10 Alignment and pitch-accent identification — Implications from F0 peak and plateau contours

Oliver Niebuhr

Abstract

The study scrutinizes the role of alignment of F0 movements in identifying two different pitch accents. Although this general issue was addressed for German, the pitch-accent contrast that was studied occurs cross-linguistically and is known as 'early' vs. 'medial' or H+L* vs. L+H*. The early pitch accent reaches the F0-peak maximum before the accented-vowel onset and hence falls into the vowel, while the medial pitch accent peaks after the vowel onset. This alignment-based identification model was recently undermined by studies that varied the slopes and ranges of the F0 movements or the extension of the F0-peak maximum. The latter parameter is taken up in the present perception experiments. Starting from a pointed rising-falling peak aligned at the accented-vowel onset, a peak and a plateau series were resynthesized by shifting either the entire peak or just the rising or falling movement into and away from the accented vowel. The peak and plateau stimuli were judged in indirect-identification and AX-discrimination tasks. While the results of the peak series are in line with the alignment model above, the plateau series showed a perceptual change from early to medial pitch accent for both the extension into and away from the vowel. The implications of the findings are discussed with regard to (a) a general atemporal model of pitch-accent identification and (b) the contrast between early and medial pitch accents at the ends of hat patterns.

10.1 Introduction

10.1.1 Setting up an alignment model for pitch-accent perception

In his early experimental investigations of German intonation, Kohler (1987) generated a stimulus continuum on the basis of the utterance *„Sie hat ja gelogen"* ('She's been lying'), in which he shifted a pointed rising-falling F0 peak across the only accented syllable *„-lo-"* of *„gelogen"* ('lying'). The F0-peak continuum, which was created by means of LPC resynthesis, was inspired by form-meaning relationships postulated in descriptive analyses of British English intonation (e.g., Halliday, 1967; O'Connor and Arnold, 1970), and represented an attempt to re-conceptualize these relationships by projecting them onto a common time axis.

A series of perception experiments was conducted that included form-related as well as directly and indirectly meaning-related judgments. The perception experiments were complemented by acoustic analyses of read-speech material from Gartenberg and Panzlaff-Reuter's study (1991). In summary, the evidence from production and perception suggested splitting up the F0-peak shift continuum across the accented syllable *„-lo-"* into three pitch accents that are defined by the location of the F0-peak maximum relative to the boundaries of the accented vowel. The first pitch accent is signalled by an F0 fall into the accented vowel, i.e. the peak maximum is reached before the vowel onset. In contrast, the second pitch accent requires an F0 rise into the accented vowel. More specifically, the peak of the rise must be reached within the vowel boundaries, but at least 30–60 ms after the vowel onset (cf. Kohler, 1987, 1991; Niebuhr, 2003, 2007a). Finally, the third pitch-accent category is defined by an F0 rise that starts within the accented vowel and continues beyond the vowel offset, typically into the vowel of the subsequent syllable.

Fig. 10.1 depicts a schematic illustration of the empirically founded alignment model for the perception of the three pitch-accent categories in German. This model provided the basis for the Kiel Intonation Model, KIM (cf. Kohler, 1991). Consequentially, alignment became a direct phonological feature in the KIM. The phonological alignment concept is also reflected in the labelling of the tripartite pitch-accent contrast as *'early'*, *'medial'*, and *'late'*. The labels refer to the F0-peak maximum location relative to the accented-vowel boundaries (cf. Fig. 10.1).

Unlike KIM, the phonology of the autosegmental-metrical (AM) approach to intonation (cf. Pierrehumbert, 1980; Ladd, 1996) is not directly built on alignment diffe-

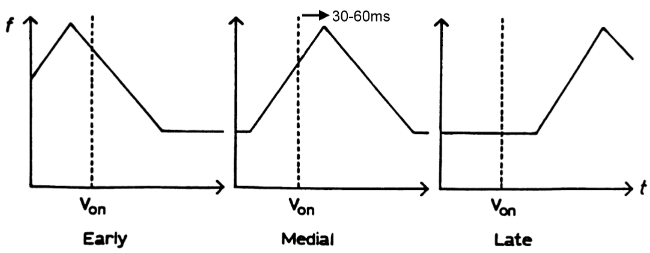

Fig. 10.1: Schematic illustration of the alignment model for pitch-accent perception that provides the basis of the Kiel Intonation Model, KIM. With reference to vowel onset (Von) and F0-peak maximum, the three pitch accents 'early', 'medial', and 'late' can be distinguished that occur cross-linguistically. The illustration was taken from Gartenberg and Panzlaff-Reuter (1991:31).

rences. Moreover, it is not the peak as a whole that constitutes the basic phonological unit in the AM approach. Rather the peak is broken down into pivotal F0 points, i.e. tones. Phonological differences result from associations of tones with accented syllables and from additional linkages between two tones. Thus the AM phonology represents the early, medial, and late pitch accents as H+L*, L+H*, and L*+H (cf. Grice and Baumann, 2000). The mandatory starred tone shows the first-order association with the accented syllable; and as in H+L* and L*+H the starred tone can be additionally linked with a leading or a trailing tone. Even though time is not part of this phonological structure, the associations and linkages of tones can be translated into certain alignment ranges on the phonetic level. Most importantly, it may be expectd that the phonologically starred tone manifests itself phonetically on the corresponding accented syllable, and an additional linkage with a leading or trailing tone is expressed in a temporally close and constant succession of the two tones. So, for example, given a rising-falling F0 peak like in Fig. 10.1, it is obvious that the alignment range of H+L* precedes the range of L+H* (although there can be some overlap between the ranges). Moreover, in explaining the phonetic exploitation of the phonologically staked alignment ranges, the accented vowel and the F0-peak maximum are used as important points of reference in AM studies (cf. Ladd et al., 2000; Rathcke and Harrington, 2006, 2007; Barnes et al. 2010; Sadat-Tehrani, 2009). So it may be claimed that, despite their different phonological interpretations of pitch-accent related F0 peaks, KIM and AM are built on a comparable understanding of pitch-accent identity, which is rooted in the notion of alignment. On this basis, Fig. 10.1 may be regarded as showing the predominant model of pitch-accent identification.

Starting from this predominant model, the present study addresses the perception of the early vs. medial pitch-accent contrast by comparing the results of F0-peak stimuli, on which the perceptual model was built, with the results F0-plateau stimuli, which up to now have been widely ignored in intonation research due to the focus on local F0 minima and maxima of pitch-accent movements. The implications of this comparison run contrary to the current alignment model.

10.1.2 The cross-linguistic use of the early vs. medial pitch-accent contrast

As for the meaning of the early and medial pitch accents, detailed examinations by Kohler (1987, 1991, 2005) as well as by Dombrowski (2003) and Niebuhr (2007a) led to the conclusion that the pitch accents are used in German as part of the speaker's argumentation structure. The early pitch accent characterizes the accented information as fixed and settled and hence as not worth further discussions. The medial pitch accent puts information up for discussion. This not only includes opening a new line of argument, but also taking up an item of information again in order to discuss or present it in a new light. It is important to note that these pitch-accent meanings do not directly relate to information structure, i.e. whether a piece of information is given or new in a discourse. For example, the KIESEL sentence „*Bier ist ausverkauft*“ ('Beer is sold out', cf. Niebuhr, 2011) was produced with an early pitch accent on „*ausverkauft*“ ('sold out'), even though this information was newly introduced into the discourse. By using the early peak, the speaker signals to the listener that the unavailability of beer is reliable information that will not change in the (near) future. Further discussions are 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 a medial pitch accent on „*Montag*“ ('Monday'), although this date had already been agreed upon in the previous discussion and, as such, is thus information that could be taken as given . The medial pitch accent was used to indicate that the given date will be dealt with from a different angle, viz. travelling. Correspondingly, the summarizing statement was followed by the question „*Und wie kommen wir da hin?*“ ('And how do we get there?'). Even though it is claimed here that the early and medial pitch accents signal argumentation and not information structure in German, there is of course some correlation between the two structures and hence also between the early and medial pitch accents on the one hand and given and new information on the other (cf. Baumann 2006). A new line of argument is frequently opened by introducing new information; and marking a piece of information as fixed and settled often originates from the fact that this piece of information has been discussed before. However, more recent evidence supports the argumentation-structure perspective by showing that information structure is more consistently related to the presence/absence of a pitch accent rather than to the type of pitch accent (cf. Röhr and Baumann, 2010).

From a cross-linguistic perspective, perception experiments based on F0-peak shift continua similar to the continuum of Kohler (1987) showed that the early vs. medial alignment contrast is very widespread, but tied to language-specific meaning contrasts. While in British and American English early and medial pitch accents have meanings comparable to those of their German counterparts (cf. Pierrehumbert and Steele, 1989; Redi, 2003; Kleber, 2006), quite a few languages such as Hungarian (cf. Gósy and Terken 1994), Bulgarian (cf. Andreeva & Barry 1999), Neapolitan Italian (D'Imperio and House, 1997, D'Imperio, 2000), and Russian (cf. Marakova, 1999; Rathcke, 2006) use the early vs. medial pitch-accent contrast for the differentiation of statements and questions. In tone or tone-accent languages like Mandarin and Stockholm Swedish, the alignment contrast serves to distinguish words (cf. Garding et al., 1985; Kohler, 1991; Ambrazaitis 2009). Against this backdrop, Kohler (1991) claims that the differentiation of early and medial pitch accents is a language universal, and that the corresponding alignment contrast shown in Fig. 10.1 is due to a general perceptual mechanism (cf. also Niebuhr and Kohler 2004).

10.1.3 Undermining the alignment model for pitch-accent perception: The issue of plateaux

In recent years, more detailed perceptual studies have revealed that pitch-accent identification is not solely a matter of the alignment of the F0-peak movements. Rather, the perception of a pitch accent can be characterized as the interplay of the alignment, shape, and range of the F0 peak. For example, using an F0 peak with a greater F0 range and/or shallower slopes in a continuum of alignment relative to the accented vowel supported the identification of the medial pitch accent in German (cf. Niebuhr, 2003, 2007a). In terms of question identifications, similar shape and range effects on medial pitch accents were found in studies on Hungarian (cf. Gósy and Terken, 1994), Neapolitan Italian (cf. D'Imperio and House, 1997), and Russian (cf. Rathcke, 2006).

For example, if the gradient of the falling slope of an F0 peak used in an alignment continuum is made less steep, the boundary between the early and the medial pitch accent will occur before the 30–60 ms region, where it would otherwise normally occur (cf. Gósy and Terken, 1994; Niebuhr, 2003, 2007a). Moreover, F0 peaks with two shallow slopes can even trigger medial pitch-accent identifications when the F0 maximum, and hence the rise, are aligned before the vowel (cf. D'Imperio and House, 1997; Niebuhr, 2003, 2007a).

