Klaus J. Kohler

Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics
1. Introduction

1.1 Hypotheses

This contribution deals with Hypotheses (2) and (3) outlined in 2.1.2 and 2.1.3 of Contribution I (Kohler, 1991b), i.e. with the alignment of an F0 peak relative to stressed vowel onset in terminal utterances containing one accent. Section 2. is concerned with the F0 peak positions in sentences that have a unique accent placement because they are made up of just one content word beside several reduced function words. Section 3. looks at F0 peaks in sentences with alternate accent places due to lexical stress oppositions or to different sentence focus. It delimits the stress and intonation functions of F0 peaks and discusses their interactions, also with reference to the data presented in Contribution IV (Hertrich, 1991a). As this will involve the perceptual ambiguity between one and two accents, peak sequences will also have to be considered briefly with reference to Contribution VI (Hertrich, 1991b).

1.2 Types of phonological structures for perceptual testing

The investigation is perceptual, aiming at the (phonological) categorization of phonetic F0 peak shift continua across a number of different syllable structures (long vs. short vowel, syllable-initial lateral vs. glide vs. glottal stop (creaky voice) vs. voiced fricative, post-nuclear voiced vs. voiceless consonant) as well as two potential accent positions in words (prefix or stem stress) and sentences (subject or verb focus).

1.3 Stimulus generation

In all cases, several naturally produced tokens of the particular sentence type under scrutiny were recorded on analogue tape (Revox A77, 19cm/s) by the same male speaker (KK, the author) under studio conditions in the Kiel Phonetics Institute. Although a medial peak position was to be the basis for stimulus manipulation in most experiments (but see 3.1 for the choice of an early peak as well), early and late peaks were also collected of each linguistic item to specify the ranges of F0 peaks from early to late that would have to be covered by the test series, and in order to provide information about the shapes of the different peaks to be taken into account.

---

1 The stimuli for 2.1.1.2-5 and 3.1-2 were generated by Michael Weinhold.
in the synthesis. The recorded data were checked auditorily for successful rendering of the intended phonetic structures, and, after A/D conversion (10 kHz, 5 kHz low-pass filter), the acceptable tokens were processed on a Data General Eclipse S230 computer with the Kiel Phonetics Institute SSP programme package (as regards the pitch algorithm, see Schäfer-Vincent, 1982, 1983). Obvious FO analysis errors (octave jumps, missing FO values in spite of clear periodicity in the signal) were corrected manually.

Then one token containing an auditorily classified medial (or early) peak was selected and its peak contour shifted along the time axis to the left and to the right in a number of steps of fixed duration determined separately for each utterance, to create new FO versions. The shift was effected either as a parallel transposition of both branches of the peak contour, or the falling branch was time-expanded in left shifts as far as the original right-hand base point, to approximate natural productions by a less steep descent and to avoid too long a low FO stretch in the LPC synthesis. The two types of left shift do not alter the basic characteristics of medial to early peak changes; the parallel transposition of the whole peak configuration simply sounds more final and categorical than the one with the flattened fall. After the shift, the tail contour was joined to the new peak position by expansion or compression, similarly the immediate precursor, and finally FO was masked in voiceless stretches. Fig. 1 illustrates the principles of generating FO peak shift versions.

The original utterances were then synthesized with the LPC analysis values and the new FO versions obtained through the peak shift parameter manipulation.

1.4. Perception experiments
Two types of discrimination and of identification tests were performed:
(1) A quick serial discrimination test, in which listeners were presented with the ordered series of peak shift stimuli from left to right or right to left and asked four questions on prepared answer sheets; for each question they heard the series at least once.
(a) Do you perceive any changes in the melody of the sentence from one stimulus to the next?
   No - one change - several changes.
Fig. 1

(a) Speech wave and fundamental frequency (linear scale) of a medial peak in the naturally produced utterance "Sie hat ja gelogen." ("She's been lying."). The end contour (on the syllable gen) was added by FO parameter manipulation because the analysis did not provide it. The time marks A₁, A₂ delimit the FO peak contour (coinciding approximately with /oː/), which was shifted left and right.

(b) The left- and right-most positions of the shifted FO peak contour on the same time scale as in (a), approximating the natural productions of early and late peaks, respectively.
(b) At which stimulus in the series has the first change occurred?
   Encircle the relevant number.
(c) At which stimuli in the series have further changes occurred?
   Encircle the relevant numbers.
(d) What are the meanings of the original utterance and of utterances
    representing the first and further changes in the series?

The test tape construction had the following format:
200-ms bleep
800-ms pause
stimulus 1 (or n)
3-s pause
stimulus 2 (or n - 1)
3-s pause
   
   
stimulus n (or 1).

(2) A formal randomized AX or XA discrimination test, in which all the pairs
of one or two-step differences, as well as of identical stimuli
(restricted to uneven rank to limit the test size), from the ordered peak shift series were presented for 'same/different' judgements on
prepared answer sheets. Two test tapes were compiled, one for the ascending and one for the descending order of arrangement of stimuli
within the pairs, and each containing a randomization of 2 repetitions
of all the different as well as the identical stimulus pairs, with the
following general format:
200-ms bleep
800-ms pause
stimulus A (or X)
2-s pause
stimulus X (or A)
4-s pause
and so on for all the stimulus pairs. After each block of 10 test items
a further 500-ms bleep was added for orientation.

(3) A natural stimuli identification test, in which three different
naturally spoken contexts were paired with sentences containing each of three naturally produced peak positions – early, medial, late – for subjects to judge, on prepared answer sheets, whether the context and (the melody of) the test item matched or not. The test tapes were compiled in a short version of 9 items (3 contexts x 3 peaks) and a long one of 90 items, with 10 repetitions of each of the 9 items. In each tape the stimuli were randomized and followed the same format as in the randomized discrimination test, with the only difference that the pause between context and test stimulus was 0.5 s.

(4) A synthesized stimuli identification test, in which one synthesized context sentence was paired with each stimulus from an early to medial FO peak shift series, for subjects to judge, on prepared answer sheets, whether context and test item matched or not. The test tape contained a randomization of 10 repetitions of each context and test stimulus combination, following the same format as in the natural stimuli identification test.

The test files were compiled on the computer and output on analogue tape. The listening tests (except those in 2.1.2 and 2.1.3; see the separate descriptions there) took place in the acoustically treated studio of the Kiel Phonetics Institute. The stimuli were presented via loudspeaker to variably sized groups of up to 8 persons, who were students of a variety of subjects including phonetics/linguistics/languages, as well as members of academic and technical staff, and "naive" outsiders, all with German of a northern variety as their native language (except for 2.1.2 and 2.1.3; see the separate descriptions there).^1

1.5 Interactive perceptual testing at the computer
The development of an intonation model for German and its RULSYS TTS implementation (see Contributions I and VII; Kohler, 1991b, d) have made it possible to check the perceptual relevance of certain changes in FO configurations very quickly by generating parametric displays and acoustic output from orthographic input (supplemented by additional symbolic markers,

^1 Michael Weinhold put together the test tapes, carried out the tests, and compiled the data, for 2.1.1.2-5 and 3.1-2.
such as @ZZ for early or @ZZZ for late peaks) and by modifying the acoustic output interactively through systematic changes in the graphic parameter representation. This can be achieved in two ways:

(a) In a graphic display of the type illustrated in Fig. 2, F0 points are moved, inserted, deleted, or changed in value, and the speech signal is regenerated with the new parameter specification for auditory evaluation, also for auditory comparison with the stored original.

(b) A pitch configuration is defined by the use of the free variables X and Y (for time and frequency) as, for example, in the rule

00.01: <VOK,FSTRESS,TERMIN> ^ <TFO=TFO+(X-100)/2.5,T2FO=T2FO+(X-100)/2.5, T3FO=T3FO+(X-100)/2.5,2F0=Y>,

which means that a (medial) peak pattern <TERMIN> associated with an accented vowel (VOK,FSTRESS> and defined by three F0 points with the time values TFO,T2FO and T3FO is to be displaced in time by adding or subtracting the same variable time value X, and/or vertically expanded or compressed by varying the frequency value of the centre F0 point (2F0). An orthographic input is then processed by the system up to this rule, when an X-Y plane as shown in Fig. 3 appears on the screen, representing 250 time frames of 10 ms along the horizontal and 250 units of 1 Hz along the vertical. A cursor can now be moved, e.g. in 5-unit steps, to feed the variables X and Y in rule 00.01 with new values for further processing. In rule 00.01, the additive time constant of -100 resets the zero point, and the factor of 1/2.5 rescales the temporal step size from 5 x 10 ms to 5 x 10/2.5 ms = 20 ms, allowing parallel shifts of all F0 points by 20 ms with one cursor step along the horizontal to the right and to the left from the (medial) zero position. The peak pattern can thus be continually shifted along the time scale and the auditory consequences tested in a quick succession from stimulus to stimulus of the same sentence type. Similar changes can be made in the frequency axis.

Both procedures (a) and (b) are very effective for quick hypothesis testing and quick checking of points left open by the more elaborate perception experiments, and have been used a good deal in the Kiel Intonation Project to confirm and expand formal test results as well as to prepare the ground for new hypotheses and their evaluation in group listening tests.
Fig. 2
RULSYS development system output of the symbolic input "Sie hat ja gelogen @ZZ." with an early FO peak. FO (in Hz; square parameter and cosine interpolation between defined FO points) and phonetic transcription aligned to the time scale (segment and cumulative durations in cs); cursor positioned on the peak value; EO = φ.

Fig. 3
X-Y plane for providing variables, defined in a TTS rule (e.g. time and frequency), with new values by moving a cursor along the horizontal and/or the vertical axis.
2. FO peak alignment

2.1. Phonetics and phonology

2.1.1 Kiel experiments on German

The first question to be asked with regard to FO peak alignment is as to how the acoustic continuum of FO maximum value position from early (well before the onset of the stressed vowel with which it is associated) to medial (around the stressed vowel centre) to late (at the end of the stressed vowel) is partitioned perceptually. Is the continual change of the temporal relation of the FO maximum to stressed vowel onset correlated with a gradual perceptual change, or are there categorical breaks corresponding to phonological switches, and how many of these have to be recognized? The second question, which is closely linked with the first one, relates to whether the perceptual organization of the physical continuum is dependent on the segmental structure of the stressed syllable, in particular the duration of the stressed vowel, the clear acoustic segmentability of stressed syllable initial consonants (laterals or fricatives vs. glides or creaky onset) and the presence of post-vocalic voicing. To find answers to these questions peak shift series were created for the following five utterances:

1. "Sie hat ja gelogen." [зи hat çã ˈloːɡoŋ] ("She's been lying.")
2. "Es ist ja gelungen." [es ist çã ˈluŋən] ("It has worked.")
3. "Sie hat ja gejodelt." [зи hat çã ˈjoːdoːlt] ("She's been yodelling.")
4. "Sie muß wohl arbeiten." [зи müs vol ˈʔarbaɪtəŋ] ("She will have to work.")
5. "Er ist ja geritten." [ɛɐ ist çã ˈɡʰiːtən] ("He's been riding.")

2.1.1.1 "Sie hat ja gelogen."

Taking the medial FO peak position of the original utterance in Fig. 1 as a point of departure, the contour A1A2 was moved along the time axis in 6 equal steps of 30 ms each to the left and 4 corresponding steps to the right. In the transposition to the right, both branches were moved in parallel, in the one to the left, only the rising branch was, the falling one being expanded between the new maximum position and the original right base point. A series with complete parallel shift also to the left was generated as well, but the LPC synthesis quality was inferior due to the long low-level FO, sounding rather "metallic", although the pitch pattern was not unnatural, conveying the meaning of greater finality in the
statement and of less room for argument. Moreover, the natural productions of early peaks in this sentence showed the same flattened F0 descent. As informal listening did not suggest a different behaviour with regard to the perceptual assessment of shifts in the peak position in the two series, the one with the adjusted falling branch was chosen for the listening experiments.

