations of each category, which can be compared with the identification profiles yielded in the perception experiments.

As regards the KIM, it should not happen that peak shape and height have an influence on the perception of early, medial, and late peaks. Although the KIM is a contour model and thus regards the rising-falling peak as a holistic unit, the peak synchronization (represented by the maximum) is postulated as the only relevant variable. Furthermore, it is assumed in the KIM that the early, medial and late peak categories are separated by the different synchronizations the peak can have in relation to the boundaries of the accented vowel. This strict link between synchronization ranges and phonological categories is undermined by the present findings. For instance, it goes against the phonological organization of the KIM that medial peaks can already be signalled by F_0 peaks with maxima *before* the accented vowel onset and that F_0 peaks with maxima clearly *after* the accented vowel onset can still be perceived as early. Nevertheless, referring the F_0 movements to the accented vowel meets the synchronization-dependent influences of peak shape and height and the category transitions within the F_0 peak shift continua more closely than the association of F_0 events with the complete syllable, as done in the AM phonological approach [Pierrehumbert, 1980]. For instance, the results point to a connection between the accented vowel and the part of the peak that influenced the peak category identification. That is, perceptual influences seem to be restricted to the peak movement that dominates in the accented vowel, and shape as well as height effects change (i.e. they appear, disappear, or are reversed) for F_0 peaks shifted across the vowel boundaries. Moreover, the perception of late peaks requires a low plateau that reaches into the accented vowel and not only into the corresponding syllable. Finally, although the vowel boundaries need not coincide with the perceptual boundaries between early, medial, and late (fig. 6), they are all still located closely around the vowel onset and offset.

On the other hand, the AM framework does at least account for some of the shape effects, since different durations of F_0 peak movements can be represented by adding trailing or leading tones to the central (starred) tone. So, the bitonality of the accent H + L* indicates a rigid temporal relationship between the two adjacent tonal events, which corresponds to a temporally close phonetic alignment [Pierrehumbert, 1980, 2000]. That is, H + L* represents a short F_0 fall, which was in fact found to support the perception of this category. In the same way, the L* + H pitch accent predicts that a fast rise is essential for the perception of this category. Again, this is in line with the perceptual findings. Nevertheless, the potential of the AM phonology to cover the complete findings of the present study is limited, since only two tonal events are allowed within a pitch accent and since these events can only be projected onto two tonal levels. Due to the latter, influences of peak height are not considered at all.

In summary, the identification profiles received in the present study point to a global signalling of the intonation categories, which may involve local events like rise onset, peak maximum, and fall offset as well as their time and frequency relations and their alignment relative to articulatory landmarks. In this way, the perception of the three categories of German intonation is in favor of phonological concepts that comprise elements of both diametrically opposed phonological approaches, the contour approach and the level tone approach. A first step to such an alternative concept was made by Niebuhr [2006].

4.3 Critical Remarks

In connection with figure 5b and d, it was pointed out that the stimuli with the high F_0 peak and with the fast rising-falling peak shape (i.e. vc>h vs. f/f) are physically almost identical. Nevertheless, stimulus 6 in the vc>h condition was judged 'not matching' in almost 80% of the cases, whereas in the f/f condition stimulus 6 only yields slightly more than 60% 'not matching' judgements. This discrepancy is likely due to the different preceding context utterances used in the two conditions, i.e. it does *not* indicate a different identification rate of the late peak category. The reason for the 20% more 'not matching' judgements in the vc>h condition is that the context utterance *Ist doch logisch* ('but that's logical') was more suitable to separate the meaning difference between the medial and late peak category than the context utterance *Ganz bestimmt* ('definitely'; see section 2.4). This interpretation entails that the subjects judged the utterance pairs in fact with regard to their meaning relations, i.e. they were not just comparing their melodies. In the latter case, there would have been no discrepancy between the judgements of the two conditions.

The effect explained above clearly shows that an appropriate context utterance is the decisive element (and the weak spot) in the function-based indirect identification task. Without an appropriate context, phonetic factors contributing to the signalling of intonational categories or even the categories themselves may remain undetected. For instance, the fact that Kohler [1987, 1991c] and Kleber [2006] failed to show the late peak category in their data of German and British English may be due to their context utterances, which were more suitable to detect the meaning change between early and medial than the one between medial and late.

Alternative experimental paradigms used in intonational and prosodic research are imitation [e.g. Pierrehumbert and Steele, 1989], prosodic restoration [Xu et al., 2004], interactive manipulation [Ambrazaitis, 2006] and semantic differentials [e.g. Uldall, 1972; Dombrowski, 2003; Ambrazaitis, 2005; Kohler, 2005]. In all of these alternative approaches, however, we need to make sure that the results indeed refer to instances of the same intonational or prosodic categories. The only way of doing this (without circular reasoning) is to make use of the fact that the units investigated are part of a means of communication, i.e. the speech code. That is, we cannot escape the critical challenge of referring to function and meaning.