Brought down to a common denominator, the cross-linguistic findings show that the more "high F0" is contained in the rising-falling peak contour, the more a stimulus continuum from early to medial is perceived as belonging to the medial pitch accent. This fact — in addition to the effects of shape and range themselves — undermines the modelling of the early vs. medial pitch-accent contrast by means of a strict alignment concept. There is no one-to-one mapping between the identification of the pitch accent and the alignment of the corresponding pivotal F0 points, including the alignment of the peak maximum relative to the vowel boundaries.

The most extreme way of amplifying the high-F0 content in rising-falling contours is not to flatten the slopes or to increase the F0 range, but to extend the F0 maximum. That is, the pointed F0 peak is changed into a plateau with a high F0 level between rise offset and fall onset. Such plateaux should have a strong predisposition towards the medial pitch accent. Initial support for this assumption comes, among others, from studies that examined the production of pitch accents. They found that particularly medial pitch accents can be realized as an F0 plateaux, cf. Silverman and Pierrehumbert (1990), House et al. (1999), as well as Knight and Nolan (2006) for English, Welby (2003) for French, Rathcke (2006) for Russian, and Grabe (1998) for German. Welby (2003) notes that up to 31% of the F0 maxima she analyzed manifested themselves as plateaux, which implies that intonational plateaux are not a side issue.

Yet, in the majority of cases, plateaux were and still are the undesirable by-products rather than the starting points of intonation studies. Plateaux interfere with the actual areas of research, namely pointed F0 peaks and peak maxima. For instance, in a critical discussion of their study on pitch-accent alignment Silverman and Pierrehumbert (1990:79) complain that plateaux were one reason why *"the time point for the F0 maximum is probably the least reliable of the measurements"*. A similar note can be found in the study of Welby and Loevenbruck (2006). On the other hand, Welby and Loevenbruck also point to the necessity for plateaux to overcome their role as disruptive elements in intonation research. They conclude that *"when or why H [...] is realized as a simple peak or as a plateau remains to be investigated"*.

So far, however, only a few studies complied with Welby and Loevenbruck's implicit demand (2006) and explicitly addressed the interplay of plateaux with other F0 parameters in the production and perception of pitch accents. As for the medial accent, these studies include Knight and Nolan's production experiment on British English as well as the perception experiments of D'Imperio and Gili Fivela and those of Barnes and colleagues on varieties of Italian and English. Their findings provide additional support for the predisposition outlined above. For example, by shifting rising-falling peak and plateaux contours into the vowel, D'Imperio (2000) showed for Neapolitan Italian that plateaux are a more effective means of signalling medial pitch accents than pointed peaks. Compared with plateaux, pointed peaks had to be shifted with their rising movements further into the vowel in order to yield similarly clearly medial pitch accent identifications (which are associated with questions in Neapolitan Italian; cf. also D'Imperio et al., 2010; Gili Fivela and D'Imperio, 2010). This finding was recently replicated by Barnes et al. (2011) for American English and refined with regard to the role of the segmental properties underlying the plateaux. Consistent with the perceptual picture, the production study of Knight and Nolan (2006) showed that the rising movements of pitch-accent plateaux do not go as far into the accented vowel than the rising movements of pitch-accent peaks, and that a plateau realization is more likely for pitch accents with a reduced F0 range (which suggests that plateaux were to compensate for the loss of high F0 content in this contexts).

In addition to representing a single pitch accent, high F0 plateaux also occur in concatenations of two pitch accents that form so-called "hat patterns" (cf. Cohen and 't Hart, 1967; Adriaens 1991). In line with Kohler (1991) and Wunderlich (1988), Ambrazaitis & Niebuhr (2008) provided experimental evidence that a plateau concatenation in a hat pattern creates a different intonational meaning than that of a concatenation of two pitch accents by an F0 dip. The experiment of Ambrazaitis & Niebuhr was carried out for German, although their findings should also be applicable to other languages such as English. Moreover, referring to a survey of the F0 patterns in the Kiel Corpus of Spontaneous Speech (cf. Peters et al., 2005), Ambrazaitis & Niebuhr postulate that the hat pattern vs. the dip pattern contrast can be realized regardless of the types of pitch accents on either side. For example, despite a concatenation by plateau, the contrast between early and medial can be maintained for the second pitch accent. It is obvious that this claim is inconsistent with the alignment model in Fig. 10.1, according to which, the perception of early or medial pitch accent depends on whether the F0 rise ends before or after the accented-vowel onset. In a concatenation by plateau, there is no separate F0 rise for the second pitch accent. So, if the distinction between early and medial pitch accent is maintained under these circumstances, it most likely depends on whether the *F0 fall* at the end of the plateau starts before or after the accented-vowel onset. However, this has yet to be shown experimentally. So far, there is only anecdotal evidence along these lines from Kohler (1991), who noted with reference to his own introspection that a perceptual change from early to medial pitch accent can in fact be achieved by extending a high F0 plateau with a final fall into the accented vowel. However, at the same time he points out that *"the medial fall in the second position of the 'hat pattern' [. . .] lacks the essential rise of a 'medial peak', and is therefore not a proper representative of this phonological category"* (Kohler 1991:323). So, overall the issue of the early vs. medial pitch-accent contrast at the end of a hat pattern must be regarded as unresolved. The experiments of the present study will contribute to the settling of this issue.

10.1.4 Research questions and hypotheses

In summary, the previous section showed that empirical evidence from a number of perception experiments carried out for different languages challenge the alignment-based model of pitch-accent identification outlined in 10.1.1. According to this model, the perceptual differentiation between the two phonological pitch-accent types 'early' and 'medial' is supposed to be linked with F0 maxima before or after the accented-vowel onset and hence with a fall or a rise into the vowel, respectively. However, amplifying the high-F0 content of a peak that is shifted into the vowel extends the alignment range that triggers medial pitch-accent identifications so that even F0 peaks with maxima before the accented-vowel onset can become medial pitch accents. Changing pointed peaks into plateaux is a very effective way of achieving an amplification of the high-F0 content; and we know from acoustic analyses that pitch accents can be realized as plateaux, either based on phonetic variation within a single pitch accent or due to a hat pattern concatenation with the preceding pitch accent. Yet, following from the focus of intonational phonology on the local F0 onsets, maxima, and offsets of pitch-accent peaks, the impact of a change from peak to plateau on medial pitch-accent perception has received scant attention so far. The fragmentary findings that do exist suggest that the effect of extending an F0 peak to a plateau fits well into the interplay of F0 parameters that may be reduced to the following simplified equation: *more high-F0 content = more medial pitch-accent identifications*.

On this basis, many important questions arise concerning the power of plateaux to signal medial pitch accents beyond the known alignment limits. For example, although it was demonstrated that the originally postulated alignment relationship between pitch-accent movements and accented-vowel onset in identifying early and medial pitch accents does not hold, the empirical evidence was still in line with the alignment model insofar as the change from early to medial occurred for shifts of rising-falling F0 peak or plateau contours to the *right*. That is, rise and peak maximum are shifted towards, and eventually into, the accented vowel. However, amplifying the high-F0 content between rise and fall is, in principle, independent of where the plateau is created and in which direction it is extended. Therefore, is a plateau contour also capable of triggering a change from early to medial pitch accent if it is extended to the *left*, i.e. *away* from the accented vowel so that the latter only contains the falling F0 movement? If so, this would be completely incompatible with the original alignment model.

In connection with the absence of a rise into the vowel, a further important question concerns the claimed contrast between early and medial for the final pitch accent in a hat pattern. Is the shift of a high and finally falling F0 plateau into the accented vowel paralleled by a perceptual shift from early to medial pitch accent? Furthermore, if this is the case, then at what point does this occur? And — with reference to the line of research on Italian started by D'Imperio (2000) — is such a high F0 plateau a better or a worse cue to medial pitch accents, compared with a pointed F0 peak that rises into the accented vowel?

The present study aims at addressing these questions in a series of perception experiments. They are built around a stimulus series in which a pointed F0 peak aligned at the accented-vowel onset is changed into a successively extended plateau contour by shifting either the falling or the rising slope into or away from the accented vowel, while the other slope remains constant. This plateau series is complemented by a peak series in which a parallel shift into and away from the accented vowel concerned not just the falling or rising slope, but the entire pointed peak. By means of the plateau and peak series the following three hypotheses are tested:

- *'Peak hypothesis'*, *H1*: The peak series will replicate the previous findings of Kohler (1987) and Niebuhr (2003, 2007a). That is, F0 peaks that are shifted with the rise more than 30–60 ms into the vowel will cause a (sudden) perceptual change from the early to the medial pitch accent, whereas pitch-accent identification will remain clearly early for the F0-peak shift in the opposite direction, i.e. away from the vowel.

- *'Plateau hypothesis'*, *H2*: Since plateaux strongly amplify the high-F0 content in the pitch-accent contour, *both* the extension of the high F0 plateau into *and* away from the accented vowel in the plateau series will shift the pitch-accent identification towards the medial pitch accent.

- *'Effectiveness hypothesis'*, *H3*: Following the impressionistic descriptions of Kohler (1991) and the cross-linguistic findings of D'Imperio (2000), D'Imperio et al. (2010) and Barnes et al. (submitted), it is expected that, for signalling medial pitch accents, the plateau extension into the vowel is more effective than the plateau extension away from the vowel, but still less effective than the F0-peak shift into the vowel.

The perception experiments are to be conducted for German. However, since the intonation contrast represented by the early and medial pitch accents is widespread across languages and even considered to be language-universal, it is reasonable to assume that the present findings will hold for many other languages as well, including those mentioned in 10.1.2. Finally, it must be noted that the following sections will continue referring to the examined pitch-accent contrast as early vs. medial. However, in view of the empirical picture, the terms will be used merely as labels of two different types of intonational form-function relationships, detached from their original temporal implications in the strict alignment model of Kohler (1987).

10.2 Indirect-identification experiment

10.2.1 Method

Stimuli

The stimuli of the perception experiment were generated on the basis of the utterance *„über Langeland“* (‘via Langeland’, [ʔyːbɐ ˈla̠ŋəla̠nt^{h}], a Danish island in the Baltic Sea). Apart from two voiceless sounds at both ends, *„über Langeland“* is characterized by continuous voicing and can hence be used for extensive F0 manipulations. Moreover, *„über Langeland“* was deliberately selected, as it was a quite short two-word utterance. This allows the use of the stimuli generated for the indirect-identification experiment also in the stimulus pairs of the subsequent AX discrimination experiment without exceeding the capacity of the auditory working memory (cf. Drenowski, 1980). The utterance *„über Langeland“* was produced by a male speaker of Northern Standard German (the author ON) with a single medial pitch accent on the syllable *„Lan-“* ([la̠ŋ]). As it was considered difficult to produce such a short and elliptic utterance in a naturally sounding way without a surrounding semantic-pragmatic context, the utterance was realized by embedding it in a short text passage. Several repetitions of this text passage were produced.