2.1.1.1 Discrimination tests
The 11 stimuli entered into both discrimination tests (1) and (2) of 1.4 in the ascending as well as the descending order.

Results

Table I presents the responses by 60 listeners in the left-right peak sequence of the serial discrimination test, Table II the responses by 33 listeners in the right-left sequence.

Table I
Frequency distribution of 'change has occurred' responses by 60 listeners in the left-right sequence of the serial discrimination test across the 11 stimuli with F0 peak shifts in "Sie hat ja gelogen." (1 = left-most, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>1</td>
<td>4</td>
<td>39</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td>1</td>
<td>5</td>
<td>11</td>
<td>15</td>
<td>21</td>
<td>22</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>4</td>
<td>40</td>
<td>21</td>
<td>11</td>
<td>15</td>
<td>21</td>
<td>22</td>
<td>11</td>
</tr>
</tbody>
</table>

Table II
Frequency distribution of 'change has occurred' responses by 33 listeners in the right-left sequence of the serial discrimination test across the 11 stimuli with F0 peak shifts in "Sie hat ja gelogen." (1 = left-most, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>5</td>
<td>7</td>
<td>13</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>16</td>
<td>6</td>
<td>10</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>1</td>
<td>11</td>
<td>15</td>
<td>10</td>
<td>19</td>
<td>7</td>
<td>10</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
The randomized paired discrimination test in the ascending ordering was carried out with a group of 39 subjects, in the descending ordering with a different group of 34 subjects; each of the two tests contained the pairings of the identical stimuli at the uneven rank numbers in the series of 11 test items described in 2.1.1.1 and ordered from left-most to right-most F0 peak position. Fig. 4 shows the results.

Fig. 4
Discrimination functions in the randomized paired discrimination test, showing percentage of 'different' judgements for utterance pairs of "Sie hat ja gelogen." with 0-step (a), 1-step (b), or 2-step (c) distances of F0 peak positions, in the ordering left-right (continuous line) or right-left (broken line). The stimulus number refers to the second stimulus in the ascending and to the first stimulus in the descending order. 73 sbs., n = 146 at each data point (a); 39 sbs., n = 78 in the left-to-right, 34 sbs., n = 68 in the right-to-left ordering of (b) and (c).
Discussion

Both types of test converge in demonstrating a major and a minor peak in the discrimination function - around stimuli 5/6 and 9/10, respectively, but also a strong order effect. On the one hand, discrimination is sharpest, and equally so in both orderings of different stimuli, if the 1-step distance is located between stimuli 5 and 6, or, correspondingly, the 2-step distance between stimuli 5 and 7 (i.e. for the pairs 5 - 6, 6 - 5, 5 - 7 and 7 - 5); on the other hand, the differentiation weakens if the distance is located at a lower position in the series for the descending sequence (5 - 4, 5 - 3) or at a higher position for the ascending one (6 - 7, 7 - 8, 6 - 8, 7 - 9). Stimulus 5 is highly discriminated if it comes second or is spanned in the pair, i.e. in 4 - 5, 3 - 5, 4 - 6, and this even occurs by way of 'false alarms' in identical pairings of stimulus 5.

So the question arises as to what there is in the signal that might mark stimulus 5 as different from all the others. Fig. 5 shows the positions of the F0 peaks in stimuli 4 and 5 in relation to the speech wave. Stimulus 5 is the first one in the series of 11 from left to right, where the F0 contour enters the accented vowel /o:/ on a rising slope; in all the preceding stimuli in the series, F0 falls throughout the vowel. In stimulus 5 the increase of acoustic energy in the transition from the consonant /l/ to the vowel /o:/ is thus coupled with a rising F0, the rising slope of the peak contour across /galo:/ being intensified over its final 30 ms. In stimulus 4 this does not happen, but a fall is intensified instead. As the peak is moved further to the right, the F0 rise becomes progressively more extensive over a progressively longer increase in acoustic energy up to the middle of the vowel, i.e. to the F0 peak position in stimulus 7, which coincides with the original production. In this continuum, distinctivity between successive stimuli will drop, if the increase in the F0 rise has reached perceptual saturation. This seems to happen after stimulus 6.

A further shift of the F0 peak to the right beyond stimulus 7 results in an increasing low F0 stretch (see Fig. 1), which receives the intensification, whereas, at the same time, the end of the rise is linked with a decrease of acoustic energy. When both parameter changes are large enough, successive stimuli have their distinctivity raised again. This seems to happen around
stimuli 9 and 10 in the ascending order, but is obviously a much weaker effect than the change from falling to rising F0 in the stressed vowel, producing much lower peaks in the response functions.

These results suggest that there is a maximum of sensitivity in the peak shift continuum in the area of stimuli 5/6. So any pairings within or progressing towards this area are discriminated best, viz. 4 - 5, 5 - 6, 6 - 5, 7 - 6 (and even 8 - 7); 3 - 5, 4 - 6, 5 - 7, 7 - 5, 8 - 6, but not 5 - 4, 6 - 7, 7 - 8, 5 - 3, 6 - 8, where the progression is away from the area of high sensitivity. A second, weaker sensitivity peak, is located at stimuli 9/10, but does not surface in the response functions for the descending order, because of the displacement to the right of the discrimination curve associated with stimuli 5/6. The acoustic continuum is thus perceptually partitioned into two clearly delimited sections with the boundary occurring between stimuli 4 and 6, and this perceptual division coincides with an acoustic change from falling to rising F0 across stressed vowel onset. Around the boundary between these two sections, discrimination

Fig. 5
F0 peaks in stimuli 4 and 5 of the series of 11 "Sie hat ja gelogen." from left-most to right-most position, in relation to the speech wave. The vertical lines mark the F0 maximum.
is sharpest, and, as will be seen in 2.1.1.1.2 and 2.2, the two perceptually
determined sections of the acoustic continuum correspond to two intonational
categories related to a semantic differentiation between 'established' and
'new'.

So it appears that we are dealing here with an example of 'categorical
perception' (see Repp, 1984), this time in the domain of pitch (Kohler,
1987a). The data point to an abrupt perceptual change when in the acoustic
continuum the F0 peak is moved into the vowel of the stressed syllable. A
further F0 peak shift along the acoustic continuum results in a more gradual
auditory change, with a minor sensitivity maximum at a point where the
initial stretch of low-level F0 and the final weakening of the rise-fall in
the stressed vowel become large enough. The data thus support Hypothesis (2)
(see Contribution I; Kohler, 1991b) as far as the abrupt vs. gradual changes
in perception are concerned. This means that an early F0 peak must
constitute a phonological category of German intonation, contrasting with a
medial peak, whereas a late peak is less clearly separated, although the
perceptual results may turn out to be different if in accordance with
natural production the F0 peak shift to late positions were accompanied by a
similar shift of the acoustic energy maximum to the right (whereas in the
stimulus manipulation the energy profile of the original medial F0 peak
utterance, synchronized with F0 on the vowel centre, was used). The minor
sensitivity maximum in the response function could then easily be boosted
(see 2.1.1.5). In 2.1.1.1.2 and 2.2, further support will be given to the
organization of the semantic functions in parallel with the perceptual and
phonological structuring of F0 peak alignment.

2.1.1.1.2 Identification tests
On the basis of the discrimination test results and of hypotheses concerning
the semantics of early, medial and late peaks, three contexts were
constructed:
(1) "Wer einmal lügt, dem glaubt man nicht, auch wenn er gleich die Wahrheit
spricht. Das gilt auch für Anna."
("Once a liar, always a liar. This also applies to Anne.")
This context sets the frame for an established fact and the summing up
of an argument, which is brought to a close.
(2) "Jetzt versteht' ich das erst."
("Now I understand.")
This context presents a new fact and opens up a new argument.

(3) "Oh!"
This context introduces emphatic surprise.

Each of these contexts was spoken naturally and paired with each of the three naturally produced peaks in the sentence "Sie hat ja gelogen." to form a natural stimuli identification test according to 1.4 (3). Furthermore, a synthesized stimuli identification test (see 1.4 (4)) was performed with pairings of context (2) ("Jetzt") and each one of the first 8 stimuli in the continuum (from left to right) of 2.1.1.1.

Results

Table III and Fig. 6 present the results of the two tests.

Table III
Percentages of 'matching' responses for combinations of 3 contexts and early, medial or late F0 peaks in the sentence "Sie hat ja gelogen." in a natural stimuli identification test. 88 subjects

<table>
<thead>
<tr>
<th>Context</th>
<th>(1) Wer</th>
<th>(2) Jetzt</th>
<th>(3) Oh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak position</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>early</td>
<td>87.5</td>
<td>27.3</td>
<td>8.0</td>
</tr>
<tr>
<td>medial</td>
<td>26.1</td>
<td>70.5</td>
<td>72.7</td>
</tr>
<tr>
<td>late</td>
<td>13.6</td>
<td>67.0</td>
<td>76.1</td>
</tr>
</tbody>
</table>

Discussion

The results of combining the 3 contexts and 3 F0 peak positions show that subjects are able to make systematic judgements because the responses are significantly different from chance, being either more than 66% or less than 30% in favour of 'matching'. This means that the different F0 peak positions must be perceptually identifiable, and since in all cases the identification of an early versus a non-early peak is far more clearly differentiated than that of a medial versus a late one, this identification test reproduces the categorization of the discrimination tests. It is only in the "Wer" context that the medial vs. late F0 peaks yield a significant difference in the
Identification function in the synthesized stimuli identification test, showing percentage 'matching' judgements for 8 stimuli "Sie hat ja gelogen." with FO peak shift from left to right in the context "Jetzt versteh ich das erst." 19 subjects; for each stimulus n = 190.

response patterns ($\chi^2 = 4.31, p = .05$). Contrariwise, the early pattern fits least into the "Oh" context (difference between "Jetzt" and "Oh" contexts $\chi^2 = 31.07, p = .001$).

The contextualization of the early to medial FO peak continuum with the "Jetzt" introduction (Fig. 6) shows an abrupt change from 'matching' to 'non-matching' judgements in spite of the gradual change along the physical dimension, and thus adds support to the assumption of a categorical perception advanced in connection with the discrimination tests. Stimuli 1 - 4 represent one perceptual identification category, stimuli 6 - 8 a
different one. They may be regarded as two phonological categories, viz. 'early' and 'medial' F0 peaks. The discrimination of stimuli is sharpest between these identification categories, which is precisely what the theory of categorical perception postulates.

2.1.1.1.3 "Sie hat gelogen."

As the objection was raised that the responses in the natural stimuli identification test of 2.1.1.1.2 might have been influenced by the modal particle "ja" ("after all"; "I see") predetermining the judgement, a new set of 9 context-peak combinations was generated by excising the signal portions corresponding to "ja" from the existing ones used in the test of 2.1.1.1.2. This splicing was easy to perform because the word was bounded by silence (= voiceless occlusions in [t] and [§]). Then two long versions of the natural stimuli identification test according to 1.4(3) were generated: one with the stimuli "Sie hat ja gelogen." and one with "Sie hat gelogen."

These two tests were run at one week's interval with two groups of subjects in the following sequence:

Group I (17 subjects) did the test with the "ja" stimuli first, the other test second, for Group II (7 subjects) the order was reversed. Table IV presents the results.

Table IV
Percentages of 'matching' responses for combinations of 3 contexts and early, medial or late F0 peaks in the sentences "Sie hat (ja) gelogen." in a natural stimuli identification test with 10 repetitions and two groups of subjects (I: 17 sbs; II: 7 sbs); A with, B without "ja"

<table>
<thead>
<tr>
<th></th>
<th>(1) Wer</th>
<th>(2) Jetzt</th>
<th>(3) Oh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>I</td>
</tr>
<tr>
<td>early</td>
<td>A</td>
<td>82.9</td>
<td>81.4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>86.4</td>
<td>85.7</td>
</tr>
<tr>
<td>medial</td>
<td>A</td>
<td>41.2</td>
<td>28.6</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>48.2</td>
<td>57.1</td>
</tr>
<tr>
<td>late</td>
<td>A</td>
<td>25.3</td>
<td>22.9</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>31.2</td>
<td>37.1</td>
</tr>
</tbody>
</table>

As in 2.1.1.1.2 (see Table III), the responses to the "ja" stimuli (= A) are in all cases either clearly positive or negative, and significantly
different from equal distribution. Again the 'medial' and 'late' peaks produce more similar judgement patterns than the 'medial' and 'early' ones, and they are only significantly different for Group I in the "Wer" context ($\chi^2 = 9.66$, $p = .01$) and in the "Oh" context ($\chi^2 = 6.44$, $p = .05$). The strong distinction between 'early' and 'medial' and the much weaker differentiation for 'medial' and 'late' has thus been confirmed. This finding once more supports the hypothesis of a categorical switch from 'early' to 'medial' and a gradual change from 'medial' to 'late' peak positions in the utterance "Sie hat ja gelogen."