Finally, the validity of the present findings needs to be discussed. In this connection, it may be pointed out that the present findings match well with the patterns observed in the production of the peak categories. Late peaks, for instance, are consistently realized with a postvocalic F_0 maximum (in the syllable structure investigated in this study), a fast rise preceded by a low plateau after the vowel onset, and a peak height larger than the ones of early and medial peaks [Gartenberg und Panzlaff-Reuter, 1991] (see also fig. 1). The same F_0 pattern seems to hold for the peak category L* + H in American English [Pierrehumbert and Steele, 1989]. In the present study, it was exactly this F_0 pattern which caused a perceptual change from medial to late. A low peak height suppressed the perception of late peaks, and the same was true for the slow-rising peak shapes s/s and s/f. As pointed out in section 2.2, all s/f and s/s peaks in the VC series are marked by a prevocalic rise onset (fig. 4), i.e. they do not show a low

F_0 plateau after the accented vowel onset. The latter, however, was present in stimuli 4–6 with the fast-rising peak shapes f/f and f/s, which were able to change the perception from medial to late. Hence, it can be concluded that a low F_0 plateau after the accented vowel onset is an essential component for the late peak identification. Furthermore, the medial peak category was most clearly signalled by F_0 peaks having their maximum immediately before the accented vowel offset, as in stimuli 1–3 of the VC series (fig. 5b, d). It is known from several acoustic analyses [e.g. Gartenberg and Panzlaff-Reuter, 1991; Grabe, 1998; Niebuhr and Ambrazaitis, 2006] that this synchronization range is typical for the realization of German medial peaks (fig. 1).

With regard to section 1.3, the interplay of F_0 peak synchronization, shape and height found in the present study not only mirrors the F_0 peak patterns that appear in the production of German early, medial, and late peaks. It is also in line with the observations in the perception of the German peak categories made by Niebuhr [2005]. Moreover, since a higher F_0 peak accelerated the perception change from early to medial (fig. 6), the present results replicated previous perceptual findings by Kohler [1991c]. The fact that the empirical picture for German is consistent can be regarded as support for the validity of the present findings. However, if the latter are related to the findings and observations from other languages, the situation is less clear. Quite a few of those findings do match well with the ones of the present study, but there are also a number of differences. The latter is, among others, reflected in the rejection of hypothesis 1c, which was derived from the perception experiments in Hungarian (see sections 1.3 and 1.4). There are many possible reasons for these differences, such as different methodological approaches or inadequate interpretations. However, it is also possible that the signalling of the German early, medial, and late peaks is not completely parallel to the peak categories in other languages. These speculations should be taken as a motive for repeating the experimental series of the present study in other languages. Such research should be guided by the perspectives sketched in the following section.

5. Research Perspectives

The considerations in section 4.3 clearly show that we need a better and more comprehensive understanding of the functions and meanings of intonational categories like the F_0 peak patterns investigated in the present study. Increasing our knowledge about what is signalled is an essential prerequisite to investigate how it is signalled [Xu, 2004].

In this connection, the present investigation revealed that the perception of the German peak categories is based on multiple, interacting cues, involving peak shape, peak height, and peak synchronization. For the latter, the accented vowel turned out to be the decisive segmental point of reference. On this basis, Niebuhr [2006] developed perceptually founded coding profiles for each of the peak categories, which consist of two co-ordinated patterns: a perceptual prominence pattern and a tonal pattern. The latter takes into account that the perceived speech melody differs in many aspects from the acoustic F_0 course, not only due to contributions of top-down knowledge, but also due to perceptual constraints like the differential representation of F_0 movements as single tones or tonal movements, depending on spectral properties and dynamics of the signal [House, 1990]. In this framework, peak category identification means pattern recognition.

Independent of the perceptual constraints, physiological constraints in the production of F_0 and its co-ordination with supraglottal articulations are postulated [Xu, 2004]. In view of such constraints, it is likely that not all of the configurations of peak parameters investigated in the present study occur in naturally produced speech. However, looking beyond the natural stimulus range can be very useful or even essential to understanding the underlying perceptual mechanisms and is therefore a frequent strategy, particularly in psychological experiments across modalities [Goldstein, 1989]. On the other hand, the insights gained by such experiments have to be projected back onto what occurs naturally. In the present case, the insights about the perception of the German peak categories have to be related to findings from acoustic analyses in order to get a complete picture of the phonological forms and the different ways to convey them in production. This may also help to explain the F_0 patterns found for different peak categories in different segmental and prosodic contexts as well as for different speakers.