After having selected and extracted the most suitable production of *„über Langeland“* from the set of text passages (e.g., in terms of a clear accentuation, a moderate F0 level, and a modal voice quality), two stimulus series were generated analogously by means of the PSOLA resynthesis algorithm in PRAAT (cf. Boersma 2001). First, the overall F0 contour of the utterance was stylized at five contour points. F0 was then linearly interpolated. Two of the points constituted the onset and offset F0 values within the voiced section of the utterance. The point at the onset was given a value of 120 Hz, which was higher than the offset value, hence taking into account the general tendency of F0 to decline across utterances (cf. Maeda, 1974). The contour point at the offset was set to a terminal F0 level of 85 Hz. The remaining three contour points were used to form a pointed rising-falling F0 peak in the middle of the utterance. The pointed F0 peak replaced the naturally produced peak, taking over the original frequency values for rise onset (115 Hz), peak maximum (170 Hz), and fall offset (112 Hz). So, the F0 peak showed ranges of 6.5 or 7.0 semitones in the rising or falling slopes, respectively. In the time domain, the rising-falling F0 peak was aligned with its maximum at the vowel onset of the accented syllable *„Lan-“*. From there, both the rising and the falling slope extended over 120 ms into the syllable-initial alveolar lateral [l] and into the accented vowel [a̠].

The utterance with the five-point F0 pattern described was resynthesized and represented the *base stimulus* of the perception experiment. According to the previous findings of Kohler (1987), Niebuhr and Kohler (2004) as well as Niebuhr (2003, 2007a), the utterance should be perceptually ambiguous between an early or a medial pitch accent. The base stimulus provided the starting point for two stimulus series: a *peak* and a *plateau* series. Each series consisted of 20 stimuli. The two series are illustrated in Fig. 10.2. As can be seen in Fig. 10.2, both the peak and the plateau series resulted from shifting the temporal alignment of F0-peak movements.

In the peak series, the *complete* rising-falling F0 peak was shifted in 10 equally-sized increments of 10ms along the time axis *to the left*, i.e. into the syllable-initial lateral [l], as well as *to the right*, i.e. into the accented vowel [a̠]. At each F0-peak position a stimulus was resynthesized. In the

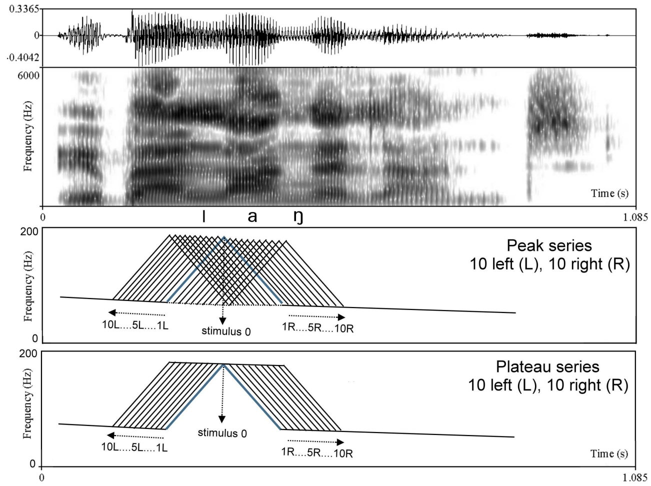

Fig. 10.2: Top panels: Oscillogram and spectrogram of the stimulus utterance *„über Langeland“*, with *„Lan-“* as the only pitch-accented syllable. Bottom panels: Schematic illustration of the 40 resynthesized intonation contours, resulting from 20 shifts, 10 to the left (L-stimuli) and to the right (R-stimuli), of the rising and/or falling F0 movements in the peak and plateau series. Additionally, the base stimulus 0 of the peak and plateau series is represented by a thicker grey line. It shows a pointed peak aligned at the accented-vowel onset.

course of the 100ms shift to the left, the F0-peak alignments (measured in terms of the peak maximum) covered the whole lateral as well as a small part of the preceding [ɐ], i.e. the vocalized /r/ in the syllable coda of *„über"*. In the opposite direction, the F0 peak was shifted to the right from the beginning to almost exactly the end of the accented vowel.

Parallel to the peak series, the plateau series also consisted of 10 shifts to the left and another 10 shifts to the right with increments of 10ms each. Limiting the length of the created plateaux to 10x10=100 ms in either direction reflects the average values of plateau durations that were provided by Welby (2003). However, unlike in the peak series, the 10 shifts to the left concerned only the *rising F0 slope*. Analogously, in the 10 shifts to the right only the *falling F0 slope* was moved. In consequence, the pointed F0 peak was changed into plateaux that were successively extended into the lateral on the one hand or into the accented vowel on the other. After each shift of the rising or falling slope a plateau stimulus was resynthesized.

While shifting the rising and/or falling F0 slopes in the peak and plateau series, the ranges of the slopes remained constant at 6.5 and 7.0 semitones. However, their absolute frequencies changed, as the values of rise onset and fall offset were adjusted to the F0 declination line (120–85 Hz) joining the initial and final contour points. In this way, the leftward and rightward moving F0 peaks and plateaux also moved slightly upward or downward along the frequency axis. Moreover, the plateaux received an F0 declination in between rise offset and fall onset that ran parallel to the declination line of the stimulus utterance. Such plateau-internal F0 declinations can also be found in natural German productions (cf. Kohler, 1991). Imitating this production pattern was to increase the naturalness of the stimuli and to avoid the auditive artefacts that result from resynthesizing sequences of physically identical F0 values.

The stimuli of the peak and plateau series were given numbers from 1 to 10 that represent the successive shift of the F0-peak movements. On this scale 10 refers to the most distant alignment from the original position. The numbers were complemented by the indices <L> or <R> that relate to the leftward and the rightward shift within each series. It must be noted in this context that those stimuli of the peak and plateau series with the same stimulus number and the index <L> are identical in terms of the aligned of the rising slope and its F0 maximum. Likewise, the peak and plateau stimuli with the same numbers and the index <R> are identically aligned in terms of the falling slope and its F0 maximum. The base stimulus with the pointed F0 peak, whose maximum coincides with the accented-vowel onset, will be referred to as 'stimulus 0'.

Task

The 2x20=40 peak and plateau stimuli that were derived from the base stimulus 0 as well as the base stimulus 0 itself were integrated into an indirect identification task. The indirect-identification paradigm has proved successful in numerous previous experiments that examined the perception of, for example, pitch accents or boundary tones (cf. Nash and Mulac, 1980; Kohler, 1987, 1991; Niebuhr and Kohler, 2004; Kleber, 2006; Niebuhr, 2003, 2007a; Ambrazaitis & Niebuhr 2008; Petrone and Niebuhr, 2009). The paradigm is based on the following idea. Instead of asking the subjects to identify pitch accents or boundary tones directly by applying a particular (meta-)linguistic label to the stimulus utterances, they judge whether the stimulus utterances match semantic-pragmatically with a constant context precursor. By selecting a context precursor that is compatible with the meaning of only one of the paradigmatically contrasting pitch accents or boundary tones in the stimulus continuum, the 'matching'–'not matching' judgments of the subjects represent indirect identifications of the accents or tones. The indirect-identification task has the advantage that intonational contrasts can be revealed and scrutinized that are difficult to grasp by simple (meta-)linguistic labels (e.g., unlike in the experiments in Neapolitan Italian, Russian or Hungarian, the subjects of the present experiment cannot simply be asked to identify the stimuli directly as either statements or questions). Moreover, since listeners are not alerted to any such labels in the indirect-identification task, it is more difficult for them to uncover the actual aim of the experiment.

In the present experiment the question *„Und wie willst du fahren?"* ('And which way do you want to go?') was used as constant context precursor. It was produced by a female speaker of Northern Standard German with an utterance-final rising intonation and two preceding pitch accents, a medial accent on *„wie"* and a late accent on *„fah-"*. The question precursor was attached to each of the 41 stimuli after a short pause of 350 ms that represents cross-linguistically a typical interval for non-overlapping changes in the turns of two dialogue partners (cf. Weilhammer and Rabold, 2003). From a semantic-pragmatic point of view, the question signals that the route of the dialogue partner was neither predetermined by a third person nor has it been mentioned before in the discourse. Rather, complying with its role as an information question as described by Huddleston (1994), the context precursor is an invitation to open a new line of argument by putting the chosen route up for discussion. Thus the context precursor was designed to be only compatible with the following stimuli, if the latter showed a medial pitch accent (cf. 10.1.2). In contrast, stimuli with an early pitch accent emphasize that the route 'via Langeland' is fixed and that there is no point in discussing any alternatives. This definite statement implies that the route has been discussed before, which clearly conflicts with the question precursor. From this point of view, 'matching' judgments can be interpreted as medial pitch-accent identifications in the stimuli, whereas 'not matching' judgments may be taken as identifications of the early pitch accent (with reference to previous production and perception studies it can be taken for granted that the created F0 continua included no other or more types of pitch accents than early and medial).

Subjects and Procedure

Five copies of the 40+1=41 context-stimulus pairs were arranged in an overall randomized order and played over loudspeakers in a sound-treated room at the Institute of Phonetics and Digital Speech Processing of the University of Kiel. The presentation of the 41x5=205 context-stimulus pairs was organized into 20 blocks of 10 and one block of five pairs. The blocks were separated by double bleeps and pauses of five seconds. The context-stimulus pairs themselves were introduced by a single bleep followed by a one-second pause. After each pair there was a pause of 4 seconds, during which the subjects made their judgments.

The context-stimulus pairs were judged by 18 native speakers of German (10 female, 8 male, average age 22.7 years). They were all undergraduate students of Phonetics and/or Linguistics at the University of Kiel, and none of them reported any hearing disorders. The 'matching'–'not matching' judgments were made by ticking boxes on prepared answer sheets. Each of the blocks of context-stimulus pairs was judged on a separate answer sheet. The subjects were instructed that they would hear repetitions of a short question-answer dialogue between a female and male speaker. While the wording of the dialogue was to remain constant, i.e. *„Und wie willst du fahren?"–„Über Langeland."*, the speech melody in the answer could differ between the repetitions. Due to these differences there could be answers that are inappropriate in the context of the preceding question. Thus the subjects' task would be to listen carefully to the question-answer dialogues and to judge after each dialogue whether question and answer did or did not match.

Before the actual experiment, the 18 subjects received a training block of 18 question-answer pairs in order to familiarize them with the task and the nature of the stimuli. The 18 pairs were composed of the 6 extreme stimuli of the two series, i.e. the 10R and 10L stimuli of the peak and plateau series as well as stimulus 0 (which is identical in both series). The 6 stimuli were presented three times in an overall randomized order. Including the instructions and the training block, the whole indirect-identification experiment took about 45 minutes.

10.2.2 Results

Fig. 10.3 and 10.4 provide a descriptive summary of the results received for the peak and plateau stimuli. Since the 18 subjects judged each stimulus five times, each data point in the two graphs represents 90 'matching'–'not matching' decisions. On this basis, it can firstly be observed in both Figures that the base stimulus 0 with the pointed peak at the accented-vowel onset was not judged clearly as either 'matching' or 'not matching'. Indeed, a χ^2-test showed that the 'matching' judgments of the 18 subjects did not differ significantly from chance level (=50%) for stimulus 0.

