Timing and communicative functions of pitch contours

Klaus J. Kohler

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

Running title: Timing and function of pitch

Prof. em. Dr. Klaus Kohler
Institut fuer Phonetik und digitale Sprachverarbeitung
der Universitaet Kiel, D-24098 Kiel
Tel.: +49 431 880 5297, Fax: +49 431 880 1578
e-mail: kk@ipds.uni-kiel.de
Abstract

A new research paradigm is applied to F0 synchronization with articulation, in peak and valley contours, under four principles: (a) Timing of F0 contours enters the definitions of the pitch categories. (b) These phonetic exponents are linked to communicative functions. (c) The listener plays a pivotal role. (d) Contextualization of test stimuli is essential for pitch data collection. Data are presented and interpreted from an experimental investigation of the substance – function relationship in the perception of peak and valley shift series by German listeners, using the semantic differential technique. The findings of the substance – function relationship are explained with reference to the frequency code (Ohala 1983) and to auditory enhancement (Diehl 1991), viz. syntagmatic contrast of high – low or low – high pitch trajectories in synchronization with articulatory landmarks of accented syllables.
Research background and new questions

The phonetic alignment of selected F0 points of pitch contours with articulatory landmarks of syllables has been investigated in a considerable number of papers, especially F0 peaks in (L+)H* pitch accents, often combined with L-L% boundary tones, i.e., in global rising-falling peak contours (cf. Atterer and Ladd 2004 for a recent account of German and English). The aim of these studies is to associate abstract and timeless pitch accents, established by the auto-segmental-metrical approach and ToBI notation, with the articulatory timing of syllabic structures, i.e., segmental anchoring of the phonetic manifestations of phonological tones. A second goal is to see whether the same pitch accent is aligned differently in different languages or in different dialects of the same language. The result of looking into this second question is, e.g., a hierarchy of progressively later peak alignment in the comparison of English/Dutch, North German, South German, Greek. These studies have been carried out on the assumption that the postulated underlying pitch accents are the same phonological entities in the prosodic systems of the compared languages, which has, however, never been ascertained independently.

In these investigations into alignment, the communicative functions of timing variations are no issue because the phonological pitch accent categories are set up outside the time scale, and are projected onto it post hoc to fill the linguistic categories with phonetic measurement in the framework of Laboratory Phonology. This is diametrically opposed to the much older investigation into pitch–articulation synchronization which aims at establishing contrastive categories of pitch alignment for different communicative functions in a particular language through experimental analysis. In this framework of Experimental Phonology, Kohler (1987, 1991) showed for German that three phonological peak positions—early, medial, late—need to be distinguished, with clearly differentiated semantic/pragmatic functions—knowing vs realising vs realising in contrast to one's expectation. It was demonstrated that the peak shift from an early F0 synchronization to a medial one produces a categorical change in perception: there is quite a sharp transition between categories, identified by the matching of test stimuli to a context sentence, and this switch is coupled with increased discrimination across the category boundary. The shift from medial to late, on the other hand, does not show this coincidence of sharp identification and discrimination functions, although opposite ends of the scale are attributed to clearly differentiated categories. This experimental paradigm in the perception of a peak shift series has been applied repeatedly and always produced the same results.

Similarly, Pierrehumbert and Steele (1987, 1989) presented experimental data for two rise-fall-rise intonation patterns in American English, which differ in their alignment with the stressed syllable—medial vs late, the former expressing 'contrastive assertion', the latter 'incredulity'. But they used a different paradigm: their subjects had to imitate the aurally presented isolated stimuli from the peak shift series, which lacked a functional setting. The categorization was then based on the clustering of the F0 contours in the subjects' productions, after having gone through the perception–production filter.

The perception paradigm was also applied to a corresponding valley shift (Niebuhr and Kohler 2004). A categorical change was absent altogether, but different category perception of stimuli from opposite ends of the early to late scale was nevertheless clearly present. This led to the postulate that categorical pitch perception is a special case of perceptual speech processing (Kohler 2004b), related to contrastive F0 trajectories found in early vs medial peaks: high-low (descending from the high focal F0 point) vs low-high (rising to the high focal F0 point) across the articulatory landmark of the consonant–vowel transition and its increase in acoustic energy. This special articulatory link is absent from the medial to late peak contin-
uum, which is defined by a low-high trajectory inside the vowel. It is also completely absent from the valley continuum because the essential pitch change occurs after the landmark: the high focal point is right at the end of voicing and the F0 trajectory from vowel onset to this high focal point determines pitch perception by its starting level and its shape (convex, linear-like, concave) in addition to its end point.