For the stimuli without "ja", in principle the same data were obtained. Of the 18 comparisons of the results for utterances with/without "ja" only four are statistically significant according to $\chi^2$ tests, the first one in Group I, the others in Group II:
(a) the 'late' peak in the "Oh" context; $\chi^2 = 5.32$, $p = .05$,
(b) the 'medial' peak in the "Wer" context; $\chi^2 = 11.67$, $p = .001$,
(c) the 'early' peak in the "Jetzt" context; $\chi^2 = 13.75$, $p = .001$,
(d) the 'late' peak in the "Jetzt" context; $\chi^2 = 6.89$, $p = .01$.

In (a), (b), (c) the difference implies an increase in 'matching' answers for stimuli without "ja", which is contrary to what would have to be expected if the objection were valid. In the remaining case (d), however, there is a decrease in 'matching' responses for stimuli without "ja", which may be taken as an indication of a strengthening through the modal particle "ja" of the meaning conveyed by intonation. But the results cannot be solely determined by the particle, all the less so since this pairwise testing increases the $\alpha$ error and may thus reject the null hypothesis of no distinction between the two utterance types, although it is correct.

A further objection might be that the order of the two tests had an influence on the results: if the "ja" stimuli are tested first the pattern would also be set for the stimuli without "ja". Group II, for which the order was reversed, should thus produce a significantly smaller number of 'matching' responses for the stimuli without "ja" more frequently than Group I, but the above data do not support this assumption. Moreover, the stimuli without "ja" do not show significant distinctions between Groups I and II, with the one exception of the 'medial' peak in the "Oh" context ($\chi^2 = 9.64$,}
p = .01). In view of the possible increase of the \( \alpha \) error, we can thus say that the test order did not have a significant influence on the response patterns, which are basically determined by an intonational phonology, i.e. by 'early' vs. 'non-early' FO peak positions - less strongly by 'medial' vs. 'late' ones -, and which may be heightened, but not replaced by, other formal means, such as modal particles.

2.1.1.2 "Es ist ja gelungen."
The question now arises as to whether the perceptually relevant timing differences between different peak positions relative to stressed vowel onset are transferable to other syllable structures and in what ways they may have to be adjusted. The first syllable structure selected was the one containing a phonologically short vowel, instead of a long one, in an otherwise comparable segment chain: "Es ist ja gelungen." (see 2.1.1). Fig. 7 shows the speech wave as well as the energy and FO contours in the natural medial-peak token selected for FO peak shift. The test stimulus generation followed the procedure of parallel shifts of both branches of the peak contour (see 1.3). The step size was 30 ms, and one peak was located at the boundary between the stressed-syllable initial consonant /l/ and the stressed vowel /u/. Fig. 8 shows the 9 different peak positions used for the stimulus generation. Only the quick serial discrimination test (see 1.4 (1)) was performed in the left-right sequence with 29 subjects.

Results

Table V presents the results.

Table V
Distribution of 'change has occurred' responses by 29 listeners in the left-right sequence of the serial discrimination test across the 9 stimuli with FO peak shifts in "Es ist ja gelungen." (1 = left-most, 9 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived at</td>
<td>5</td>
<td>19</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived at</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>21</td>
<td>10</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
Speech wave, energy and FO contours (linear scale) in the natural medial-peak token of "Es ist ja gelungen." selected for FO peak shift. The time marks indicate on- and offsets of /g/, /a/, /V and /u/. The broken lines mark the left and right base points as well as the maximum of the peak configuration to be shifted.

Discussion

There is again an abrupt change in the response pattern as the FO peak is moved into the stressed vowel. The absolute timing of positions 5 and 6 after vowel onset, i.e. 30 ms and 60 ms, respectively, is exactly the same as in the stimuli "Sie hat ja gelogen." (see 2.1.1.1.1). These data point to an absolute time span of up to 60 ms into the stressed vowel that is responsible for a phonological change from 'early' to 'medial' peak, independent of the phonological vowel quantity and consequently of vowel duration following the FO peak, at least in disyllables. This finding means that the 'medial' FO peak has a later relative position in a short vowel than in a long one, viz. closer to its offset, and this ties in with the production and perception data in Contribution II (Gartenberg & Panzlaff-Reuter,
Fig. 8
Speech wave and F0 contour (linear scale) in "Es ist ja gelungen." with time marks indicating the 9 F0 positions for complete-contour shift from left to right.

This means, furthermore, that the series of 9 stimuli did not include a proper 'late' peak: it would have had to be located well into the unstressed vowel /a/.

2.1.1.3 "Sie hat ja gejodelt."
The next syllable structure to be considered contains a long stressed vowel /o:/, as in "Sie hat ja gelogen.", but a glide /j/ with a much more gradual articulatory/acoustic transition in the initial position of the stressed syllable, instead of the more abrupt change associated with the initial lateral /l/: "Sie hat ja gejodelt." (see 2.1.1). The question is as to whether the more gradual spectral transition influences the perception of the F0 transition into the stressed vowel, because the F0 peak position relative to vowel onset can be less clearly assessed. Fig. 9 shows the speech wave as well as the energy, F0 and spectrum displays in the natural
Fig. 9
Speech wave, energy, FO (linear scale) and spectral displays in the natural medial-peak token of "Sie hat ja gejodelt." selected for FO peak shift. The time marks indicate the left base point (appr. in the temporal centre of the F2 transition for /ajo:/), the maximum FO value and the right base point.

medial-peak token selected for FO peak shift. The test stimulus generation followed the same procedure as in 2.1.1.2, with a step size of 35 ms, and one peak (nr 5) being located at the temporal centre of the F2 formant transition in /ajo:/.

Fig. 10 shows the 11 different peak positions used for the stimulus generation. Only the quick serial discrimination test (see 1.4 (1)) was performed in the left-right sequence with 24 subjects.
Fig. 10
Speech wave and F0 contour (linear scale) in "Sie hat ja gejodelt." with time marks indicating the 11 F0 peak positions for complete-contour shift from left to right.

Results

Table VI presents the results.

Table VI
Distribution of 'change has occurred' responses by 24 listeners in the left-right sequence of the serial discrimination test across the 11 stimuli with F0 peak shifts in "Sie hat ja gejodelt." (1 = leftmost, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>10</td>
<td>16</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Discussion

In this case the first change occurs less abruptly although it is still clearly marked and coincides with the temporal half-way position of the F0 peak in the F2 transition. Further changes in the perceptual profile also occur earlier than in the other stimulus types tested so far. All this goes to show that a glide transition does interfere with the categorization of F0 peaks, but the general pattern of a phonological separation of 'early' and 'medial' peaks and a gradual switch from 'medial' to 'late' stays.

2.1.1.4 "Sie müß wohl arbeiten."

The next syllable structure chosen has creaky voice (the phonetic realisation of a syllable-initial vowel prefixed by a glottal stop) before a

![Speech wave, energy and F0 contours](image)

**Fig. 11**

Speech wave, energy and F0 contours (linear scale) in the natural medial-peak token of "Sie müß wohl arbeiten." (with creaky voice transition instead of a glottal stop interruption of voicing) selected for F0 peak shift. The time marks delimit the F0 peak configuration that was shifted (left and right base points, and maximum).
stressed long vowel: "Sie müß wohl arbeiten." (see 2.1.1). The question is as to whether a creaky voice onset has the same effect on FO peak categorization as a glide. Fig. 11 shows the speech wave as well as the energy and FO contours in the token selected for FO peak shift. The test stimulus generation followed the same procedure as in 2.1.1.2, with a step size of 35 ms and one peak (nr 5) being located at the onset of the more regular glottal vibrations at the transition from /l/ to /ae/. Fig. 12 shows the 11 different peak positions used for the stimulus generation. Only the quick serial discrimination test (see 1.3 (1)) was performed in the left-right sequence with 24 subjects.

**Fig. 12**
Speech wave and FO contour (linear scale) in "Sie müß wohl arbeiten" with time marks indicating the 11 FO peak positions for complete-contour shift from left to right.
Results

Table VII presents the results.

Table VII
Distribution of 'change has occurred' responses by 24 listeners in the left-right sequence of the serial discrimination test across the 11 stimuli with FO peak shifts in "Sie muß wohl arbeiten." (1 = left-most, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>3</td>
<td>20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>20</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discussion

The first change occurs very abruptly in stimulus 5, i.e. about 35 ms after the creaky voice transition. The perception of later changes is spread over the remainder of the continuum without clear peaks in the response function. There is a minor maximum at stimulus 10, i.e. at a similar position as in the continuum across "Sie hat ja gelogen." In every respect "Sie muß wohl arbeiten." thus patterns with the latter, rather than with the case of a glide transition in "Sie hat ja gejodelt." What seems to be important for FO peak perception is the abrupt articulatory change in the transfer function from /I/ to the stressed vowel (in "gelungen" as well as in "wohl arbeiten"), not the gradual change in phonation from voice to creak to voice, superimposed on the articulatory switch.

2.1.1.5 "Er ist ja geritten."

In the stimuli examined so far the course of the peak contour has been manifested in the observable FO values. This changes when the post- and/or pre-vocalic consonant in the stressed syllable associated with the peak is voiceless. Now part of the contour has to be reconstructed before a peak shift becomes possible. Fig. 13 shows the speech wave as well as the energy and FO contours in a natural medial-peak token of "Er ist ja geritten." selected for FO peak shift (see 2.1.1). The test stimulus generation
followed the same procedure as in 2.1.1.2 with a step size of 30 ms and 15 peak positions from left to right starting at the beginning of /a/. The quick serial discrimination test (see 1.4 (1)) was performed informally in the left-right sequence by the experimenter. Again there was an abrupt change in perception as the peak entered the stressed vowel. But as the peak was moved into the voiceless section of /t/ it lost its characteristics, becoming lower and lower in pitch. This proves that the maximum value of a peak contour must be present in the signal for identification: it is not reconstructed by a listener from surrounding values of the rising and falling branches, whereas a low right base point may be missing due to FO contour truncation before voicelessness in final syllables (see Gartenberg & Panzlaff-Reuter, 1991, 3.) without detriment to the peak characteristics (on the contrary, there must be truncation in certain contexts to guarantee pattern identity).

Fig. 13
Speech wave, energy and FO contours (linear scale) in the natural medial-peak token of "Er ist ja geritten." selected for FO peak shift. The time marks indicate on- and offsets of /g/, /a/, /w/, /i/, /t/, /n/. The dotted line represents the reconstructed FO interpolation of the right branch of the peak contour. The broken lines mark the left and right base points as well as the maximum of the peak configuration to be shifted.
A further shift of the peak contour maximum to the onset of voicing in the final /n/ approximates the FO configuration found in natural productions of late peaks (see Fig. 14), but the auditory impression is still that of a medial peak, not of a late one. A comparison of Figs. 13 and 14 shows that the final nasals in medial and late peaks differ in amplitude and mode of vocal fold vibration: In medial peaks (and the same would apply to early ones), the low FO fall at the end of an utterance is accompanied by a drop in source amplitude, which weakens unstressed vowels and sonorants considerably, often reducing them to creaky voice and to irregular breathy glottal pulses. In late peaks, this decline is moved to the right following the later FO fall, thus keeping a high source amplitude at the onset of unstressed vowels and syllabic sonorants; on the other hand the low FO stretch in the stressed vowel before the peak gets its intensity reduced. So there is a natural parallelism in the time courses of FO, source amplitude and sound intensity for the three peak contours. If it is destroyed, the

![Speech wave, energy and FO contours](image)

**Fig. 14**

Speech wave, energy and FO contours (linear scale) in a natural late-peak token of "Er ist ja geritten." The time marks indicate on- and offsets of /ɡ/, /ŋ/, /n/, /ɪ/, /t/, /n/. 