Furthermore, as regards the peak stimuli 1R-10R Fig. 10.3 displays that shifting the rising-falling peak to the right *into the vowel* of the accented syllable *„Lan-"* increases the number of 'matching' judgments rapidly to about 85%. This clear transition in the judgments is already complete for stimulus 3R, whose F0 peak is aligned with the maximum 30 ms after the accented-vowel onset. Correspondingly, it can be seen in Fig. 10.3 that the grey response curve of the remaining stimuli 4R-10R, in which the peak is shifted further into the vowel, runs almost flat. In contrast, the stimuli 1L–10L, which represent the F0-peak shift away from the accented-vowel onset to the left, yielded a clear decrease in 'matching' judgments. This decrease continues until a 'matching' level of less than 10%. This very low level is already reached at stimulus 4L, in which the F0 peak is aligned 40 ms before the vowel onset. Stimuli 8L did not receive any 'matching' judgments at all. In other words, the peak stimuli 4L–10L were almost exclusively judged as 'not matching'.

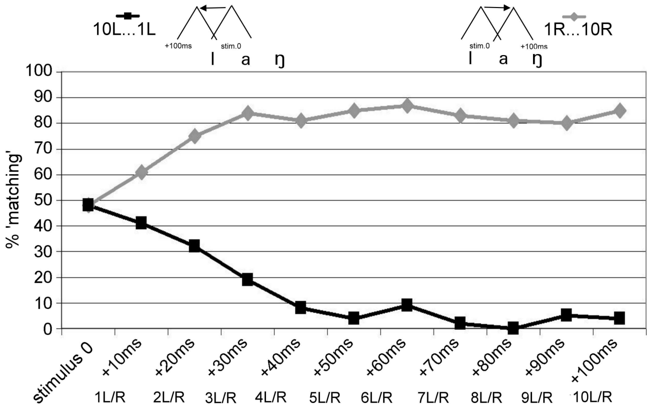

Fig. 10.3: Percentages of 'matching' judgments for the base stimulus 0 and the peak stimuli 1R–10R and 1L–10L in which the entire F0 peak was shifted in constant increments of 10 ms into the accented vowel (L, grey) or into the syllable-initial lateral (R, black). Each data point corresponds to 90 judgments.

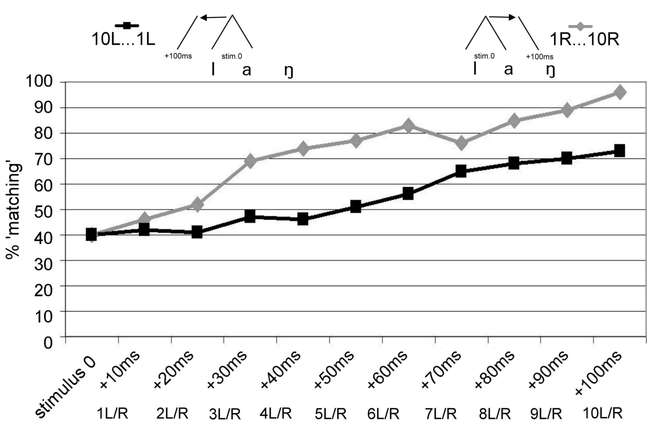

Fig. 10.4: Percentages of 'matching' judgments for the base stimulus 0 and the plateau stimuli 1R–10R and 1L–10L in which the plateau was successively extended into the accented vowel or into the syllable-initial lateral by shifting the fall or rise in constant increments of 10 ms to the left (L, grey) or to the right (R, black), respectively. Each data point corresponds to 90 judgments.

The diametrically opposed judgment behaviour for the left and the right branch of the peak series, i.e. 1R–10R and 1L–10L, is also expressed in two statistical measures. First, a t-test was run with the direction of the peak-shift as the independent variable and the sums of 'matching' judgments of each subject across 1R–10R and 1L–10L as the dependent variable. The t-test showed that the 1R–10R stimuli yielded highly significantly more 'matching' judgments than the 1L–10L stimuli (t=17.005, df=17; p<0.0001***). Additionally, there is a significant negative correlation (in terms of Pearson's product-moment coefficient) between the 'matching' sums across the 1R–10R stimuli on the one hand and the 1L–10L stimuli on the other (r=-0.728; df=8; p<0.01**). So, while the stepwise 10 ms peak shifts to the right increase 'matching' judgments, they decrease to a similar extent across the 10 ms steps of the leftward peak shift.

Fig. 10.4 shows for the plateau stimuli 1R–10R that not only the shift of a pointed peak to the right into the accented vowel was able to bring about an increase in the number of 'matching' judgments. The rightward shift of the falling peak movement and the concomitant extension of the high F0 plateau into the accented vowel had a similar effect. However, compared with the peak stimuli 1R–10R the 'matching' judgments increased more slowly across the plateau stimuli 1R–10R. For example, the plateau had to go 60 ms into the vowel in order to yield about 85% 'matching' judgments (cf. plateau stimulus 6R), whereas the pointed F0 peak already reached this 'matching' level 30 ms after vowel onset (cf. peak stimulus 3R, Fig. 10.3). But while the peak stimuli stagnated at a 'matching' level of about 85%, the 'matching' level in the plateau series increased beyond 85% for further plateau extensions to the right. Stimulus 10R with the longest plateau that spanned the whole accented vowel was almost exclusively judged as 'matching' (>95%, cf. Fig. 10.4). So, across the entire right branches of the peak and plateau stimuli, there is a crossing over in 'matching' frequencies. Due to this crossing over, the corresponding t-test that compared the 'matching' sums of 1R–10R between the peak and plateau series just missed significance (t=-1.844; df=17; 0.05<p<0.1).

In addition to this difference in quantitative detail, the major qualitative difference between the peak and the plateau series lies in the shift to the left, i.e. away from the accented-vowel onset. In the plateau stimuli 1L–10L, this leftward shift, which concerned the rising F0-peak movement and extended the high F0 plateau successively across the syllable-initial lateral, did not cause a decrease in 'matching' judgments, as was the case in the peak series. In fact, it increased the 'matching' judgments. However, this increase was slower again than that of the rightward plateau extension in stimuli 1R–10R. For example, the 'matching' level of the plateau stimulus 10L with the most leftward extended plateau, which included the whole lateral, was about 20% lower than the 'matching' level of 10R. That is, 10L received approximately 75% 'matching' judgments. Correspondingly, a t-test that compared the 'matching' sums of the plateau stimuli 1R–10R with the sums of the plateau stimuli 1L–10L produced the same result as the analogous t-test with the peak stimuli. The rightward shift in 1R–10R yielded highly significantly more 'matching' judgments than the leftward shift in 1L–10L (t=7.623; df=17; p<0.0001***). However, this overall difference in 'matching' judgments does not go along with negatively correlated 'matching' frequencies across 1R–10R and 1L–10L. Rather, the correlation is highly significantly positive (r=0.875; df=8; p<0.001***). So, the plateau extensions to the right and to the left both increase 'matching' judgments.

The test statistics of the t-tests for the remaining crosswise comparisons between the left and right branches of the peak and plateau series, which were not primarily for the hypotheses, are summarized in the following. The plateau stimuli 1L–10L yielded overall clearly more 'matching' judgments than the peak stimuli 1L–10L (t=5.790; df=17; p<0.0001***). Additionally, the overall number of 'matching' judgments was significantly higher for the peak stimuli 1R–10R than for the plateau stimuli 1L–10L (t=4.313; df=17; p<0.001***), and significantly lower for the peak stimuli 1L-10R than for the plateau stimuli 1R–10R (t=-12.602; df=17; p<0.0001***).

10.2.3 Conclusions

The peak and plateau stimuli of the experiment were attached to the context precursor *„Und wie willst Du fahren?"*. The information question implies that the route of the dialogue partner was neither predetermined by a third person nor had it been mentioned before in the discourse. Instead, the dialogue partner is prompted to present his route to the questioner. This semantic-pragmatic framework matches with the stimulus utterance *„über Langeland"* if the latter contains a medial pitch accent on *„Lan-"*. In contrast, with an early pitch accent on *„Lan-"* the stimulus utterance expresses that the speaker marks the route 'via Langeland' as fixed, in this way, for example, correcting or confirming previously discussed information. Such an answer is inadequate in the given question context. If the 'matching'–'not matching' decisions are interpreted in terms of this pitch-accent dependent semantic-pragmatic compatibility, the following four conclusions can be drawn from the results.

First, the base stimulus 0 was in fact ambiguous as to the signalling of the early and medial pitch accents. Secondly, the shift of the pointed F0 peak to the right into the accented vowel of *„Lan-"* causes a change from early to medial pitch accent. In terms of the majority of judgments, this change occurs immediately after the F0-peak maximum and hence part of the rising movement is located inside the vowel. When the F0 peak is shifted to the left away from the vowel, pitch-accent identification becomes clear early. Thirdly, the change from the early to the medial pitch accent can also be brought about by extending the peak-maximum F0 level in the form of a high plateau into the accented vowel by moving only the fall to the right. That is, the rising F0 movement itself need not go into the accented vowel. But the fourth and most important conclusion is that the perceptual transition to the identification of the medial pitch accent takes place even when the extension of the F0-peak maximum in the form of a high plateau has gone in the opposite direction, i.e. away from the vowel. In this case, the accented vowel contained neither a rising nor a high-level, but solely a falling F0 movement.

10.3 AX discrimination experiment

Although the previous identification experiment is complemented by an AX discrimination experiment, it is important to see that this complementation does not aim at addressing the issue of categorical speech perception in the sense of Liberman et al. (1957) and Repp (1984). Rather, the discrimination experiment is used to assess pitch-accent identity from the direction of intonational form, in this way providing additional support for the conclusions of the indirect-identification experiment that were drawn from the direction of intonational meaning. Specifically, the AX discrimination experiment aims at demonstrating that those peak and plateau stimuli that approximated each other in meaning-based 'matching' judgments also approximate each other in form-based 'sameness' judgments and may hence be regarded as instances of the same intonational form-meaning relationship referred to as the medial pitch accent. Thus the focus of the AX discrimination experiment is not on category boundaries and related discrimination maxima but rather on within-category discrimination minima.

As for the latter, Niebuhr and Kohler (2004) argue with reference to theoretical aspects of perception and cognition in general (cf. Massaro, 1987, 1998) that, unlike identifications, the nature of discrimination judgments is not psychophonetic in the sense that they follow from the decoding of the speech code. Rather, discrimination judgments originate from a pre-linguistic, psychoacoustic level. Consequently, two stimuli that represent the same category — like in the present case the medial pitch accent — need not be perceptually indistinguishable, even though it is likely that their discriminability is lower than that of two stimuli with different pitch-accent categories. Therefore, the question whether two stimuli do or do not contain instances of the same pitch accent should not be based on the presence or absence of discriminability *per se*, but on relative changes in the level of discriminability. However, if it was found that two stimuli are in fact indistinguishable from one another, this would be a strong argument in favour of the same pitch accent, since it is impossible for two signals to convey different meanings if they cannot even be discriminated on an early psychoacoustic level.