Categorical perception is thus not necessary for the cognitive evaluation of intonation. In spite of the absence of sharpened sensitivity at some point in the medial to late peak and early to late valley continua, the end points are clearly attributed to different semantics and pragmatics. What is required are different ranges of prosodic alignment associated with different semantic/pragmatic categories. These ranges are the phonetic variants of different phonological synchronizations, with some variants acting as prototypical exponents of the linguistic and communicative categories.

It has also been found in informal listening to data (first pointed out by Oliver Niebuhr in the author's Prosody Colloquium at IPDS Kiel) that the medial to late peak sequence contains a further prototypical late-medial domain, which has a more substantial F0 rise in the accented vowel than the prototypical medial peak but lacks the low precursor to the late rise in the accented vowel for the prototypical late peak. This late-medial peak, like the late peak, is associated with contrast to one’s expectation, but signals matter-of-fact statement of this contrast, whereas in the late peak the speaker adds a personal expressive evaluation to the contrast. For example, in the situational context of looking at old photos, the utterances Er war mal schlank. "He used to be slim.", Sie war mal zierlich. "She used to be neat.", Sie war mal hübsch. "She used to be pretty." convey either an observation or a contrast or a contrast with expressive evaluation, depending on the realization with a medial, a late-medial or a late F0 peak (cf. audio and graphic illustrations, as well as interpretations, in Kohler 2004a).

There are strong indications that the same perceptual and semantic-pragmatic partitioning of the medial to late peak series also applies to English, over and above the separation into early and non-early. The importance of F0 timing in relation to spectral change and spectral stability of accented vowels for pitch category formation by the listener was also investigated by House (1990) from a general psychophonetic point of view with logatome material.

Against this research background this paper asks new questions with three goals.

1. It applies a new research paradigm to the synchronization of peak and valley contours with articulation under four principles: (a) Timing of F0 contours enters the definitions of the pitch categories. (b) These phonetic exponents are linked to communicative functions. (c) The listener plays a pivotal role. (d) Contextualization of test stimuli is essential for pitch data collection.

2. It presents and interprets data from an experimental investigation of the substance – function relationship in the perception of peak and valley shift series by German listeners, using the semantic differential technique. Specific questions are:

- Is the partitioning of a peak synchronization continuum into early – medial – late-medial – late mapped onto semantic scales, especially as regards such features as "concluding", "accepted", "unexpected", "contrastive", "surprised"?

- In what ways do medium and high rising valleys differ semantically?
- How do valley synchronization and rise shape influence the semantic interpretation of valley contours?
- How do height of rise end point and synchronization/shape of rise interact in their semantic interpretation?

3. It aims at explaining the substance – function relationship with reference to the *frequency code* (Ohala 1983) and to *auditory enhancement* (Diehl 1991), viz. syntagmatic contrast of high – low or low – high pitch trajectories in synchronization with articulatory landmarks of accented syllables.

2 Method

2.1 Generation of test stimuli
The German sentence *Er war mal mager.* "He used to be thin." was produced naturally by a male speaker (the author) with a single pitch accent on the adjective. The pitch accents were a medial peak and a late low-to-high valley. Three stimulus series, a peak and two (medium and high ending) valley shift continua across the early-to-late scale, were generated in praat, from the corresponding falling and rising originals, respectively. The two original stimuli selected for the generation of the falling and rising series have very similar durations of the total sentence and of the word *mager*: 990 ms and 546 ms for the falling pattern, 1040 ms and 560 ms for the rising one. They differ more substantially in the duration of the accented vowel [aː]: 260 ms and 220 ms for falling and rising, respectively. This means that the durations of the stressed and the unstressed syllable are distributed differently, a longer stressed vowel going with the substantial fall in the first syllable (endpoint 83 Hz), a longer unstressed vowel with the extensive rise on the unstressed syllable (endpoint 243 Hz).