144
perceptual pattern identity may be lost.

Thus a late peak, positioned at the sonorant voicing onset after a voiceless obstruent, can only be successfully reconstructed by a listener if the FO descent to the terminal low level has a high enough source amplitude to guarantee sufficient intensity in the final sonorant for the high falling FO contour to be auditorily monitored. But a natural medial peak utterance with its low final intensity and glottal irregularity lacks these attributes and cannot be turned into a late peak percept simply by an FO shift into the appropriate location. The amplitude and duration of the final sonorant have to be raised considerably at the same time and the mode of vibration changed. This can be achieved by transferring the final /n/ from the late peak stimulus. Contrariwise, with a late peak stimulus as point of departure a perceptually convincing medial peak pattern can only be generated if in addition to the peak shift the final sonorant is drastically lowered in amplitude and shortened. This has also been reproduced in a RULSYS TTS formant synthesis-by-rule (Kohler, 1991f).

2.1.1.6. "Sie hat ja gestritten."
If a short stressed vowel is not only followed but also preceded by voiceless obstruents the masking of peak height as the maximum value is moved into the voiceless section arises syllable-initially as well. Fig. 15 shows the speech wave as well as the energy and FO contours in a natural medial-peak token of "Sie hat ja gestritten." [z̥ hat ja ˈʃaˌʃaʃtən] ("She's been quarrelling.") where FO sets in higher in the stressed vowel compared with "geritten" of 2.1.1.5, because the initial cluster /ʃt/ is much longer than the initial /ʃ/, and FO, rising from the left base point in /a/, has thus reached a higher value at vowel onset. In addition, there is a CFO increase caused by the preceding voiceless fricative (see Gartenberg & Panzlaff-Reuter, 1991, 3.). "geritten" and "gestritten" converge, however, in having their FO maximum close to vowel offset, as is usual for medial peaks in short stressed vowels before an unstressed syllable (see loc.cit., 5.2.).

The peak shift in "Sie hat ja gestritten." was tested interactively using the TTS research tool (see 1.5 (b)) with the rule-generated medial peak position as a point of departure (see Fig. 16a) and a step size of 20 ms in complete parallel shift. Significant FO values in the rule-generated
Fig. 15
Speech wave, energy and FO contours (linear scale) in a natural medial-peak token of "Sie hat ja gestritten." The time marks indicate on- and offsets of /g/, /a/, /j/, /t/, /u/, /i/, /t/, /n/.

The utterance at /a/ and /i/ on- and offsets are 84 Hz, 88 Hz, 144 Hz and 140 Hz, respectively. When the peak is located 40 ms before the beginning of /i/, the peak is clearly 'early'; at /i/ onset (see Fig. 16b) it has changed to 'medial'. The corresponding significant FO values in these two positions are 88 Hz, 104 Hz, 138 Hz, 86 Hz, and 84 Hz, 94 Hz, 148 Hz, 108 Hz. So the change from 'early' to 'medial' occurs quite abruptly in this syllable structure as well when the FO rise across the voiceless cluster becomes more extensive than the fall, and the FO offset in the stressed vowel is in the middle of the FO range between maximum and minimum values in the utterance. Thus in this sentence, the switch from 'early' to 'medial' occurs before there is an initial FO rise in the stressed vowel, which is different from all the other syllable structures, with initial voiced consonants, analysed so far. The reason for this difference lies in the CFO interference, which is obviously accounted for in the perception process.
Fig. 16
(a) RULSYS output of "Sie hat ja gestritten." (= (default) medial peak). (b) 60 ms peak shift to the left (= first clear medial peak position in left-right move). FO (in Hz; square parameter and cosine interpolation between set FO points) and phonetic transcription aligned to the time scale (segment and cumulative durations in cs); EO = ø, SH = j.
2.1.1.7 Conclusion

The discrimination and identification tests of 2.1.1.1-5 all point in the same direction, viz. the perceptual exploitation of different FO peak synchronizations with stressed vowel onsets and of the ensuing low (falling) vs. high (rising) FO as a psychophonetic basis for phonological categorization at the level of intonation. For an 'early' peak, FO is low in the stressed vowel because it is on its descent at the vowel onset and, in complete parallel shift, also reaches its low end point early. If there is voicing before the stressed vowel, the FO point at vowel onset is preceded by a higher FO value so that FO falls into the accented syllable. If there is no previous voicing in utterance-initial position of a stressed syllable beginning with voiceless consonants, FO at vowel onset has as low a value as would result from an FO descent across the stressed syllable periphery to strengthen the low FO level in the accented vowel. The 'early' peak is thus characterized by a high prenuclear FO - either directly observable or by extrapolation from the FO start in the stressed syllable nucleus - and by a low FO in the latter.

Contrariwise, the 'medial' peak has a low prenuclear FO, an FO rise (of at least 2 semitones from nuclear vowel onset to peak value, according to interactive testing), and a subsequent descent to a low FO at a later point in time than in an 'early' peak. The amount of descent depends on syllable structures, and the rise may be absent because of CFO interference, resulting in a higher FO starting point at nucleus onset. So in all cases, the 'medial' peak accentuates a higher FO level in the stressed vowel than the 'early' peak. In a 'late' peak the rise is extended because it occurs later, but it is also prefixed by a stretch of low level FO.

Interactive perceptual testing (see 1.5) has further shown that the 'early' and 'medial' peak patterns do not lose their characteristic auditory difference if their right base points have the same FO value at the same time after nucleus onset (due to a flattening of the FO descent in the 'early' peak). It is thus the FO differentiation in the initial part of the nuclear vowel that counts as the distinctive feature. If in a RULSYS generated 'early' peak of "Sie hat ja gelogen.", FO is kept at the peak maximum value up to a point including the first 3 FO frames of 10 ms each in the stressed vowel, instead of having an immediate fall, the auditory
characteristics of the 'early' peak are not lost. On the other hand, if, starting from a 'medial' peak configuration in the above utterance (with /l/ onset = 88 Hz, /l/ offset = 116 Hz, stressed vowel onset consisting of the F0 sequence 122 - 128 - 128 - 130 Hz), the /l/ offset and all the vowel onset frames are raised to 130 Hz, the 'medial' peak is changed in the direction of an 'early' one although the only difference between the two patterns now lies in the rise being completed in the /l/ rather than continuing into the accented vowel, i.e. in the presence or absence of a nucleus-initial rise. Admittedly, the difference between the 'early' and 'medial' patterns is clearly weakened by this modification, but it shows that a 'medial' peak needs an F0 rise in the nucleus after a sonorant.

2.1.2 Munich experiments on German

The results of the Kiel experiments naturally prompted the question as to how widespread the phonological categorization of F0 peak positions is in the intonation system of German in general. Therefore the serial discrimination and randomized paired discrimination tests in the ascending ordering (2.1.1.1.1) as well as the synthesized stimuli identification test (2.1.1.1.2) were repeated in the Phonetics Institute of Munich University with groups of listeners of a Bavarian dialect background. The tests were performed in the Institute language laboratory and the stimuli were presented over headphones. Tapes and instructions were identical to the ones in the Kiel experiments. 11 listeners participated in the serial discrimination test and in the identification test, 14 in the randomized paired discrimination test.

Results

Table VIII presents the results of the serial discrimination test.

---

3 I wish to thank Dr Anton Batliner for organizing the test runs.
Table VIII

Distribution of 'change has occurred' responses of 11 Munich listeners in the left-right sequence of the serial discrimination test across the 11 stimuli with FO peak shifts in "Sie hat ja gelogen." (1 = left-most, 11 = right-most)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 17 shows the results of the randomized paired discrimination test.

![Graph](image.png)

Fig. 17

Discrimination functions in the randomized paired discrimination test, showing percentage of 'different' judgements for utterance pairs of "Sie hat ja gelogen." with 0-step (a), 1-step (b), or 2-step (c) distances of FO peak positions, in the ordering left-right. The stimulus numbers refer to the second stimulus. 14 sbs., n = 28 at each data point in (a), (b), (c).
Fig. 18 shows the results of the synthesized stimuli identification test.

Identification function in the synthesized stimuli identification test, showing percentage 'matching' judgements for 8 stimuli "Sie hat ja gelogen." with FO peak shift from left to right in the context "Jetzt versteh' ich das erst." 11 subjects; for each stimulus n = 110.

Discussion

The comparison of Tables I and VIII shows that the Munich group has the same type of response pattern with a maximum for stimulus 5 and a large scatter for further changes in the series from stimulus 7 to 11. However, due to the much smaller number of subjects in the Munich group, the minor peak in the response curve does not show up so clearly. The results of the serial discrimination test are supported by those of the randomized paired discrimination test of Fig. 17 (in comparison with Fig. 4). There is again a maximum of sensitivity in the FO peak shift continuum in the area of stimuli 5/6 and a second, weaker sensitivity peak at stimulus 9, but the sensitivity area is narrower, with the pairings 5 - 6 and 3 - 5 not being included in
the maxima, and there is no peak of 'false alarms' for the 5 - 5 pair. The identification function of Fig. 18 (in comparison with Fig. 6) points to the same two perceptual identification categories comprising stimuli 1 - 4, on the one hand, and stimuli 6 - 8, on the other, but with a lot more noise (an offset of about 20% - 30%) in the first category and at stimulus 5, the boundary between the two. We may again associate this partitioning with the two phonological categories of 'early' and 'medial' FO peaks. The Munich results are thus in agreement with the Kiel data and allow the generalization of a perceptual and phonological categorization of FO peak positions relative to stressed vowel onset for the intonation of German across regional varieties.

2.1.3 Experiments on other languages
What remained an open issue after the very clear results of the experiments on different varieties of German was whether we are here dealing with a phonological categorization of German, albeit on a psychophonetic basis, or whether the phenomenon is more widespread or even a language universal, based on a specific feature of human speech perception in general. The hypothesis that such a general psychophonetic principle does operate in the perception of FO patterns in human speech, irrespective of the phonological categorization and the linguistic functions it may serve in any particular language, leads to the assumption that native speakers of other languages than German listening to German utterances should be able to detect changes in FO peak positions in relation to general human consonant-vowel sequences, even without knowing any German at all, and therefore without assessing the stimuli semantically, but simply on the basis of general phonetic properties of human speech. If the results of such listening tests were to coincide with the results for German, this would be a strong indication of a language-independent psychophonetic mechanism. As a first step in this direction, the serial discrimination test in the ascending ordering of 2.1.1.1.1 was run with two groups of non-German speakers: (a) 25 Russian speakers in Leningrad⁴, who had no knowledge of German and who either worked on Russian, English or French phonetics (11) or were students in their first or second year in the Philological Faculty (14).

⁴ I wish to thank Prof. Natalia Svetozarova of Leningrad University for administering the test in her Phonetics Laboratory.
A copy, on standard cassette, of the original series of 11 stimuli of "Sie hat ja gelogen." with FO peak shifts from left to right was provided. The subjects listened to the series twice and then had to cross, on a prepared answer sheet, the number of the stimulus in the series that they perceived as being most clearly different from the rest.