10.3.1 Method

Stimulus pairings

The AX discrimination experiment includes two experimental series and a control series of physically identical stimuli. The two experimental series will be referred to as the *left-plateau* and the *right-plateau series*. In the left-plateau series, the element 'A' is constant and corresponds to the plateau stimulus 10L with the longest leftward extension up to about the [l] onset of *„Lan-“*. The 'X' elements were each of the 10 plateau stimuli 1R–10R, in which the F0 fall was successively shifted to the right, extending the plateau into the accented vowel of *„Lan-“*. So, the left-plateau series is made up of 10 AX pairs (10L+1R, 10L+2R, 10L+3R etc., cf. Fig. 10.2). It is shown in Figure 4 that the 10L stimulus of the plateau series yielded around 75% 'matching' judgments. The same holds approximately for the stimuli 3R and 4R of the plateau series. If this correspondence in 'matchin' judgments is due to the perception of qualitatively comparable instances of the same pitch accent, then there will be a discrimination minimum for the 10L+3R and 10L+4R stimulus pairs. That is, across all 10 AX pairs of the left-plateau series, the discrimination curve should take a V-like shape.

The right-plateau series relates a plateau stimulus to those peak stimuli that are known to be robust indicators of the medial pitch accent. This concerns the 10 peak stimuli that show an alignment in the accented vowel, i.e. 1R–10R. They function as the 'X' elements. The plateau stimulus 6R from the centre of the plateau extension into the accented vowel is taken as the constant 'A' element. Correspondingly, it is paired with each of the 10 peak stimuli 1R-10R, whereby the right-plateau series also consists of 10 AX pairs. If the plateau stimulus 6R resembles intonationally one or more of the peak stimuli 1R–10R, this would clearly support the previous conclusion that the high numbers of 'matching' judgments yielded by both peaks and plateaux result from medial pitch-accent identifications. In particular, if similar 'matching' levels of peak and plateau stimuli are an indicator of intonational resemblance, then the AX discrimination curve of the right-plateau series should have a clear minimum at the stimulus pair 6R+3R. Furthermore, the results of the right-plateau series can be related to the 'effectiveness hypothesis' (H3, cf. 10.1.4), which concerns the effectiveness of medial pitch-accent signalling by peaks and plateaux in the accented vowel. Finally, by sorting out the matter of the type of pitch accent that is cues by plateau stimulus 6R, the right-plateau series sheds further indirect light on the pitch accent conveyed by those plateaux that lead away from the accented vowel, since the plateau stimulus 6R is also part of the left-plateau series.

As a supplement to the left-plateau and the right-plateau AX series, some *physically identical* pairs were included in the discrimination experiment. They were composed of each of the 10 peak and the 10 plateaux stimuli with even stimulus numbers (i.e. 2, 4, 6, 8, 10 with <R> and <L> indices) as well as of the base stimulus 0. The resulting total number of 21 identical — or AA — pairs represented the control condition, and they served to counterbalance the expected number of 'same' and 'different' judgments in the experiment.

Subjects and Procedure

The general procedural framework of the discrimination experiment was identical with the framework of the indirect-identification experiment. That is, the stimulus pairs were presented several times in an overall randomized order that was organized in blocks of 10. However, in the discrimination experiment the number repeated presentations was reduced from five to three, as the presentation of all psychically different pairs was already done in both possible orders, AX and XA. So, the three randomized repetitions concerned 20+20=40 AX/XA pairs and another 21 physically identical AA pairs, which resulted in 61x3=183 pairs. The blocks of stimulus pairs were separated by double bleeps and pauses of 5 seconds. The stimulus pairs within each block separated by pauses of 4 seconds. Following the summary of Gerrits and Schouten (2004) on experimental experiences with comparisons of speech stimuli, the inter-stimulus interval was set to one second. Pairs were introduced by a short single bleep. Judgments were again made on prepared answer sheets that followed the block-like organization of the stimulus pairs.

A total of 31 undergraduate students with normal hearing from the Department of General and Comparative Linguistics at the University of Kiel (21 female, 10 male, average age 21.2 years) participated in the experiment. None of them took part in the previous indirect-identification experiment. The 31 subjects received the instruction to listen carefully to the paired presentations of the short statement utterances *„über Langeland“*, and to decide afterwards whether the two utterances were identical or different in terms of speech melody. The stimulus pairs were played over loudspeakers in a silent room at the University of Kiel.

Prior to the actual experiment, the subjects were asked to judge a training block consisting of the eight AX/XA pairs that involved the extreme peak and plateau stimuli 1R and 10R. Moreover, the training block included the physically identical AA pairs that were composed of 2R, 2L, 10R, 10L, or stimulus 0. As in the main discrimination experiment, the pairs of the training block were presented three times in an overall randomized order. Including the instructions and the training block, the entire AX discrimination experiment took about 40 minutes.

10.3.2 Results

The results of the discrimination experiment are illustrated in Fig. 10.5(a)–(b). Each data point represents 31*6=186 judgments. As already implied by this number, the judgments of the AX and XA pairs were pooled in the Figures and the statistical analyses. It can clearly be seen that the subjects' ability to discriminate the stimulus pairs changes across both the left-plateau (Fig. 10.5a) and the right-plateau (Fig. 10.5b) series. The discrimination curves of the left-plateau and right-plateau series have a pronounced V-like shape. So, the 'X' stimuli with the longest or shortest plateaux into the accented vowel or with F0 peaks aligned at the edges of the accented vowel can be better discriminated from the constant 'A' references 10L (left-plateau series, Fig. 10.5a) or 6R (right-plateau series, Fig. 10.5b) than the peak and plateau stimuli from the centre of the corresponding F0 continua. Accordingly, two separate χ^2-tests (with expected frequency = total number of 'different' judgments / 10) showed for each series that the 'different' judgments are not equally distributed across the stimulus pairs (left-plateau series: χ^2=126.699; df=9; p<0.0001***; right-plateau series: χ^2=112.961; df=9; p<0.0001***). The opposite holds for the 21 pairs of physically identical stimuli, i.e. AA. According to their χ^2-test, the 'different' judgments may be regarded as statistically equally distributed across the pairs. On average, the percentage of "false alarms", i.e. the frequency with which identical stimuli were judged erroneously as 'different', amounts to about 11%. This base level of false alarms is indicated in Fig. 10.5(a)–(b) by a grey dotted line.

Furthermore, sets of t-tests were calculated for the physically different stimulus pairs of the left-plateau and the right-plateau series. Separate t-tests compared for each pair whether the (relative) number of 'different' judgments is significantly above the base level of the physically identical pairs. At a significance level of p<0.05 this was true for all but three stimulus pairs. The three exceptions are 10L+3R and 10L+4R in the left-plateau series as well as 6R+3R in the right-plateau series. With p-values >0.25 the comparisons of the three AX pairs with the AA pairs miss the significance threshold very clearly. This supports that 10L+3R, 10L+4R (left-plateau series) and 6R+3R (right-plateau series) are perceptually as indistinguishable as the stimuli in the physically identical pairs. It is crucial to note that the indistinguishable stimuli within each of the three pairs are also similar in terms of their 'matching' judgments.

The close correspondence between the discriminability of the stimuli in the AX pairs and the 'matching' judgments they evoked is further supported by Pearson's correlations. They were calculated separately for the left-plateau and the right-plateau series and related the discrimination level of the stimuli in a pair to their difference in 'matching' levels. For example, the stimulus pair 10L+2R had a discrimination level of 53% (cf. Fig. 10.5a) and a difference in 'matching' levels of 89%-70%=19%. The following pair in the series, i.e. 10L+3R, had a discrimination level of 7% and a difference in 'matching' level of 73%-69%=4%. On this basis, significantly positive correlations resulted for the stimulus pairs in the left-plateau series (r=0.690; df=8; p<0.05*) and in the right-plateau series (r=0.612; df=8; p<0.05*). So, the worse the discrimination, the smaller the difference in 'matching' judgments was between the two stimuli. However, it is also important to see that — even though they are significant — the two positive correlations are relatively weak. This is primarily due to the fact that the stimuli with high stimulus numbers were easily discriminable, although they were all judged as 'matching' in the vast majority of cases. Apart from experimental artefacts like ceiling effects in the judgments towards the ends of the stimulus continua that

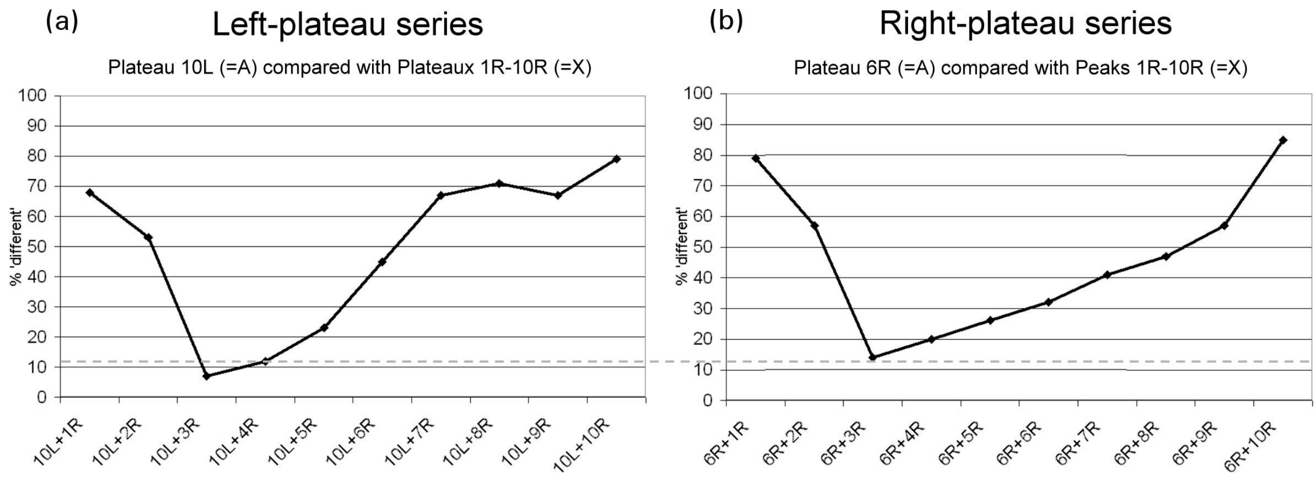

Fig. 10.5: Percentages for the discrimination of the 10 stimulus pairs in the left-plateau series (a) and in the right-plateau series (b). The grey dotted line refers to the mean "false alarm" (i.e. discrimination) level of the physically identical pairs in the discrimination experiment. Each data point represents 186 'same'–'different' decisions.

may come into play here, the correlations are consistent with the argument of Niebuhr and Kohler (2004). That is, two instances of the same (e.g., the medial) pitch accent may be less easily discriminable than instances of difference pitch accents (e.g., early and medial). But, comparing instances of the same pitch accent does not per se preclude discriminability.