In the F0 generation the falling pattern has the following precursor before the peak contour, adapted to microprosodic variation: 112 – 98 – 110 – 90 Hz. The peak is symmetrical 90 – 150 – 90 Hz, with 120 ms duration of both rise and fall. F0 then trails off to 73 Hz at the end of voicing. In the starting position for the F0 shift, the peak maximum is at the boundary between the initial consonant [m] and the vowel [aː] of the accented syllable. The complete peak contour is then shifted to the right in 8 steps of 30 ms each, and to the left in one step of 60 ms. This procedure creates a peak-shift series of 10 stimuli (p01 – p10), as illustrated by the F0 traces of Figure 1, synchronized with the spectrogram. They span the previously established prosodic categories of *early – medial – late-medial – late* peaks, which are associated with the accented word "mager" and in turn with its stressed syllable. But the temporal position of these patterns is not such that the contour is completed within a set domain. All that can be said is that the peaks are synchronized in relation to the accented syllable and its vowel nucleus: ‘early peak’ means that the high point is before the accented vowel onset, and it may even be in a preceding syllable, depending on the timing of voicing in the CV transition. In the experiment, the high point was near the beginning of /m/. In a late peak the high point is at the end of the accented syllable, and it may not occur until following unstressed syllables, again depending on the timing of articulation (vowel length, voicing of consonants). The other two medial peak types fall in between. The decisive differences between these peak categories are their relational earlier - later synchronizations with the accented syllable; their absolute positions are variable within certain ranges and depend on the articulatory timing constraints.

In the F0 generation of the rising pattern, the precursers is 130 – 120 – 125 – 110 Hz. The valley contour ends either at 180 Hz or at 300 Hz and has one of three shapes:
(a) 110 – 85 – 180/300 Hz, valley minimum at boundary between [m] and [a:], 100 ms duration of the F0 descent 110 – 85 Hz: linear rise

(b) 110 – 85 – 90 – 180/300 Hz, valley minimum at boundary between [m] and [a:], 100 ms durations of the F0 descent 110 – 85 Hz and of the F0 trough 85 – 90 Hz: concave rise

(c) 110 – 180/300 Hz, 110 Hz at boundary between [m] and [a:]: linear rise from higher level.

The valley descent + trough 110 – 85 – 90 Hz in (b) is shifted to the right in 2 steps of 50 ms each, i.e. in the third position, the 110 Hz point is at the consonant-vowel boundary. This procedure creates two valley series of 5 stimuli each. The two series differ in a medium vs a high endpoint, but are otherwise exactly parallel. The 3 stimuli of (b) constitute a shift from early to late position of a valley trough. (a) and the first stimulus of (b) have the same early synchronization of the F0 minimum (85 Hz), but they differ in the shapes of their rises, the latter with a trough, the former without, reinforcing low and high pitch, respectively. (c) and the third stimulus of (b) have the same synchronization of the 110 Hz point, but they, too, differ in the shapes of their rises, the latter with a descent + trough, the former without, reinforcing low and high pitch, respectively, even more strongly. In the generation of all three series, the F0 points are connected by linear interpolation. Figure 2 shows the F0 patterns, in synchronization with the spectrogram, for the medium rise. The high-rise patterns are identical, except for the higher F0 end point of 300 Hz instead of 180 Hz.

2.2 Semantic scales

The association of the domains of peak and valley synchronization and of valley shapes with pragmatic functions is investigated using the technique of the semantic differential (Osgood, Suci and Tannenbaum 1967). It has occasionally been applied to the analysis of intonation, either with the aim to extract dimensions from the semantic scaling data by factor analysis (Uldall 1960, 1964), or hypothesis-driven with subsequent analyses of variance (Dombrowski 2003, Ambrazaitis 2005). The latter application will be followed in this study.

7-point scales are used to capture the semantic space covered by the early-to-late peak and valley stimulus series. The semantic scaling is expected to produce groupings of stimuli that can be linked to the phonological peak and valley categories, established by previous analyses, as outlined above. Two scales are selected from each of four semantic and pragmatic domains:

(a) the speaker acting in dialogue
scale 1: abschließend – weiterweisend (concluding – continuing)
scale 2: fragend – feststellend (questioning – asserting)

(b) the relationship speaker – hearer/situation
scale 3: endgültig – kompromissbereit (final – ready to compromise)
scale 4: akzeptiert – kontrastiv (accepted – contrastive)

(c) information assessment by the speaker
scale 5: bekannt – unerwartet (known – unexpected)
scale 6: sachlich – erstaunt (matter-of-fact – surprised)
(d) the speaker's expressiveness

scale 7: emphatisch – unemphatisch (emphatic – unemphatic)
scale 8: engagiert – gleichgültig (involved – indifferent)

Each stimulus presented in the listening test is linked to one of eight scales in a corresponding questionnaire, with the following format:

<table>
<thead>
<tr>
<th>fragend</th>
<th>feststellend</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 2 1 0 1 2 3</td>
<td></td>
</tr>
</tbody>
</table>

("The utterance sounds questioning – asserting."")

The scale points were defined as follows:
- centre point 0 = undecided between the poles
- scale end points 3 = strong in the direction of one pole
- intermediate values 1 or 2 = weak or moderate in the direction of one pole.