To extend the link between systematic perceptual and acoustic investigations is one of the tasks for future research. These efforts should be complemented by perception research aiming at the development of theories of the perception of speech melody [e.g. Niebuhr, 2006; Dilley, 2005], whose structures and measures offer a basis for phonological considerations on empirical grounds.

Moreover, further perception experiments are necessary, which go beyond F_0 and investigate the contribution of the intensity course as well as of segment durations and qualities to the signalling of intonational (peak) categories. Finally, this line of research has to include other intonation categories and languages.

Acknowledgements

I am very grateful to Klaus Kohler for lots of helpful and thoughtful comments on earlier versions of this paper. Furthermore, I would like to thank the two reviewers, Francis Nolan and Yu Xu, for their insightful suggestions and for giving me research ideas which even go beyond the immediate scope of this paper. Finally, thanks are due to the members of the former Kiel Prosody Group for many inspiring discussions.

References

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

Ambrazaitis, G.I.: Prosodic signalling of (un)expected information in South Swedish – an interactive manipulation experiment. Proc. 3rd Int. Conf. Speech Prosody, Dresden 2006, pp. 911–944.

Ambrazaitis, G.; Bruce, G.: Perception of South Swedish word accents. Lund Univ., Centre Lang. Lit., Dept. Ling. Phonet., Working Papers *52:* 5–8 (2006).

Andreeva, B.; Barry, W.J.: Intonation von ‘checks’ in der Sofia-Varietät des Bulgarischen. Res. Rep. Inst. Phonet., Univ. Saarland, Phonus *4:* 1–13 (1999).

Barry, W.J.: Prosodic functions revisited again! Phonetica *38:* 120–134 (1981).

Baumann, S.; Hadelich, K.: Accent type and giveness: an experiment with auditory and visual priming. Proc. 15th ICPhS, Barcelona 2003, pp. 1811–1814.

Bolinger, D.: Around the edge of language: intonation; in Bolinger, Intonation – Selected readings, pp. 19–29 (Penguin, Harmondsworth 1972).

Bruce, G.: Swedish word accents in sentence perspective (Gleerup, Lund 1977).

Chafe, W: Givenness, contrastiveness, definiteness, subjects, topics, and point of view; in Li, Subject and topic, pp. 27–55 (Academic Press, New York 1976).

Chen, A.: Reaction time as an indicator to discrete intonational contrasts in English. Proc. Eurospeech 2003 Conf., Geneva 2003, pp. 97–100.

Cohen, A.; ‘t Hart, J.: On the anatomy of intonation. Lingua *19:* 177–192 (1967).

David, H.A.; Hartley, H.O.; Pearson, E.S.: The distribution of the ratio, in a single normal sample, of range and standard deviation. Biometrika *41:* 482–493 (1954).