Finally, an additional t-test was calculated based on the 40 physically different stimulus pairs of both the left-plateau and the right-plateau series. It compared the 'different' sums yielded by the AX pairs with the sums of the corresponding XA pairs for the 31 subjects. A clear and statistically highly significant effect of the order of presentation was found ($t=13.733$; $df=30$; $p<0.0001$***). The order effect is in line with those reported in previous studies on the perception of intonation. That is, discrimination performance is better when the physical change from the first to the second stimulus increases rather than decreases F0 (cf. Kohler, 1987; Chuang and Wang, 1978; Hellström et al., 1994; Ladd and Morton, 1997; Remijsen and van Heuven, 1999; Schneider et al., 2009; cf. also Cutting and Rosner, 1974; Gerrits and Schouten, 2004; Hwang et al., 2008). In the case of the present stimulus pairs, this means, for example, that the discrimination was better when the pitch accent in the second stimulus showed a longer high plateau and/or a later alignment. It is possible that this order-of-presentation effect is due to the fact that speech utterances are typically characterized by an F0 declination (cf. Maeda, 1974). Thus listeners may react more sensitively to the more marked event of a change towards a higher F0 (cf. also the 'production code' of Gussenhoven, 2002).

10.3.3 Conclusions

The results of the AX discrimination experiment demonstrate for the left-plateau series that shifting the alignment of F0 movements in opposite directions does not necessarily lead to intonational dissimilarity. Two stimuli can become intonationally similar to each other although their high F0 plateaux are extended across the accented vowel and away from the accented vowel across the preceding syllable-initial (lateral) sonorant. The stimulus 10L, in which the high plateau spanned the whole lateral and did not enter the accented vowel at all, can also be considered as indistinguishable from stimuli 3R and 4R, whose high plateaux covered around one third of the accented vowel. The increase in similarity that culminated in the *indistinguishability* of those stimuli which received also similarly high numbers of 'matching' judgments is a strong support for the conclusion that the diametrically opposed plateau extensions lead towards the same pitch accent.

As for the type of this pitch accent, the results of the right-plateau series show that plateaux extending into the accented vowel — represented by stimulus 6R — can also become similar to or even indistinguishable from pointed F0 peaks that are aligned within the accented vowel. It is known that F0 peaks with this alignment are a robust indicator of the medial pitch accent. Thus, the intonational similarity or *indistinguishability* of the 'medial' F0 peaks and the F0 plateau 6R confirms the interpretation that 'matching' judgments yielded by plateaux into the vowel are due to medial pitch-accent identifications. Moreover, since the plateau stimulus 6R was among those stimuli in the right-plateau series which could not be discriminated from the plateau stimulus 10L in the majority of cases, the results of the right-plateau series provide indirect support for the claim that 10L, in which the plateau leads away from the accented vowel, also indeed conveys the medial pitch accent.

10.4 General Discussion

10.4.1 Fitting the results into the empirical picture

The AX discrimination experiment supported the conclusions drawn from the matching judgments of the peak and plateau stimuli in the indirect-identification experiment. Taken together, this means that the hypotheses put forward in 10.1.4 were confirmed by the results of the present study. Firstly, in line with the 'peak hypothesis' (H1) the present study replicated the previous findings of Kohler (1987) and Niebuhr (2003, 2007a) for German. Shifting a pointed rising-falling F0 peak across the vowel onset of the accented syllable resulted in an abrupt perceptual change in intonational form and meaning. With reference to the predominant alignment model in Fig. 10.1 (cf. Kohler, 1987), the form-meaning relationships were denoted as early and medial pitch accents. In the AM model for German, the pitch accents are labelled as H+L* and L+H* (cf. Grice and Baumann, 2000). In terms of the majority of matching judgments the change from early to medial pitch accents took place immediately (i.e. 10 ms) after the F0 maximum and hence the rising slope entered the accented vowel, which is slightly earlier than in the studies of Kohler (1987) and Niebuhr (2003, 2007a), who located the boundary between early and medial 30–60 ms after the accented-vowel onset. This difference may be due to the fact that the acoustic transition from the syllable-initial lateral [l] to the vowel [a̲] in „*Lan-*" is less abrupt in the present stimuli than in the stimuli used by Kohler and Niebuhr, particularly in terms of the increase in intensity (i.e. short-term energy) around the CV boundary of the accented syllable. In the stimuli of Kohler and Niebuhr, there was a steep increase in intensity after the vowel onset, whereas in the present stimuli the increase in intensity was shallower and set in already in the syllable-initial [l] (cf. Fig. 10.2; cf. also 10.4.2).

Furthermore, the plateau series confirmed the 'plateau hypothesis' (H2), according to which both the extension of the F0-peak maximum from the accented-vowel onset to the right *and* to the left, i.e. into *and* away from the accented vowel, causes a change in the pitch-accent identification towards medial accent type. Firstly, this means that an F0 rise into the vowel is not a prerequisite for perceiving medial pitch accents. More importantly, however, the change towards medial for the *leftward* plateau extension is completely incompatible with the alignment-based model of pitch-accent perception and hence with the original temporal implications of 'early' and 'medial'. This issue is discussed in more detail in 10.4.2. The change from early to medial brought about by the rightward plateau extension supports the argumentation of Ambrazaitis & Niebuhr (2008). They claim with reference to Kohler (1991) that the contrast between the early and medial type of pitch accent is maintained at the ends of hat patterns, where the pitch accent is concatenated with the previous accent by a high F0 plateau and hence does not show a separate F0 rise. Additional phonetic details of the hat pattern final pitch-accent contrast are discussed in 10.4.3.

As regards the power of peak and plateau stimuli in signalling the medial pitch accent, both the matching and the discrimination results support the 'effectiveness hypothesis' (H3). Plateaux leading to the left, i.e. away from the accented vowel, are worse in triggering the medial pitch accent than plateaux leading to the right, i.e. into the accented vowel. In order to be perceptually equivalent to a plateau that extends over just 30 ms into the accented vowel, a plateau that reaches away from the vowel must extend over 100ms and span the entire syllable-initial lateral. A more complex picture emerged for the comparison of peak and plateau shifts into the vowel. D'Imperio (2000) found for Neapolitan Italian that a rising-falling F0 peak had to be shifted with its rise further into the vowel than a rising-falling plateau in order yield a similar level of medial pitch-accent perceptions (cf. also D'Imperio et al., 2010). This finding was replicated by Barnes et al. (submitted) for American English. Against this backdrop, hypothesis (H3) predicted that plateaux extended into the vowel are more powerful in cueing the medial pitch accent than peaks shifted into the vowel. Simply, the present findings neither support nor contradict this prediction. Overall, the corresponding t test lacked significance. However, the non-significant outcome is due to a crossing over of the increasing medial pitch-accent identifications yielded by the peak and plateau stimuli. While the medial pitch-accent identifications in the second half of the peak-shift continuum into the vowel stagnated at around 85%, they increase further up to almost 100% in the corresponding second half of the plateau extension into the vowel (cf. Fig. 10.3 vs. Fig. 10.4). This finding supports hypotheses (H3). In contrast, the results received for the first half of the peak and plateau shifts into the vowel are inconsistent with hypothesis (H3). Compared with short F0 plateaux into the vowel, F0 peaks aligned shortly after the accented-vowel onset triggered clearly more medial pitch accents.

This inconsistency with hypothesis (H3) is likely due to the fact that the plateaux in the present stimuli were not preceded by a rise into the accented vowel, whereas this rise was there in the peak stimuli; and it was also there in the plateau stimuli of D'Imperio (2000) and Barnes et al. (submitted), from which (H3) was derived. So, a rise into the accented vowel seems to be a more powerful cue for the medial pitch accent than a small amount of high F0 in the vowel. But when the "high-F0 content" within a rising-falling contour is considerably increased by means of a long plateau, it outweighs the rise in terms of cueing the medial pitch accent. In summary, this refined hypothesis must be tested in a subsequent experiment, while the original hypothesis (H3) can partly be accepted. Initial support for the refined hypotheses comes from the recent study of Barnes et al. (submitted). They shifted entire F0 peak and plateau contours with identically aligned rises into the vowel and found that, early in the vowel, peaks and plateau do equally well in cueing the medial pitch accent. But, later in the vowel, plateaux become better medial pitch-accent cues than peaks, which is in line with the present findings.

Finally, in view the exact phonological meanings of H+L* and L+H*, the experimental findings on all three hypotheses demonstrated clearly that the signalling of the early and medial pitch accents (in German) cannot be conceptualized with exclusive reference to the alignment of either the rising or the falling slope. If this was the case, then the stimuli with the identical numbers in the left (<L>) and right (<R>) branches of the peak and plateau series would have evoked equal 'matching' responses. However, that an F0 fall triggers not only the corresponding H+L* accent, but also the L+*H accent without an alignment change would not be expected in the AM framework. Likewise, it would not be expected that L+H* perceptions can be turned into H+L* perceptions without changing the alignment of the rising F0 movement. This point is also emphasized by Barnes et al. (submitted). The implications of this and other points for modelling pitch-accent perception are outlined in the following section.

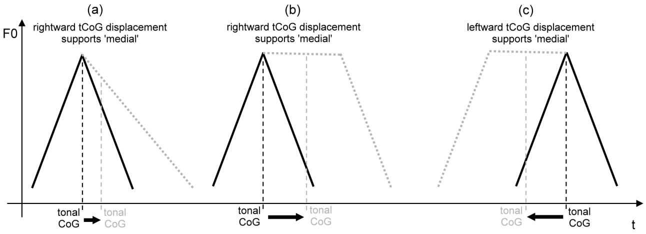

Fig. 10.6: Changes in the slope of the falling F0 movements and in the extension of the F0-peak maximum expressed as change in the temporal alignment of the tonal centre of gravity (tCoG).

10.4.2 Towards a new model of pitch-accent perception

The perception of the early vs. medial pitch-accent contrast is traditionally associated with an alignment contrast that refers to the temporal position of the F0-peak maximum. The alignment range of the medial pitch accent follows the range of the early accent. The perceptual boundary between the two alignment ranges is located around the accented-vowel onset. Thus, any F0-peak shift to the left is supposed to change the pitch-accent identification towards the early pitch accent, whereas a change in the identification towards the medial accent should require that the F0 peak is moved to the right.