(b) 40 native speakers of 13 different languages attending German language courses at beginners or advanced level at Kiel University. The original test tape was presented to them over loudspeaker in four subgroups (twice 14 and twice 6 listeners) in their relatively quiet but acoustically non-treated classroom. The answer-sheets and the procedure were the same as for the corresponding test with German listeners in 2.1.1.1.1. A great deal of time and care was spent on explaining the

<table>
<thead>
<tr>
<th>Native language</th>
<th>Native country</th>
<th>Beginners</th>
<th>Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Farsi</td>
<td>Iran</td>
<td>9</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Polish</td>
<td></td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Portuguese</td>
<td>Brazil</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Korean</td>
<td></td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Spanish</td>
<td>Chile</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Spanish</td>
<td>Argentina</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>English</td>
<td>USA</td>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>English</td>
<td>England</td>
<td></td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Arabic</td>
<td>Israel</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Japanese</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Thai</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Nepali</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Singhalese (Sri Lanka)</td>
<td></td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
</tbody>
</table>

Table IX
Background information about the 40 foreign listeners in the *serial discrimination test*
test instructions in German. Table IX provides the background information about the 40 listeners.

Results

Table X presents the results of the Russian group. Although the instruction demanded a single response, some subjects indicated more than one stimulus as being clearly different.

Table X

Frequency distribution of 'clearly different' responses by 25 Russian listeners without any knowledge of German in the left - right sequence of the serial discrimination test across the 11 stimuli with FO peak shifts in "Sie hat ja gelogen." (1 = left-most, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneticians</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Non-phoneticians</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>18</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table XI presents the results of the multilanguage group, restricted to the perception of the first change in the series. The one Chinese, one Farsi and one Korean speaker did not perceive any change at all, although the other three Korean speakers did.

Table XI

Frequency distribution of 'first change has occurred' responses by 40 listeners of 13 different languages, in the left - right sequence of the serial discrimination test across the 11 stimuli with FO peak shifts in "Sie hat ja gelogen." (1 = left-most, 11 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>first change perceived</td>
<td>2</td>
<td>9</td>
<td>19</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

(3 listeners perceived no change at all.)

5 Robert Gartenberg carried out the tests and compiled the data.
Discussion

Both groups, in spite of their language diversity, converge in having a clear maximum of the response function for stimulus 6. This is a higher position than for the German listeners, who favoured stimulus 5, but who also provided a substantial portion of their answers for stimulus 6. These results are a very strong indication that the dichotomy between an 'early' and a 'medial' peak position is indeed a general psychophonetic, language-independent phenomenon, which may then be incorporated into the language-specific phonology at different levels.

Thus in Mandarin Chinese (see Gårding, Kratochvil, Svantesson & Zhang, 1985) it is put to use in the tone system, differentiating between the continuously (low) falling F0 of tone 3 (e.g. in ma3 'horse') and the (high) rising-falling F0 of tone 4 (e.g. in ma4 'to curse'). It is worth noting in this connection how a Chinese speaker (Dr Chilin Shih, research worker at Bell Labs, Murray Hill in 1986) classified the 11 stimuli of the left-right series "Sie hat ja gelogen." without any knowledge of German. Without the slightest doubt she associated stimuli 1 - 4 with tone 3, stimulus 5 with tone 4; later in the series tone 4 changed to the combined tone 2 - 4; but whereas the switch from tone 3 to tone 4 occurred abruptly in the succession of stimuli 4 and 5, the change from tone 4 to tone 2 - 4 was gradual and could be less easily located (at stimulus 9 the change had definitely taken place). This informal test shows (a) that tones 3 and 4 in Mandarin Chinese are differentiated by the F0 maximum relative to the vowel onset, a prenuclear F0 peak signalling the former, a nuclear F0 peak the latter, and (b) that these categorizations are possible on the language-independent basis of human speech perception in general.

A different case of exploiting the perceptual relevance of earlier vs. later peaks are the acute and grave tonal word accents in Norwegian and Swedish (Gårding, 1979, 1982). And finally intonation languages like German, English and French make use of the different peak patterns in their intonation phonologies and relate them to semantic distinctions along the 'closed/open to argument' dimension (see 2.2). English and French can differentiate in the same way as German in their corresponding sentences "She's been lying." and "Elle a menti." It is an interesting research
objective for the future to investigate the different linguistic functions the dichotomy can be put to in the world's languages.

2.2 Semantics
The question to be pursued now is what linguistic functions are carried by the phonological contrasts of 'early' vs. 'medial' vs. 'late' peaks in German. In particular, it is to be ascertained whether the categorical change from 'early' to 'medial' and of the more gradual change from 'medial' to 'late' peak positions are mapped onto a semantic space in a congruent fashion. Some insight was gained from the data obtained through controlled dialogues (Gartenberg & Hertrich, 1988). Furthermore, in the Kiel and Munich serial discrimination tests (2.1.1.1 and 2.1.2) with "Sie hat ja gelogen."

subjects were also asked to paraphrase the meanings of the utterances corresponding to the three peak positions (see 1.4 (1)). Here are some of the answers.

<table>
<thead>
<tr>
<th>Kiel</th>
<th>(a) original utterance</th>
<th>(b) first change</th>
<th>(c) further change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement of a fact or end of an argumentation.</td>
<td>Introductory statement, beginning of argumentation.</td>
<td>As (b), but greater insistence.</td>
<td></td>
</tr>
<tr>
<td>Report to a third party that she has been lying; the speaker stresses a fact resulting from the environment.</td>
<td>Indignant statement.</td>
<td>Surprise statement.</td>
<td></td>
</tr>
<tr>
<td>Statement, explanation.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement without surprise.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement of a fact, e.g. in the context &quot;the punishment is justified.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

157
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Matter-of-fact statement.</td>
<td>Statement with the expression of indignation.</td>
<td>Statement with the expression of astonishment.</td>
</tr>
<tr>
<td>Declarative, expected, matter-of-fact.</td>
<td>Unexpected, indignant.</td>
<td>&quot;I can't believe it.&quot;</td>
</tr>
<tr>
<td>Confirmation of a fact; it is obvious that she's been lying.</td>
<td>Surprising fact for the speaker, the lie is unexpected.</td>
<td></td>
</tr>
<tr>
<td>Statement of fact that the speaker discovered long ago.</td>
<td>Explanation of a fact.</td>
<td>Indignation.</td>
</tr>
</tbody>
</table>

**Munich**

(a) original utterance

<table>
<thead>
<tr>
<th>Confirmation of what is already known.</th>
<th>Surprised statement, reproachful undertone (&quot;... I would not have expected that.&quot;)</th>
<th>Pure astonishment.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple statement of a fact with which the speaker seems to be familiar.</td>
<td>Surprise, indignation, speaker's reaction to a fact he did not know before.</td>
<td></td>
</tr>
<tr>
<td>But we knew that before anyway.</td>
<td>Amazement.</td>
<td></td>
</tr>
<tr>
<td>Matter-of-fact statement, there was not really any doubt about her behaviour.</td>
<td>It was not certain if she would tell the truth or not.</td>
<td>Contrary to expectation she has been lying, comes as a surprise; gradual transition from (b) to (c).</td>
</tr>
</tbody>
</table>

158
Another paraphrasing experiment was carried out with the following sentences containing first an 'early' and then a 'late' peak (the underlined word received the peak accent):

1. "Wer war das?" ("Who did that?", "Who was that?")
2. "Mach' bitte das Fenster zu!" ("Shut the window, please.")
3. "Was ist denn ein Atom?" ("What's an atom?")
4. "Sehen wir uns also morgen?" ("So we are going to see each other tomorrow.")

Each pair of 'early/medial' peak utterances was selected from 10 naturally produced repetitions (speaker KK) and played to listeners as often as they liked. They had to write down their assessment of the situation or speaker attitude that fitted each sentence and peak pattern. Here are some of the answers:

'early'

1. Several people are asked: which of you was it?  
   In sense of "who did that?"
   The speaker asks the listener a question, he knows the answer himself and urges the listener to give him the right answer.
   Reproachful question: the person concerned has to expect a reprimand for some mischief.
   Speaker sounds superior, demanding, tries to be distant.

'medial'

1. Speaker A asks speaker B the name of a third person.
2. Somebody unknown to the speaker is passing by and the speaker asks a listener the name of the unknown person.
   The speaker wants to know something unknown to him, e.g. he has just seen somebody whose name he does not know.
   Neutral, positive question, e.g. a teacher's question: "Charles the Great, who was that?"
   Speaker asks in a familiar way, is on the same level as the person spoken to.

2. Speaker, rather annoyed, asks somebody standing at the open window to shut it.
   Order to shut the window at once.
   Request to shut the window, not the door.
   Friendly request to shut the window, not the door, because, e.g., otherwise grandmother might catch a cold.
Tone of a command, slightly threatening, repeated order to the naughty son, object of shutting is self-evident, could be left unmentioned.

Opposed to "shut the door please," object has to be defined specially.

(3) Speaker asks the listener for specific information to test him, i.e. the speaker knows the answer to the question himself.

Speaker does not know the answer himself and asks the listener to give him the information.

Speaker knows the answer already, e.g. could be a teacher.

Speaker does not know the answer, asks a real question.

Teacher to his class, rhetorical.

After the teacher has provided the explanation a pupil not having heard it asks what an atom is.

Exam question.

Continuation in a chain of questions.

(4) Statement, all settled, routine utterance.

Tomorrow, not today or the day after tomorrow or any other day the speaker might prefer.

At the end of an ordinary conversation, routine.

Speaker mentioned tomorrow for the next meeting and then changed it; to make sure he repeats the new arrangement at the end of the conversation.

Statement.

Confirmation of tomorrow as against the day after.

The meanings that may be abstracted from the dialogue data and the paraphrases for the three peaks are:

(a) *early*: established fact; no room for discussion; final summing up of argument

(b) *medial*: new fact; open for discussion; starting a new argument

(c) *late*: emphasis on a new fact and contrast to what should exist or exists in the speaker's or hearer's idea.

The FO peak differences are thus not associated with stress, which remains the same in all three cases, but with intonation, which is in turn linked to semantic categories expressing the speakers evaluation of facts in respect of expectations. As regards the distinction between 'medial' and 'late' peaks similar categorizations have been proposed for English, the 'late' peak expressing the speaker's incredulity or his uncertainty (Ward &
The phonetic differentiation between the three peaks and the associated changes of meaning point to another instance of what Ohala (1983, 1984) has called the frequency code: low frequencies signal domination, high ones submissiveness. Of course, in the case under discussion, this link has been given linguistic plasticity in two ways:
- the synchronization with the syllable structure, i.e. with human sound articulation,
- a semantic denotation, rather than an expressive meaning.
But the semantics of 'closed vs. open to argumentation' are intimately related to 'domination vs. submissiveness'. It is, however, not necessarily the domination or submissiveness of the speaker that is signalled here, it may be that of the situation or of other communicative partners setting an established fact or leaving the door open for change and new things. These are the basic, underlying meanings of 'early' vs. 'non-early' peaks. The actual meanings observable on the surface in individual utterances and contexts depend on the interplay of these basic semantics of intonation contours with the semantics at the levels of syntactic structures, within and across sentences, and of the lexicon.

If an early peak is used in questions, whose semantics suggest openness, then the question gets special connotations in keeping with the semantics of the early peak intonation: the question is asked with a presumed knowledge of the answer, as in
- the teacher's question "Wer war das?" ("Who did that?" = I'll find out anyway; possible threat)
- the exam question "Was ist ein Phonem?" ("What's a phoneme?)
- the résumé asking for confirmation "Das Phonem ist also eine Lautklasse." ("So the phoneme is a sound class." = Can we keep that in mind and start from there, moving to the next question?)

If an imperative construction gets an early peak, there is again a contradiction between the signalling, through intonation, of the expected completion of an action, and, through syntax, of the order to carry it out. This contradiction produces the connotation of annoyance and impatience at the delay of an action. "Mach' bitte das Fenster zu." ("Shut the window,
please.

The early peak can also get the connotation of resignation because nothing can be done to alter the established facts: "Nun gut. Wie Sie wollen." ("Alright. As you like."). The resignation is all the greater the earlier the F0 fall and the longer the low F0 tail on "gut" and "wollen". In either-or questions, an early peak in second position signals a choice within a closed set of alternatives, whereas a succession of medial peaks with low F0 in between refers to an open set of alternatives, which are simply given as possible examples from a longer list: "Willst du Tee oder Kaffee?" ("Would you like tea or coffee?"). Rising patterns instead of medial peaks convey the same open set but sound less categorical and more friendly.