Dilley, L.: The phonetics and phonology of tonal systems; PhD diss. MIT (2005).
D'Imperio, M.; House, D.: Perception of questions and statements in Neapolitan Italian. Proc. Eurospeech '97, Rhodes 1997, pp. 251–254.
Dombrowski, E.: Semantic features of accent contours: effects of F_0 peak position and F_0 time shape. Proc. 15th ICPhS, Barcelona 2003, pp. 1217–1220.
Gartenberg, R.; Panzlaff-Reuter, C.: Production and perception of f_0 peak patterns in German. Arbeitsber. Inst. Phonet. dig. Sprachverarbeitung *25:* 29–115 (1991).
Goldstein, E.B.: Sensation and perception (Waldsworth, Belmont 1989).
Gósy, M.; Terken J.: Question marking in Hungarian: timing and height of pitch peaks. J. Phonet. *22:* 269–281 (1994).
Grabe, E.: Comparative intonational phonology: English and German; PhD diss. Universität Nijmegen (1998).
Grice, M.; Baumann, S.: Deutsche Intonation und GToBI. Ling. Ber. *181:* 1–33 (2000).
Halliday, M.A.K.: Intonation and grammar in British English (Mouton, The Hague 1967).
House, D.: Tonal perception in speech. Trav. Inst. Ling. Lund *24:* 7–163 (1990).
Kleber, F.: Form and function of falling pitch contours in English. Proc. 3rd Int. Conf. Speech Prosody, Dresden 2006, pp. 61–64.
Kohler, K.J.: Categorical pitch perception. Proc. 11th ICPhS, Tallinn 1987, pp. 331–333.
Kohler, K.J.: A model of German intonation. Arbeitsber. Inst. Phonet. dig. Sprachverarbeitung *25:* 295–360 (1991a).
Kohler, K.J.: Prosody in speech synthesis: the interplay between basic research and TTS application. J. Phonet. *19:* 121–138 (1991b).
Kohler, K.J.: Terminal intonation patterns in single-accent utterances in German: phonetics, phonology and semantics. Arbeitsber. Inst. Phonet. dig. Sprachverarbeitung *25:* 117–185 (1991c).
Kohler, K.J.: Timing and communicative functions of pitch contours. Phonetica *62:* 88–105 (2005).
Kohler, K.J.; Niebuhr, O.: The phonetics of emphasis. Proc. 16th ICPhS, Saarbrücken 2007.
Koiso, H.; Horiuchi, Y.; Tutiya, S.; Ichikawa, A; Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Lang. Speech *41:* 295–321 (1998).
Ladd, D.R.: Intonational phonology (Cambridge University Press, Cambridge 1996).
Ladd, D.R.; Morton, M.: The perception of intonational emphasis: continuous or categorical? J. Phonet. *25:* 313–342 (1997).
Nábelek I.; Hirsh, I.J.: On the discrimination of frequency transitions. J. acoust. Soc. Am. *45:* 1510–1519 (1969).
Nash, R.; Mulac, A.: The intonation of verifiability; in Waugh, van Schonneveld, The melody of language: intonation and prosody, pp. 219–241 (University Park Press, Baltimore 1980).
Niebuhr, O.: Perzeptorische Untersuchung zu Zeitvariablen in Grundfrequenzgipfeln; MA thesis Universität Kiel (2003a).
Niebuhr, O.: Perceptual study of timing variables in F_0 peaks. Proc. 15th ICPhS, Barcelona 2003b, pp. 1225–1228.
Niebuhr, O.: Sequenzen phonologischer Intonationsgipfel – Theoretische Möglichkeiten und empirische Realisierungen in Abhängigkeit vom segmentellen Kontext. Arbeitsber. Inst. Phonet. dig. Sprachverarbeitung *35a:* 55–95 (2005).
Niebuhr, O.: Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen; PhD diss. University Kiel (2006).
Niebuhr, O.; Ambrazaitis, G.: Alignment of medial and late peaks in German spontaneous speech. Proc. 3rd Int. Conf. Speech Prosody, Dresden 2006, pp. 161–164.
Niebuhr, O.; Kohler, K.J.: Perception and cognitive processing of tonal alignment in German. Proc. Int. Symp. Tonal Aspects Lang. Emphasis Tone Lang., Beijing 2004, pp. 155–158.
O'Connor, J.D.; Arnold, J.F.: Intonation of colloquial English (Longman, London 1970).
Ohala, J.J.: Prosody and phonology. Proc. 2nd Int. Conf. Speech Prosody, Nara 2004, p. 161.
Peters, B.: Form und Funktion prosodischer Grenzen im Gespräch; PhD diss. Universiät Kiel (2006).
Pierrehumbert, J.B.: The phonology and phonetics of English intonation; PhD diss. MIT (1980).
Pierrehumbert, J.B.: Tonal elements and their alignment; in Horne, Prosody: theory and experiment. Studies presented to Gösta Bruce, pp. 11–26 (Kluwer, Dordrecht 2000).
Pierrehumbert, J.B.; Steele, S.A.: Categories of tonal alignment in English. Phonetica *46:* 181–196 (1989).
Redi, L.: Categorical effects in the production of pitch contours in English. Proc. 15th ICPhS, Barcelona 2003, pp. 2921–2924.
't Hart, J.: Differential sensitivity to pitch distance, particularly in speech. J. acoust. Soc. Am. *69:* 811–821 (1981).
't Hart, J.; Collier, R.; Cohen, A.: A perceptual study of intonation. An experimental-phonetic approach to speech melody (Cambridge University Press, Cambridge 1990).
Uldall, E.: Dimensions of meaning in intonation; in Bolinger, Intonation – Selected readings, pp. 250–259 (Penguin, Harmondsworth 1972).
Xu, Y.: The PENTA model of speech melody: transmitting multiple communicative functions in parallel. Symp. 'From sound to sense', Cambridge 2004, pp. 91–96.
Xu, Y.; Sun, X.: Maximum speed of pitch change and how it may relate to speech. J. acoust. Soc. Am. *111:* 1399–1413 (2002).
Xu, Y.; Xu, C.X.; Sun, X.: On the temporal domain of focus. Proc. 2nd Int. Conf. Speech Prosody, Nara 2004, pp. 81–84.