On the basis of this alignment model, it poses a problem that changing the slopes of the F0 peak can affect pitch-accent identification, despite a constant position of the peak maximum. For example, Niebuhr (2003, 2007a) demonstrated for German that a shallower fall supports the perception of the medial pitch accent (cf. also Gósy and Terken, 1994; D'Imperio, 2000 for other languages). Addressing evidence like this, Barnes et al. (2010) detached the alignment model from concrete F0 turning points and introduced a more abstract and perception-driven value: the *Tonal Centre of Gravity* (TCoG). The TCoG was, among others, inspired by the 'pitch integral' that Nolan and Segerup (2005) developed in order to account for the interplay of timing and height of F0 falls in the production and perception of the word-accent distinction in Gothenburg Swedish. The TCoG of Barnes et al. (2010) is the weighted average of the F0 values within the pitch-accent related peak, projected onto the time axis. The points in time that constitute the peak are summed up, whereby each point is multiplied with its corresponding F0 value. The total sum is then divided by the number of points to get the average. In this way, TCoG is able to translate dispositions of the "bulk of mass" within the F0 peak, which are, for example, caused by the F0-peak shape, to changes in the temporal alignment of the TCoG. These TCoG alignments are then again in line with the original link between the direction of the alignment change and pitch-accent identification. For example, Fig. 10.6(a) shows that a shallower fall shifts the TCoG to the right. This matches with a perceptual support of the medial pitch accent in the original alignment model.

Modelling pitch accent perception in terms of TCoG alignment is also able to account for the present finding that the extension of the high F0 plateau into the accented vowel caused a change from early to medial pitch accent. As in the case of a shallower fall, it can be seen in Fig. 10.6(b) that the extension of the plateau into the vowel shifts the TCoG to the right, and this accords with a change towards medial pitch-accent identification in the original alignment model. However, like the original alignment model, the modified TCoG-based alignment model is unable to explain, why also an extension of the high F0 plateau in the opposite direction, i.e. away from the accented vowel, changed the pitch-accent perception towards medial. As is illustrated in Fig. 10.6(c), extending the plateau away from the vowel inevitably pulls the TCoG to the *left* and hence in the wrong direction (that should lead towards the early pitch accent).

Niebuhr and Pfitzinger (2010) have pointed out that the TCoG approach cannot account for effects on pitch-accent identification due to changes in peak height or symmetrical changes in peak shape (e.g. an equal flattening of the rising and falling slope) as they leave the alignment of the TCoG constant. The latter holds even more for changes that go beyond F0. For example, continuing previous studies of Niebuhr (2006, 2007b), Niebuhr and Pfitzinger (2010) showed that changes in the intensity course of the segmental string that underlies the F0 peak contour also affect pitch-accent identification.

The failure of alignment models to cope with the entirety of F0 and non-F0 effects on pitch-accent identification suggests that pitch-accent perception should be modelled on the basis of an *atemporal* approach. Within the more general framework of the *contrast theory of the perception of speech melody*, which was developed in analogy to visual perception (particularly in terms of contrast coding, contrast enhancement and Gestalt rules), such an atemporal approach was provided by Niebuhr (2007c). Going beyond the TCoG approach, it relocates the characteristic features of pitch accents from the acoustic to the perceptual domain by representing pitch accents in terms of two Gestalt-like patterns. The first Gestalt is the *perceived pitch pattern*. It is derived from breaking down the rising and/or falling F0 slopes of pitch-accent peaks into a string of tonal events. The number of tonal events within the string can change due to the durations of the F0 slopes and the listeners' ability to detect and to follow pitch movements along with spectral changes in the segmental string (cf. 't Hart et al., 1990; House, 1990; Bregman 1990). For example, if an F0 slope is longer and/or extends over more speech segments, the number of tonal events that the slope generates in perception will increase. The tonal events themselves can be both static and dynamic. However, irrespective of the number and type of tonal events that a pitch Gestalt includes, it can remain, for example, overall rising-falling.

Furthermore, each tonal event within the pitch Gestalt has a separate perceptual prominence. The individual prominences together constitute the second Gestalt-like pattern, the *perceived prominence pattern*. The prominence of a tonal event depends (a) on whether it is associated with a static or a dynamic tonal event and (b) on how strongly it contrasts syntagmatically with the preceding tonal event. The magnitude of the contrast of a given tonal event with the preceding event is, amongst others, determined by the direction and the size of the pitch change. Additionally, the durations and intensity levels of the tonal events are taken into account (cf. Fry, 1958 and Kohler, 2008 for the relationships between duration, intensity, and perceptual prominence); and ultimately the magnitude of the syntagmatic contrast decreases with an increasing temporal distance between the tonal events. The prominence of a syllable, including the accented syllable, is the sum of the prominences of all tonal events it coincides with (which results in a positive correlation between number of tonal events, i.e. pitch changes, and syllable prominence). Within a prominence Gestalt, the prominences across the individual tonal events can remain identical or they can increase or decrease gradually. However, if a change in prominence becomes too strong (which is defined in terms of the Gestalt rule of "good continuation"), an additional syllable will be perceived. A single abrupt switch from a low to a high pitch level or vice versa can be sufficient to perceive an otherwise constant vowel as disyllabic.

For explaining the present findings with the outlined atemporal approach, we can reduce the dynamic complexity in the Gestalt patterns. As regards the pitch Gestalt, it is sufficient to simply think of the rising-falling peak and plateau contours as tripartite LHL patterns. Hence, the difference between the early and medial pitch accents lies in the prominence patterns. For 'early', the prominence levels increase from the initial to the final L (LHL), whereas 'medial' is identified when the H stands out in perception against the surrounding Ls (LHL). Shifting an F0 peak, i.e. the LHL pitch pattern, with its maximum into the high-intensity area of the accented vowel will rapidly increase the prominence level of the H. At the same time, the prominence level of the final L will rapidly decrease, since the F0-peak shift pushes the falling movement successively beyond the offset of the accented vowel. The resulting inversion of the H and L prominence levels explains, why the peak series of the present study replicated previous cross-linguistic findings by causing a quick perceptual change from early to medial pitch accent after the F0 maximum entered and the fall offset left the accented vowel (cf. Fig. 10.2). However, as was also noted in 10.4.1, the actual abruptness of the perceptual change from early to medial, i.e. from LHL to LHL, depends on how steeply the intensity increases and decreases at the vowel boundaries (cf. Niebuhr, 2007b).

As regards the results of the plateau stimuli, the prominence effect of duration must be additionally taken into account. That is, extending the duration of the H in the form of F0 plateaux successively increases the prominence level of this tonal event. At a certain point, the H prominence outweighs the prominences of the surrounding Ls, irrespective of plateau alignment. This explains not only why a perceptual change from early to medial pitch accent occurred for the both the plateau extensions to the right and to the left but also why the plateau extension to the left was less effective in signalling the medial pitch accent than the plateau extension to the right. While both the rightward and the leftward plateau extensions make the H event more prominent relative to the Ls, the H prominences of the rightward extended plateaux benefit additionally from the higher intensity level of the accented vowel, compared with the lower intensity level of the preceding lateral [l].

The fact that, right after the vowel onset, pointed peaks were better triggers of the medial pitch accent than short plateaux can be accounted for in the contrast theory in terms of the inherently greater prominence of dynamic tonal events compared with static tonal events. The crucial difference between the short plateaux and the pointed peaks was that the pointed peaks rose into the vowel; and as the vowel was a spectrally stable sound section (cf. House, 1990), the F0 rise also manifested itself in perception and gave the H part of the pitch Gestalt a greater prominence than the static H event that was derived from the short F0 plateau into the vowel. However, as is implied in the plateau-related explanations, due to the combined contribution of duration and intensity to the prominence of the static H event, *long* high plateaux across the vowel can become more powerful triggers of the medial pitch accent than pointed peaks that rise into the vowel (whereas these pointed peaks remained more powerful than long high plateaux across low-intensity consonants).

It is worth pointing out that the explanations above fit in well with the perceptual impressions of Welby and Loevenbruck (2006), according to which, a long plateau across the vowel enhances the high pitch target in terms of perceptual salience compared with a pointed peak. This impression is supported by experiments of Knight (2008), who showed that long plateau contours are overall perceived as higher and more prominent than pointed peak contours. Beyond the scope of the present findings, it can also be understood in the outlined theoretical framework, why variations in the durations and the types of consonants and vowels in the accented syllable have (cross-linguistically similar) effects on the alignments of the pitch-accent F0 movements (cf. Silverman and Pierrehumbert, 1990; Gartenberg and Panzlaff-Reuter, 1991; van Santen and Hirschberg, 1994; Grabe, 1998; Ladd et al., 2000). The alignment effects are adaptations to the altered prominence conditions provided by the segmental string. For example, it is well known that F0 peaks of medial pitch accents (or H*) are aligned more central in the accented vowel for phonologically short than for phonologically long vowels. Compared with long vowels, short vowels have an overall lower intensity level and are delimited by steeper intensity changes. This reduced potential to increase the prominence of the H event beyond the prominences of the adjacent L events is compensated for by moving the F0 peak, which is a major part of the H event, away from the steep intensity decrease at the vowel offset and towards the intensity peak in the centre of the vowel. Ultimately, the presence of a vowel is not even necessary for distinguishing pitch accents in the outlined theoretical framework. Clearly defined intensity courses are sufficient, which matches with the fact that pitch accents can be distinguished without any problems if they occur on single long vowels or nasals like [ɔː] or [mː]. Under such conditions, the concept of alignment is not even applicable. For more details on and examples of the contrast theory of the perception of speech melody the reader is referred to Niebuhr (2007c).

Finally, it should be emphasized that the term 'atemporal' does not mean that the contrast theory of the perception of speech melody by Niebuhr (2007c) completely disregards temporal aspects. In fact, quite the opposite is true. Pitch and prominence Gestalts do not only unfold over time, the theoretical framework also includes anticipatory top-down effects. That is, pitch and prominence levels of following tonal events are to some extent shaped by the levels of the previous tonal events. In the opposite direction, due to syntagmatic contrasts and contrast enhancements new tonal events continuously cause backward modifications of the pitch and prominence levels of the preceding tonal events. One of the consequences of these forward and backward modifications is that the perceived locations and types of pitch accents in a stretch of utterance can change drastically depending on whether this stretch is presented to listeners in isolation or embedded in the surrounding utterance context (experienced prosodic labellers surely know this phenomenon). As regards the pitch accents themselves, 'atemporal' just means that acoustically defined alignment patterns are not the essential characteristics of pitch accents. Rather, the essential characteristics of pitch accents are perceptually defined pitch and prominence Gestalts. But, the most efficient strategy to create a particular prominence Gestalt for a given pitch Gestalt is to coordinate the tonal events with the intrinsically given intensity properties of the underlying segmental string. In terms of this temporal coordination, the outlined atemporal approach still acknowledges the major role of alignment in the production and perception of pitch accents. However, the approach also leaves room for other (partly speaker-specific) strategies to achieve the same perceptual goal. For example, speakers additionally adapt the segmental intensity levels in and around the accented syllable to the type of pitch accent (cf. Niebuhr and Pfitzinger, 2010) and some speakers are more "shapers" or "rangers" than "aligners" in the sense that they primarily vary the shape or the range of the F0 peak instead of peak-maximum alignment, see Nolan and Segerup (2005) for the speaker-specific use of range and alignment in distinguishing word accent 1 and 2 in Gothenburg Swedish.