In the late peak, the preceding low F0 interferes with the openness connotation of the rise and introduces the speaker's difference of opinion, which is rated very high in relation to observable facts. The speaker stresses the difference between his opinion or way of assessing things and the opinion of others or facts or beliefs as to how things should be. This leads to meanings of surprise, incredulity, "that can't be true", insinuation, talking down, changing in degree according to the amount of peak shift to the right. Very often the late peak is combined with modal particles, reinforcing their meanings, such as (word with late peak accent underlined)

"ja" in exclamations
"Da steht ja eine Kirche!" ("Oh, there's a church!"), expressing surprise because reality differs from the speaker's view,

"doch" in statements and imperatives/requests
"Er ist doch gekommen." ("He's come, what are you going on about."), "Setzen Sie sich doch." ("Do sit down."), "You are still standing, it is my opinion that you should be sitting."), expressing opposition to what the speaker is confronted with,

"etwa" in questions
"Hast du das etwa gekauft?" ("You did not buy that, did you."), expressing incredulity, which is all the stronger the greater the emphasis signalled by peak height.

In these examples the modal particle may be missing, but the presence of a late peak still conveys the meaning of a contrast between the speaker's
observation and his opinion on it. In utterances, such as "Ja." ("Yes.") or "Natürlich." ("Of course.") the speaker stresses his own opinion and rejects any opinion to the contrary, producing a supercilious, arrogant, presumptuous undertone. Talking to a child, "Wie heißt du denn?" ("What's your name?") stresses the distance between the speaker and the addressee and gives the impression of talking down. In a sentence like "Hast du bei Christine übernachtet?" ("Did you spend the night with Christine?") there are indications that the addressee has done just what the speaker suggests, but should not have because this clashes with moral standards which the speaker purports to hold, resulting in reproach or insinuation; combined with a high peak it suggests incredulity.

The important lesson to be learnt from these data is that there is a direct link between particular FO contours and specific meanings, but this link is not one on the surface, but underlies the actual meanings, which are the result of an interaction of various meaning levels. Social psychologists (e.g. Scherer, 1985) who have been concerned with these direct substance/expressive meaning relations, have often lacked a detailed insight into the phonetic and semantic structures of language as a prerequisite to a successful interpretation. The corollary of the phonetic-semantic explanations offered for the use of different FO peaks in intonation is that these phonological intonation categories in their association with meanings relatable in one form or another to the basic ones given must be at least widespread in languages, provided the phonological dichotomy has not already been booked at some other level, e.g. tone or word accent.

2.3 General discussion concerning Hypothesis (2)
The perception experiments of 2.1 and the semantic evaluation derived from paraphrasing tasks in 2.2 have largely confirmed Hypothesis (2) of Contribution I (Kohler, 1991b): the shift of an FO peak in a single-accent terminal utterance between a prenucleus and a nucleus position results in a categorical change of perception, which is correlated with an equally categorical semantic switch along the dimension 'established/new' or 'closed/open to argumentation'; the corresponding realignment to the right produces a gradual auditory change correlated with a semantic continuum expressing degrees of distance which the speaker establishes between himself and the world as it presents itself to him. This degree of distance rather
than the degree of emphasis, as formulated in Hypothesis (2), is the semantic basis of the 'medial' to 'late' peak positions, emphasis being correlated with peak height.

3. Intonation and stress

It has already been pointed out that the three FO peak positions discussed in Section 2. represent different phonological categories of intonation associated with the same stressed syllable. So intonation must be differentiated from stress, through which a syllable in a chain is selected and marked for an intonation peak (or valley) to be hooked onto. But the stress feature may be chosen for different syllables in a sequence, and thus a shift of an FO peak (or valley) position from one syllable to another can also change the stress position in a syllable chain, not just the intonation peak (or valley) associated with it. FO peaks can therefore become cues to stress beside being cues to intonation. Then two questions arise:

(a) Under what conditions is an FO peak shift (without concomitant changes in sound duration and intensity) sufficient to shift stress to a different syllable? Two cases have to be distinguished: the stress pattern changes, but the peak pattern stays, or both change. In principle, at each stress position three intonation peaks are possible.

(b) How can the stress and intonation functions of FO peaks be differentiated, and in what ways do they interact?

These questions relate to the level of lexical stress or of sentence stress because words in sentences do not all retain their stresses. 3.1 deals with the former, 3.3 with the latter. In 3.2 the importance of duration for the signalling of stress, in addition to FO, will be discussed. Finally, 3.4 will deal with the perceptual ambiguity between one and two accents combined with conflicting intonation patterns, and 3.5 will enquire into the relevance of intensity for the cuing of stress and intonation.

3.1 Lexical stress

German offers good examples for testing the issues of stress signalled by FO peak position and of stress and intonation interaction at the lexical level because it has minimal verb pairs, with either prefix or stem stress, which can occur in the same natural sentence frame, e.g. "Er wird’s wohl umlagern." [er vierts vol 'uml:agen (uml:a:gen)], with stress either on the
prefix "um-", meaning "verlagern" ("He is presumably going to shift it to another place."), or on "-la-", meaning "belagern" ("He is presumably going to besiege it.").

Utterances of the above two sentences, (a) with stress on "um-" and a 'medial' intonation peak on this syllable, and (b) with stress on "-la-" and an 'early' intonation peak, which is actually located on the syllable "um-", were analysed and Fig. 19 presents the waveforms together with their FO displays. The FO peak positions in the two utterances are practically identical in relation to the syllable structures of "umlagern": they occur at more or less the same time interval just before the beginning of /I/. The differences between the two are in the shapes of the FO peak contours and in the syllable durations. In the utterance with stem stress in Fig. 19b the post-peak FO descent is more gradual, the syllable "um-" shorter (135 ms in Fig. 19b vs. 222 ms in Fig. 19a) and therefore the FO rise faster, starting at a structurally earlier point (beginning of the /I/ in "wohl" rather than at the "um-" syllable onset, as is the case in the utterance with prefix stress). The "-la-" syllables in the two utterances, on the other hand, have very similar durations in the stem and prefix stress words (268 ms in Fig. 19b vs. 258 ms in Fig. 19a). Two further stimuli were generated from the two illustrated in Fig. 19 by exchanging the FO contours (see Fig. 20). These four stimuli (ST1 - ST4) were the basis for creating four series of FO peak positions (P1 - P4):

P1 A series of 12: 6 left shifts (parallel transposition of the left branch and time expansion of the right branch) and 5 complete parallel right shifts of 30 ms each in the utterance of Fig. 19a.

P2 A series of 9: 8 complete parallel left shifts of 30 ms each in the utterance of Fig. 19b.

P3 A series of 12 in the utterance of Fig. 20a, following the procedure in P1.

P4 A series of 9 in the utterance of Fig. 20b, following the procedure in P2.

P1 and P3 are based on the original prefix stress, P4 and P2 on the original stem stress utterance, and in each pairing the series form an opposition between more abruptly and slowly falling FO peak contours, respectively. From these four sets of stimuli two tests were compiled: Test I combined
Fig. 19
Speech waves and FO contours (linear scale) of the original (a) prefix stress with 'medial' peak and (b) stem-stress with early peak in "Er wird's wohl umlagern." A, B, C mark the FO base and peak points for peak contour shift.
As in Fig. 19, but with exchanged F0 contours, adjusted to the different timing of the new utterance.
the more sharply falling sets P1 and P4, Test II the slowly falling sets P2 and P3. Subjects were asked to identify the stimuli with the meanings of either "belagern" (stem stress) or "verlagern" (prefix stress). Further details about test stimulus generation, test tape construction and test administration can be found in Kohler (1990c).

In P1 and P3 the series of FO peak positions straddle the syllable structures where a change from prefix to stem stress is to be expected if FO is a sufficient cue. The two sets differ in that the peak shape of P3, but not of P1, approximates the more slowly descending FO configuration found in the early peak of the original stem-stress utterance (cf. Fig. 19b). It is hypothesized, therefore, that if stress is perceptually shifted at all in P1 and P3, there will be a more clear-cut change in P1 because there is a higher probability in P3 that an FO peak position on "um-" can not only be perceived as a 'medial' or 'late' peak prefix stress but also as an 'early' peak stem stress. Similarly, there would be a greater likelihood in P2 than in P4 for an 'early' peak stem stress to interfere with a 'medial' or 'late' peak prefix stress because of the slower FO descent and its time expansion in the left shift of P2 as against P4.

Results and Discussion

Figs. 21 and 22 present the data of the two identification tests for the original prefix and stem stress series, respectively, each with slow and more sharply falling peak contours.

In the shift of the more sharply falling FO peak contour through the original prefix-stress utterance there is a clear change from initial to stem stress, in spite of the duration of "um-" pointing to the former. FO can thus override duration, particularly since the duration of the unstressed "-la-" syllable in the original utterance is very close to its duration under stress. In stimulus 10, which is the first in the ordering from 1 to 12 to yield an unequivocal stem-stress categorization with over 80% positive responses, the FO peak position is 30 ms into the vowel of the syllable "-la-". This corresponds to the data discussed in Section 2., concerning the change from an 'early' to a 'medial' intonation peak on the stressed syllable. The fact that the change from one stress position to the
Fig. 21
Percentage stem-stress responses for "umlagern" (= "belagern", i.e. stem stress) in the series of 12 F0 peak positions (from left to right) combined with the original prefix-stress utterance of "Er wird's wohl umlagern." (nr. 7 appr. original peak position). Broken line = P3, slowly falling peak contour (n = 80 at each data point), continuous line = P1, sharply falling peak contour (n = 185 at each data point), dotted line = P1, sharply falling peak contour, but in Test III of 3.2, see text (n = 170 at each data point).

other is gradual rather than categorical can be related to a residue of the duration cue. But we also have to consider some interaction of the stress and intonation functions of F0 because the F0 peak assumes positions before the beginning of the syllable nucleus /a:/ of "-la-" which can simultaneously function as the 'medial' or 'late' intonation peak in stressed "um-" and as the 'early' intonation peak in stressed "-la-". The relevance of this intonation interference with stress is confirmed by the finding that when the more slowly falling F0 peak is substituted the initial-stress category is not so clearly represented: the interpretation of
Fig. 22
Percentage stem-stress responses for "umlagern" (= "belagern", i.e. stem stress) in the series of 9 FO peak positions (from left to right) combined with the original stem-stress utterance "Er wird's wohl umlagern." (nr. 9 appr. original peak position). Broken line = P2, slowly falling peak contour (n = 80 at each data point), continuous line = P4, sharply falling peak contour (n = 185 at each data point), dotted line = P4', sharply falling peak contour and durations of prefix stress (cf. Test III of 3.2, n = 170 at each data point).

an 'early' intonation peak for stem stress is then never completely precluded.

When an FO peak contour is shifted through the original stem-stress utterance there is no change between the stress categories (Fig. 22): the answers remain predominantly in favour of stem stress. In this case, FO cannot override the duration cue completely because "um-" is too short in relation to "-la-" to signal initial stress. There is some effect of FO when
the more sharply falling FO peak occurs within the syllable "um-". In stimuli 1 to 5 the FO peak has been shifted leftward all the way into the preceding syllable "wohl", whereas in 6 to 8 it has been moved only as far back as some point within the prefix syllable "um-", and in these stimuli there are up to 30% judgements of prefix stress. This pattern suggests that the overriding salience of duration in the original prefix stress stimulus is checked somewhat when the characteristic sharply falling contour occurs in the relevant syllable and is more narrowly limited to it, allowing the interpretation of a 'medial' or 'late' peak on "um-", rather than an 'early' one on the following "-la-". In the other series, however, the slowly falling and time-expanded FO contour reduces the probability of interpreting the peak as a 'medial' or 'late' peak for a prefix stress, because of the stronger interference from an 'early' peak interpretation on "-la-", due to the wider span of the FO peak descent.

The questions asked initially can now be answered as follows:
(a) An FO peak shift by itself is sufficient to bring about a clear change from one stress position to another, provided the duration of the stressed-syllable-to-be toward which the FO peak is shifted is not too short. But even when it is, there is a residual FO effect.
(b) The intonation function of FO interferes with its stress function if the latter is not supported by duration. This finds its expression in a gradual change from one stress position to another in abutting syllables where an ambiguity can arise between a 'medial' or 'late' intonation peak in one stressed syllable and an 'early' intonation peak related to a subsequent stressed syllable. This interaction is strengthened when the shape of the FO peak contour approximates the more slowly falling one of the 'early' intonation peak of a later stress.

3.2 Duration as a feature in stress perception
It has been shown in 3.1 that although FO is a strong cue in stress perception, duration can become an additional distinctive feature when vowels and postvocalic sonorants are shorter than would be associated with the production of a stressed syllable. On the other hand, if they are longer than would be associated with an unstressed syllable, the FO cue may be disturbed, but never dominated by the duration cue.
3.2.1 Duration increase for inducing stress perception in F0 peaks

The importance of duration for stress perception was further investigated in an experiment that repeated Test I of 3.1 by using the peak series P1 and a modified peak series P4', i.e. the sets of stimuli based on the original prefix and stem stress utterances, respectively, both combined with the more sharply falling F0 contour derived from the prefix-stress utterance (see Figs. 19a and 20b). But this time a new basis stimulus ST4' for a series P4' was created by adjusting the durations of the syllable "um-" [um] and the vowel [a:] of the syllable "-la-" in the basis stimulus ST4 to the same values as in the basis stimulus ST1. By repeating some periods in [um] and deleting some in [a:], [u] was lengthened from 70 ms to 117 ms, [m] from 65 ms to 105 ms, and [a:] reduced from 210 ms to 189 ms. Then the F0 contour of the basis stimulus ST1 was transferred - sound segment by sound segment - to the modified basis stimulus ST4'. The series P4' was generated by shifting the F0 peak to the left as for P4.

Series P1 and P4' were then compiled to a new Test III, which only differs from Test I in the segment durations of P4' vs. P4. The first 7 stimuli of P1 and the last 7 of P4' occupy the same ranges of F0 peak positions, have very similar segment durations (with [um] and [a:] being identical) and comparable F0 contours, but they differ in the basis stimulus, which is either the original prefix-stress utterance in P1 or the original stem-stress utterance in P4', implying spectral and intensity differences. The hypothesis connected with Test III was that the change of segment durations in P4' vs. P4 would be sufficient to reverse judgement from stem stress to prefix stress in all cases of the series, resulting in similar response functions for stimuli 1 - 7 of P1 and for stimuli 3 - 9 of P4', and would thus point to the low relevance of spectral and intensity features in German stress perception. Test III was run with 34 listeners.

Results and Discussion

The dotted lines in Figs. 21 and 22 present the results of identification Test III. The hypothesis of the complete reversal of judgements has been confirmed by P4 and P4' in Fig. 22 yielding ca. 80% and 20% "belagern" responses, respectively. The left shift of the response function for the identical P1 series in Test III, compared with Test I, may be due to the
test design: the decrease of the number of clear stem-stress cases and the increase of the number of clear prefix-stress cases by swapping P4' for P4 may have pushed the responses to the more ambivalent cases in P1 in the direction of stem stress, but there is also more noise in the P1 response curve of Test III, as is shown by the offset of 10% - 20%.

3.2.2 Duration decrease for eliminating stress perception in FO peaks
Parallel to generating ST4' from ST4, a new ST1' was generated from ST1 by shortening the durations of [um] to 70 ms - 65 ms (from 117 ms - 105 ms) and of [a:] to 210 ms (from 189 ms), applying the same period splicing procedure. Then the same peak shifts to the left and right were performed as in P1, resulting in P1' with 12 peak positions and sharply falling FO contours. Informal listening to the series P1' by phoneticians established that all the 12 stimuli were unequivocally perceived as stem stressed, even when the FO peak position was on "um-". Because of this very clear evidence no further formal test was run. These results prove again that if the duration of a stressed-syllable-to-be is too short the FO cue may not be sufficient to signal stress.

3.2.3 Conclusion
In German, stress is cued by two features, FO and duration, which may be expressed in a distinctive feature notation as ±FSTRESS, ±DSTRESS. The FO cue clearly dominates if the duration is not too short for stressed syllables; otherwise longer duration is required to signal stress. Syllables are thus marked as stressed/unstressed by the two stress features: (1) -FSTRESS, -DSTRESS = unstressed, (2) -FSTRESS, +DSTRESS = secondary stress, e.g. in non-initial components of compounds ("Ausfahrt" ['aus fa:vt] ("exit"), which receive increased duration, but no intonation peak (or valley), (3) +FSTRESS, +DSTRESS = primary stress, where the intonation points are hooked. The intonation associated with stressed syllables is, among other things, defined according to different peak positions, which may again be expressed in distinctive feature notation taking the primary dichotomy between 'early' and 'non-early' into account: ±EARLY, and -EARLY may then be ±LATE.

At each potential stress position +FSTRESS, three intonation peaks are possible. But since the FO of these peaks serves to signal the stressed

173
syllable - as a stress cue - and at the same time the peak position in relation to such a stressed syllable - as an intonation cue, there may be interference between the two cue functions leading to ambiguity, if the temporal distance between successive potential stresses, as in lexical items of the type "umlagern", is small, particularly because of a lack of intervening unstressed syllables (e.g. containing /a/) and even more so in the case of abutting syllables with short quantity vowels.

3.3 Sentence stress
In sentences not every lexical item gets a +FSTRESS marking for the association with intonation peaks (and valleys), although at a more abstract level it has lexical stress, i.e. at least one syllable is phonologically marked as having the potential of receiving the features +FSTRESS and +DSTRESS. The rules of grammar and pragmatics determine which lexical-stress syllables are given the feature combinations +FSTRESS, +DSTRESS or -FSTRESS, +DSTRESS in sentences. In a sentence such as "Aber der Leo säuft." [aːɐ dɐ 'leːoː 'zɔːft] ("But Leo drinks.")⁶ either the subject "Leo" or the verb "säuft" may be in focus, receiving the features +FSTRESS, +DSTRESS, or both elements may be so characterized simultaneously. The question to be answered is whether the findings at the lexical level in 3.1 - 2 can be replicated at the sentence level, viz. whether a switch from one stress position to another can be brought about simply by F0 peak shift through the sentence. In this case it will also have to be checked whether at some intermediate section of the peak shift scale both stresses are realised. And finally, there is the issue of the perceptual manifestation of different intonation peaks ('early', 'medial', 'late') at each stress position, in parallel to what was found in the sentences of Section 2. with only one potential accent.

3.3.1 Stimulus preparation for perception experiments
A natural production of the utterance "Aber der Leo säuft." with sentence stress and 'medial' intonation peak on "Leo" was used for stimulus generation. Fig. 23 shows the speech wave, energy and F0 contours. A series

⁶ This sentence played an important role in some experiments of the Munich Intonation Project (see Altmann et al., 1989) and was taken as the basis of further experiments in the Kiel Intonation Project for purposes of cross-reference.
Fig. 23
Speech wave, energy and FO contours (linear scale) of the utterance "Aber der Leo säuft." with subject stress and 'medial' peak. The time marks indicate the FO base and peak points for peak contour shift.

of 7 left shifts (parallel transposition of the left branch and time expansion of the right branch) and of 11 complete parallel right shifts of 30 ms each were generated on the basis of the utterance in Fig. 23. An informal assessment of the series detected a poor quality in the synthesis of the segment /z/ and of too strong a final aspiration; furthermore, the last stimuli of the series, from 15 to 19, with the accent on "säuft" sounded too strong at the beginning and husky at the end, obviously due to the wrong energy contour for a final FO peak position, i.e. to a desynchronization of FO and energy (see 2.1.1.5). To remedy these defects and to create as natural synthetic versions as possible, almost the entire [z] was devoiced, the final aspiration reduced by lowering the dB-values, and the discrepancy between energy and FO eliminated by lowering the energy
As Fig. 23, but with the 19 peak position points marked. The peak series was then regenerated with these parameter modifications of the stimulus in Fig. 23; it formed the basis for identification and serial discrimination tests.

3.3.2 Identification test
Five repetitions of the 19 stimuli were randomized and presented (in the format of 1.4 (2) for single stimuli) to 31 listeners with the instruction to decide whether "Leo" or "säuft" was more strongly stressed.

Results and Discussion
Fig. 25 presents the results of the identification test, which demonstrate very clearly that a simple FO peak shift causes a change from subject to verb stress. The transition in the response function between the two stress positions indicates - as was confirmed in phonetic expert listening - that
as the peak is moved into the fricative [z] and therefore spans both words, giving a late F0 rise to "Leo" and an early F0 fall to "säuft", the perception of double stress results, which disappears again when the peak is located at the beginning of the vowel of the verb and the impression of focus stress on the latter is created.

Fig. 25
Identification function showing percentage 'subject stress' judgements for 19 stimuli "Aber der Leo sauft." with F0 peak shift from left to right. (nr 8 appr. original peak position). n = 155 at each data point.

3.3.3 Serial discrimination tests
The series of 19 stimuli was partitioned into two sub-series: (a) stimuli 1 - 10 representing clear instances of the category of subject stress and (b) stimuli 14 - 19 representing clear instances of the category of verb stress, according to the results of the identification test. Each set in ascending (numerical) ordering was presented to 32 subjects for evaluating at which
stimulus in the series the first and further changes in the speech melody had occurred.

Results and Discussion

Tables XII and XIII present the results of the serial discrimination tests (a) and (b).

Table XII
Frequency distribution of 'change has occurred' responses of 32 listeners in the left-right sequence of the serial discrimination test across the first 10 stimuli with F0 peak shifts in "Aber der Leo säuft." (1 = left-most, 10 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>1</td>
<td>5</td>
<td>12</td>
<td>10</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Further changes perceived</td>
<td></td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1</td>
<td>5</td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

(2 listeners perceived no change at all.)

Table XIII
Frequency distribution of 'change has occurred' responses of 32 listeners in the left-right sequence of the serial discrimination test across the last 6 stimuli with F0 peak shifts in "Aber der Leo säuft." (14 = left-most, 19 = right-most position)

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
</tr>
</thead>
<tbody>
<tr>
<td>First change perceived</td>
<td>16</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Further changes perceived</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>16</td>
<td>3</td>
<td>6</td>
</tr>
</tbody>
</table>

(3 listeners perceived no change at all.)

In both series the first perceptual change has a maximum frequency at the stimulus in which the F0 peak occupies the first position within the respective syllable nucleus (nr 5 in (a) and nr 16 in (b)). This result coincides with the data obtained in the peak alignment test in utterances
containing a single potential accent (cf. 2.1.1). It points to the change from an 'early' to a 'medial' peak within each stress position.

A corresponding clear-cut switch was not observed in the "umlagern" series of 3.1. The reason for this difference lies in the shorter duration of [um] vs. [le:o:], which allows less separation of the intonation peak and stress positions and causes the F0 configuration to straddle both potential accent syllables, given the width of the shifted peak contour, across a greater number of stimuli. The more gradual transition from prefix to stem stress in the response function of Fig. 21, compared with that in Fig. 25, is a further indication of this stronger stress/intonation interaction across segment durations that are insufficient for restricting the chosen peak timing to. To achieve a greater separation of the different intonation peaks within each accent, the peak descent would at least have to be faster to encroach less on the other peak and stress positions.