10.4.3 The early vs. medial pitch-accent contrast at the end of hat patterns

The findings of the present study showed contrary to the introspection of Kohler (1991, cf. 10.1.3) that high F0 plateaux, which reach into, but do not rise into the accented vowel, can be proper indicators of the medial pitch accent. This holds most clearly for high plateaux that span more or less the entire accented vowel. However, high plateaux were also capable of conveying the medial pitch accent when they were extended in the opposite direction over the preceding (sonorant) consonant. Taken together, the present findings have the following implications for the realization of the contrast between early and medial pitch accents at the ends of hat patterns. In the case of a medial pitch accent at the end of a hat pattern the onset of the F0 fall should be located after the accented-vowel offset so that the high level F0 goes across the whole vowel; and in order to convey an early pitch accent at the end of a hat pattern, the F0 plateau should not reach far into the (sonorant) consonant preceding the accented vowel. Rather, the F0 fall should set in before the (sonorant) consonant.

Based on these implications, a small sample of hat patterns was extracted from the Kiel Corpus of Spontaneous Speech. The sample consisted of 20 hat patterns; 10 of them ended in early pitch accents, the other 10 in medial pitch accents, according to the classifications of Peters et al. (2005). The hat pattern sample was analyzed acoustically with regard to the fall-onset alignment of the second pitch accent (disregarding microprosodic perturbations, cf. Kohler, 1990). The sample was restricted to phrase-final hat patterns that were composed of two pitch-accented syllables (i.e. hat patterns containing additional accents in the middle of the plateau were excluded). Moreover, as the phonological structure of the accented syllable is known to affect alignment (cf. Gartenberg and Panzlaff-Reuter, 1991; Ladd et al., 2000; van Santen and Hirschberg, 1994), and in order to avoid alignment effects due to tonal crowding (cf. Caspers and van Heuven, 1993), the two accented syllables had to be separated by at least one unaccented syllable, and the second (i.e. the target) syllable had to have a voiced rhyme with a phonologically short vowel. The analyses showed that the differences in fall-onset alignment between hat-pattern final early and medial pitch accents match well with the implications of the present study. That is, in 9 of the 10 hat patterns ending in early pitch accents (i.e. 90%) the corresponding F0 fall set in well before the (sonorant) consonant of the ac-

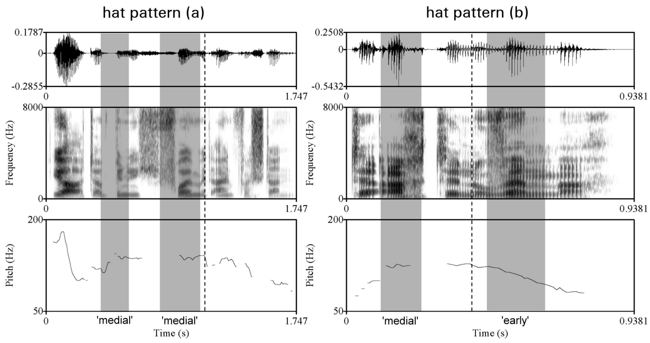

Fig. 10.7: Oscillograms (top), spectrograms (middle), and F0 contours (bottom) of the two utterances *„Ja, da bin ich voll mit einverstanden“* ('Yes, I completely agree with this', left) and *„Der achte November“* ('the eighth of November', right, accented syllables underlined), produced naturally by male speakers (g071a) in the 'Kiel Corpus of Spontaneous Speech'. Each utterance contains a phrase-final, two-accent hat pattern. Hat pattern (a) consists of two medial pitch accents, whereas hat (b) represents a 'medial'–'early' pitch-accent sequence. Accented syllables are shaded in grey. Dotted lines refer to the onsets of the F0 fall which are aligned around the end or beginning of the accented word.

cented syllable. In contrast, in the hat patterns that ended in medial pitch accents the final fall was delayed until after the accented vowel. Fig. 10.7(a)–(b) illustrate the two fall-onset alignments based a pair of examples from the Kiel Corpus of Spontaneous Speech.

Further correspondences between the present perception results and real-speech alignment data show up in comparisons of peaks and plateaux. For example, production studies found that if the early pitch accent is realized as a pointed F0 peak, then the transition into the fall (i.e. the peak maximum) occurs typically immediately before the accented-vowel onset (cf. Gartenberg and Panzlaff-Reuter, 1991; Rathcke and Harrington, 2006, 2007) and hence clearly later than in the plateau realizations of the early pitch accent at the ends of the analyzed hat patterns. This matches with the fact that — unlike plateaux — pointed peaks were also clearly identified as the early pitch accent when they occurred in the syllable-initial [l]. In the case of the medial pitch accent, the production data of Gartenberg and Panzlaff-Reuter (1991) shows that pointed peak realizations have their F0 maxima about 20–60% before the accented-vowel offset, whereas the plateau realizations of the medial pitch accent in the analyzed hat-pattern sample all started falling well after the accented-vowel offset. This is in line with the present perceptual findings that the high F0 levels of the plateaux had to be extended further into the vowel than pointed F0 peaks in order to become clear triggers of the medial pitch accent.

10.5 Outlook

The present findings suggest that extending the F0-peak maximum in the form of plateaux dovetails with the interplay of F0 parameters that relate the amplification of "high-F0 content" to the perceptual support of the medial pitch accent. Consequently, one obvious task of follow-up studies in the domains of perception as well as production will be to analyze the interactions of the peak vs. plateau contrast with variations in the shape, duration, and range of the rising and falling pitch-accent movements in more detail. For example, it is possible that the alignment difference of the final F0 fall that conveys the contrast between early and medial at the end of hat patterns (cf. Fig. 10.7) interacts with the F0 range of the hat. In view of the fact that the F0 range is lowered and narrowed in the course of utterances due to declination (cf. Maeda, 1974), maintaining the early vs. medial pitch-accent contrast could require different fall-onset alignments in utterance-initial than in utterance-final hat patterns (note that similar alignment differences are well known for pointed F0 peaks in prenuclear and nuclear position, cf. Grabe, 1998). It seems even questionable that a difference in fall-onset alignment (alone) is able to preserve the difference between early and medial when F0 levels and ranges of the hat patterns fall below a certain threshold. Kohler (1991:326) points out in this context that if hat patterns "*reached a fairly low F0 [. . .] then a final 'medial peak' fall is [. . .] something much closer to an 'early peak'* ". For the opposite end of hat patterns, it should be investigated whether hat pattern initial medial pitch accents, in which the rise is followed by a plateau instead of fall, show different rise-offset alignments than medial pitch accents that are concatenated with the following accent by a dip, and that are hence represented by a separate pointed peak.

Furthermore, beyond plateaux in hat patterns, follow-up studies must deal with the use of plateaux in the realization of individual pitch accents, again with respect to the concomitant variation in other F0 parameters. From a perceptual point of view, this means continuing the line of research by D'Imperio (2000), Gili Fivela and D'Imperio (2010), D'Imperio et al. (2010) and Barnes et al. (submitted). Instead of extending plateaux from a constant point in time into and away from the vowel, entire rising-falling contours with varied F0 ranges, slopes, and plateau durations should be shifted in either direction across the accented syllable. Moreover, as in the study of Barnes et al., the plateau research must take also other types of pitch accents than early and medial into account.

Finally, the whole research on the role of plateaux in signalling pitch accents — including, and going beyond, early and medial types — must be done on a comparative, cross-linguistic basis, taking into account that multiparametric cueing of meaning-related speech units typically goes along with language-specific trading relations between the individual cues (for instance, cf. the studies of Koreman et al., 2008 and Abramson & Tingsabadh (1999) on perceptual prominence or the identification of voicing contrasts). For example, the use of plateaux in coding types of pitch accents could be different in languages that show an extensive use of hat patterns than in languages that use hat patterns only under very restricted conditions or that do not know them at all.

In this holistic research perspective the experiments must start from an atemporal perceptual model like the one that is contained in the contrast theory of the perception of speech melody (cf. Niebuhr, 2007c). At the end of their plateaux study, Barnes et al. (submitted) outline a research perspective that would bring the TCoG model and the contrast theory closer together in terms of perceptually defined tones and the interrelation of F0 and segments (or their intensities). Specifically, Barnes et al. acknowledge that the problem of the TCoG *"lies in the very notion of a single 'peak analog' point [...] that can be derived without referring to the segmental string upon the tones in question are realized."*

10.6 Acknowledgements

First of all, I would like to thank Mariapaola D'Imperio, Francesco Cangemi, Barbara Gili Fivela, Jonathan Barnes, Gilbert Ambrazaitis, Rachael-Anne Knight and Klaus J. Kohler for many fruitful and inspiring discussions on pitch accent perception and the intonational contrast between peaks and plateaux. Furthermore, I am greatly indebted to Francis Nolan and Jan P. H. van Santen for their positive, very constructive comments on an earlier draft of this paper. Finally, thanks are due to Hartmut Pfitzinger for providing the experimental facilities and to Declan Donaghey for his thoruough proofreading of earlier drafts of this paper.

References

Abramson, A. S. & Tingsabadh, K. (1999). Thai final stops: Cross-language perception. *Phonetica 56*, 111–122.

Adriaens, L. M. H. (1991). *Ein Modell deutscher Intonation. Eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text*. Ph. D. thesis, Technical University Eindhoven.

Ambrazaitis, G. (2009). Nuclear intonation in Swedish — evidence from experimental-phonetic studies and a comparison with German. Travaux de l'Institut de linguistique de Lund 49, Lund, 1–225.

Ambrazaitis, G. I. & Niebuhr, O. (2008). Dip and hat pattern: a phonological contrast of German? In: *Proc. of the 4th Int. Conf. of Speech Prosody*. Campinas, Brazil, 269–272.

Andreeva, B. & Barry, W. J. (1999). Intonation von Checks in der Sofia-Varietät des Bulgarischen. Forschungsbericht (Phonus) 4, Inst. of Phonetics, Univ. of the Saarland, 1–13.

Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel, S. (2010). The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. In: *Proc. of the 5th Int. Conf. on Speech Prosody*. Chicago, USA.

Barnes, J., Veilleux, N. & Shattuck-Hufnagel, S. (2011). On the nature of perceptual differences between accentual peaks and plateaux. In: Niebuhr, O. & Pfitzinger, H. R. (Eds.), *Prosodies — Function, Context, Communication*. Berlin, New York: de Gruyter. *Submitted*.

Baumann, S. (2006). *The intonation of givenness. Evidence from German*. Tübingen: Max Niemeyer Verlag.

Boersma, P. (2001). Speak and unSpeak with PRAAT. *Glot International 5*(9/10), 341–347.

Bregman, A. S. (1990). *Auditory scene analysis. The perceptual organisation of sound*. Cambridge; Massachusetts, London: MIT Press.