3.4 Perceptual ambiguity between single and double accent
In spite of the more adequate temporal structure in "Aber der Leo säuft.", for separating the theoretically possible peak and stress positions, there is still an ambiguous transition period between the two potential accents, as shown in Fig. 25. And as was argued in 3.3.2, this ambivalence is not so much between either subject or verb focus stress, but between subject focus and double stress. In the latter case, the late rise on "Leo", followed by an early fall on "säuft", may be interpreted as belonging to two F0 peak configurations - 'late' followed by 'early' -, without an intervening dip between the two, or as a single 'late' F0 peak on the subject. In the first case, two accents are perceived, in the second only one. Because of the still close temporal proximity between the two potential stress positions, there must be a stretch along the peak shift scale where the signal is ambivalent between these two interpretations. That we are here dealing with a confusion of subject focus stress and double stress is proved by expert listening to the series of 19 F0 peak shifts in "Aber der Leo säuft.", establishing stress on "Leo" in stimuli 12 - 14, which may or may not be

---

7 The relevant serial discrimination tests were carried out but are not reported here in detail. The results were negative so that the summarising statement is considered sufficient.
accompanied by stress on "säuft". In stimulus 15, however, the change to focus stress on the verb has taken place: the peak rise is now far enough away from the potential accent syllable in "Leo" and therefore no longer associated with the subject, FO being low during the whole of the word "Leo".

The perceptual ambiguity between a single 'late' peak and a 'late' + 'early' peak combination is even stronger in cases where two potential accent syllables abut and the first contains a short vowel, as in "Der Ring glänzt.", as is shown in Contribution IV (Hertrich, 1991a). Even when in abutting accents the first vowel is long, or when a short or long vowel in the first potential accent position is followed by one unstressed vowel (/ə/ or /ɛ/), as in "Die Uhr tickt.", "Die Bremse quietscht.", "Die Maler malen." (see Hertrich, 1991a), a perceptual confusion between the two categories is possible. The confusion can be avoided if for the single 'late' peak the descent is rapid to avoid trespassing on the second accent syllable domain, as was demonstrated for "Die Maler malen." (loc. cit.). So if the temporal distance between two potential accents is short enough, the FO peak shift through the sequence produces perceptual changes from subject focus stress to dual stress to verb focus stress. And in the transition area between the two focus stresses, perception may be ambiguous between double and single first accents. This ambiguity disappears as the distance between potential accents gets longer, as in "Die Bäcker haben gebacken." or "Die Sekretärin hat die Briefe geschrieben." (loc. cit.).

In accent sequences at longer distances from each other double stress does not occur by simple FO peak shift through the utterance; the peak contour has to be broadened at the same time to realise a 'medial' or 'late' rise on one accent syllable and an 'early' fall on the next one. In between these two intonation turns - rise and fall - associated with two stressed syllables, there may be an FO dip of various degrees of extension, to generate two properly manifested FO peak contours, or the two peak points are joined by a plateau or a slight monotone descent/ascent, creating a 'hat pattern' (cf. Cohen & 't Hart, 1967). Although the 'hat pattern' is perceptually and semantically different from a succession of complete peaks (as is shown in Contribution VI, Hertrich, 1991b, see also Contribution VII, Kohler, 1991d), there are strong arguments in favour of treating a 'hat
'Pattern' as a succession of two peaks without an F0 dip:

1. The timing of the initial rise is exactly the same as the rising part in a 'medial' or 'late' peak. There are rising patterns that are timed more slowly and have rises up to the beginning of the next stressed syllable (see Contribution VII, Kohler 1991d). They have to be recognised as separate entities. So we would have to set up two rising patterns - slow and fast - but since the latter coincides with the rising part of the peak pattern it is more economical to have no new units 'fast rises'. The complementary solution to regard 'medial' or 'late' peaks, too, as being composed of two tonal entities each - rise and fall - is ruled out by the fact that they constitute one stress, whereas the 'hat pattern' rises and falls represent two stresses.

2. The timing and syllable alignments of the final fall coincide with the falling section of an 'early' (or 'medial') peak.

3. 'Hat patterns' can be derived from the corresponding dipped peak sequences by general phonetic rules changing the prominence relationships between the first and the second peak as a consequence of removing phonetic features characteristic of the definitions of the different F0 peaks. Two cases can be distinguished:

(a) In the sequence 'medial' (or 'late') + 'early' peaks, the elimination of the F0 dip does not affect the essential feature of the low falling F0 in the 'early' peak and also preserves the characteristic (low level +) rise in the 'medial' (or 'late') peak (see 2.1.1.7), but it modifies the complete manifestation of the latter by removing the separate F0 descent, thereby reducing its prominence.

(b) In the sequence 'medial' (or 'late') + 'medial' (or 'late') peaks, the elimination of the F0 dip results in a loss of the 'medial' or 'late' characteristics of the second peak because in a derived 'hat pattern' it lacks the essential F0 rise in the syllable nucleus (see 2.1.1.7), and since it cannot be associated with an 'early' peak either, not having the early low fall, it lacks the prominence-lending feature of the 'medial' peak rise as well as of the 'early' peak fall. But since on the other hand, the first peak has it, the prominence of the second one is subordinated. Thus a principled relationship can be established between 'hat patterns' and peak sequences on the basis of general phonetic rules modifying the relative prominences of the peaks.
In both cases (3a) and (3b), the generation of a 'hat pattern' from a dipped peak sequence does not change the number of accents, but only the prominence relations between them. Thus when the sentence "Die Wählerinnen wählen." is combined either with a 'hat pattern' consisting of a medial (or late) rise on "Wählerinnen" plus a medial fall on "wählen", or with a single 'medial' (or 'late') peak on "Wählerinnen", only the second intonation represents focus stress on the subject and deaccentuation of the verb (see also Contribution VI, Hertrich, 1991b).

3.5 Intensity in the cuing of stress and intonation

The question now arises as to whether it is possible to change stress perception simply by varying intensity. Two test cases may be distinguished:

(a) Utterances that are ambiguous between one and two stresses in FO peak shifts, such as "Aber der Leo säuft." in 3.4,
(b) 'hat patterns' in which a medial (or late) FO rise is immediately followed by a medial FO fall, reducing the prominence of the second stress compared with the sequence of two complete peaks (cf. 3.4).

If intensity alone can change stress perception, then it should be possible in (a) to produce a switch from double to initial focus stress simply by reducing the intensity in the second accent syllable and by simultaneously raising it in the first. Similarly in (b), it should be possible to alter the prominence relation by a comparable intensity adjustment in the two accent syllables.

The issue has been tested interactively by changing the source amplitude values accordingly in the RULSYS TTS synthesis-by-rule. The result has been negative: the focussing, and consequently the number of stresses or the prominence relation, does not change. It is more the relative loudness that is affected (see also Kohler, 1991f). This is further support for the long-established finding that intensity has a low signalling value for stress compared with FO and duration (Fry, 1958).

The situation is different as regards the contribution of intensity to the perception of intonation. Again two cases may be distinguished:

(a) It has already been discussed in 2.1.1.5 that a late FO peak pattern requires a parallel timing of the intensity course to guarantee its perceptual identity.
(b) It is argued in Contribution V (Kohler & Gartenberg, 1991) that lower intensities around the F0 peaks in 'early' and 'late' patterns vis à vis 'medial' ones have to be offset by higher F0 to provide the same prominence across the different intonations. On the other hand, the 'early' peak pattern, which accentuates low F0, has its characteristics strengthened by not having a lower intensity around its prenucleus F0 maximum compensated for in a higher F0 peak value.

Finally, the disruption of the natural parallelism in the time courses of F0, source amplitude and sound intensity for the three terminal peak contours, as it is caused by the synthesis of F0 peak shifts across an original 'medial' peak utterance, may result in a degraded acoustic output quality. So, when a natural 'medial' peak speech signal of "Sie hat ja gelogen." is taken as a point of departure for LPC synthesis with a 'late' peak, the stress and intonation categories are signalled correctly, but the utterance sounds husky at the end and overloaded in the middle because F0 and intensity diverge in opposite directions in these two places. To improve the synthesis quality of 'late' peaks appropriate corrections at these points in the intensity curve had to be carried out for "Aber der Leo säuft." in 3.3.1 (see also Kohler, 1991f).

3.6 General discussion concerning Hypothesis (3)
The perception experiments of 3.1-5 have largely confirmed Hypothesis (3) and its corollaries of Contribution I (Kohler, 1991b). If there is more than one potential accent in a single-accent terminal utterance - either at the lexical or at the sentence level - three phonological intonation categories - 'early', 'medial', 'late' peaks - are distinguished at each stress position, provided the temporal distance between the accent places allows the separation of the F0 peak configurations. Furthermore, an F0 peak shift alters the stress position as well, which can result in an interaction of stress and intonation if two accent syllables occur at such a short duration interval that the rising and falling branches of a peak contour can be at the time associated with a single peak on the first accent syllable or with a succession of two peaks on two successive accent syllables, not separated by an F0 dip. This ambivalence of a stimulus between single and double stress results in a perceptual ambiguity between, e.g., prefix and stem word stress at the lexical level, or subject and verb stress at the sentence
level. It is only when the F0 peak is moved out of the joint domains of both accent syllables in order to be exclusively in that of the second one that the ambiguity is resolved and second position focus stress results. A link has thus been established between 'hat patterns' and dipped F0 peak sequences, based on prominence relationships, as postulated by Hypothesis (4) in Contribution I. This point will be further discussed in Contributions VI (Hertrich, 1991b) and VII (Kohler, 1991d).

Duration is a further cue to stress in German, but usually subordinated to F0, unless it is too short for what is to be expected of stressed vowels. Intensity and spectral characteristics, on the other hand, do not seem to play a role in stress perception. Intensity intervenes as an important cue to intonation identity and to voice (speech) quality when the usually parallel time courses of F0 and intensity are disrupted, and it is, of course, the signal attribute of loudness. Finally, the height of an F0 peak cues prominence at the perceptual and emphasis at the semantic level (see 2.3, and Contributions I, V and VII, Gartenberg & Panzlaff-Reuter, 1991, Section 6.; Kohler & Gartenberg, 1991; Kohler, 1991d).

4. Conclusions for the Kiel Intonation Model of German (KIM)
The results of the experiments discussed in this Contribution III suggest a number of points that have to be taken into account in KIM as regards the intonation peak component of the model.
1. KIM must comprise the phonetic intonation model proper and the syntactic, semantic and pragmatic environment providing interpretations for symbolic representations of sentences as input to the model.
2. In particular, this environment must specify the lexical items that are to receive sentence stress, and it must provide semantic interpretations along the dimensions 'established/new', 'degree of distance between the speaker and the world', and 'emphasis'.
3. The basic categories of the phonetic model include
   (a) a feature specification of stress with reference to the signal properties F0 and duration: ±FSTRESS, ±DSTRESS,
   (b) a feature specification of intonation with reference to F0 peak position: ±EARLY, and ±LATE within -EARLY,
   (c) the timing of the intonation peaks depending on syllable structures (mono/polysyllables, long/short vowels, voiced/voiceless consonant
environment),
(d) a numerical scale of peak height with reference to degrees of prominence,
(e) IFO and CFO modifications of the basic peak contours,
(f) intensity adjustments to guarantee parallelism with FO time course.
4. After the introduction of the peak categories the model has to deal with their concatenation.
(a) An FO descent from a peak position can be fast or slow. In the latter case double accentuation may result, or a main accent followed by a secondary one, e.g. in "Er hat einen Brief geschrieben." ("He's written a letter.") the final participle is deaccented in relation to "Brief", which gets the main nuclear sentence stress. But the deaccentuation may result in a default secondary stress, or in no stress at all suggesting a contrast between, for example, "Brief" and "Karte" ("card"). This is the same phenomenon as what Kingdon (1965, p. 195) has called 'semantic partial stress' with reference to compounds of different degrees of semantic unity, e.g. "butter cup" (cup for butter) with secondary stress on "cup" vs. "buttercup" (ranunculus) with unstressed "cup". The phonetic manifestation of this difference is not only one of duration, but, first and foremost, of different timings of the FO fall from the FO peak.
(b) Besides peak sequences various 'hat patterns' have to be generated and the semantic and pragmatic differences evaluated.

These points will be developed in Contribution VII (Kohler, 1991d), supplemented by further model components derived from the empirical data collections in the other contributions and from interactive RULSYS TTS experimentation.