In the listening test, several subjects seem to have had problems with scale 7 "emphatisch" – "unemphatisch". After the trial run (see 2.5), which illustrated this scale by a very late peak (position p10, see 2.1), they enquired what "emphatisch" meant. In one or two cases they even appeared not to know the word in their German, in other cases it may have been the ambiguity of the word between "forceful" and "emotional", the latter of which being intended by the scale. Moreover, this is the only scale where one pole is named by the verbal negation of the other, which makes it difficult to place 0. The assessments along this scale will be presented, but due to the uncertainty created by the choice of words, no conclusions will be drawn from these results.

2.3 Contextualization of stimuli

The listening test is put in a scenario of "Looking at old photos": subjects are to put themselves in a situation in which they look at old photos together with a friend; they come across a photo of a mutual acquaintance, to which the friend reacts by saying ... The subject then has to assess the friend's remark. This general contextualization of all the stimuli in the listening test is presented to the subjects in the instructions preceding the test.

2.4 Test file and questionnaire

Each of the 20 stimuli (10 peak, 5 medium and 5 high rising valley contours) is presented once for assessment on each of the 8 scales, i.e. the test consists of 160 utterances Er war mal mager. They are labelled as peak\text{n}m\text{p}, valley\text{n}M\text{p} and valley\text{n}H\text{p}, with \text{m} = 01 ... 10, \text{n} = 1 ... 5, \text{p} = 1 ... 8, and this label list is randomised. An IPDS-developed software associates the randomised label list with the wav files of the 20 stimuli and generates the test file. The stimulus sequence in the audio file corresponds to the scale sequence in the questionnaire. To reinforce the auditory impression of the utterances for semantic judgement, each stimulus is presented as a repetition of itself with a 1s-pause in between. This is followed by a 5s-pause for the response. Each stimulus pair is introduced by a 500 Hz, 500 ms bleep. After each block of 10 stimulus pairs, there is an additional 500 Hz, 1000 ms bleep for orientation: one such block corresponds to a page of the questionnaire, so this is the signal to turn the page.

The test file is preceded by a recording of the test instructions and by a trial run with eight stimuli, one for each semantic scale. The instructions are also presented in writing at the be-
ginning of the questionnaire, together with the scales for the trial run. At the end of the ques-
tionnaire there are questions about the ease/difficulty of the task, about the natural/unnatural 
sound of the utterances and about biographic data (age, gender, name, native language = 
German?).

2.5 Listening test
The listening test was carried out in the sound-treated studio of the IPDS of Kiel University in 
several sessions with groups of 2 to 8 subjects, who were given the prepared questionnaire. 
Instructions, trial run and listening test were presented through loudspeakers. Subjects were 
allowed questions after the trial run. The total test took about 45 minutes, with 25 minutes for 
the actual scaling test. 46 native German speakers, students of phonetics/linguistics/psychol-
ogy, took part. Three questionnaires had to be excluded from the analysis because of missing 
answers. This left 6880 responses from 30 female and 13 male listeners (average age 23) - 20 
x 8 x 43.

3 Hypotheses
The experimental design has 3 factors:

(1) direction of post-accentual F0: peak vs valley

(2) end point of valley rise: medium vs high

(3) synchronization and shape of peak or valley, respectively
   (3a) early vs medial vs late-medial vs late: 1 + 3 + 3 + 3 peak positions
   (3b) 3 early to late positions of a concave valley rise
   (3c) early position of linear valley rise from 2 different levels, pairable with the first 
        and the third concave valley shift position for F0 shape effect.

There is a fourth factor, which is, however, not pursued in this investigation: listener gender.

This design is based on the following hypotheses:

(I) There is semantic category formation corresponding to the ranges of early vs medial 
    vs late-medial vs late peak positions through semantic feature bundles of the 5 scales 
    "concluding" – "continuing", "accepted" – "contrastive", "known" – "unexpected", 
    "matter-of-fact" – "surprised", "emphatic" – "unemphatic".

(II) The two valley series produce different category formations on each semantic scale: 
    the high-ending valley points to higher degrees of "continuing", "questioning", "ready 
    to compromise", "contrastive", "unexpected", "surprised", "involved", "emphatic"

(III) The negative results of the valley shift experiment (Niebuhr and Kohler 2004) suggest 
    that the position factor is weak in valleys and, in view of hypothesis (II), interacts with 
    the end point of the rise.

(IV) Following on from hypothesis (III), the shape effect in valleys is also weak and again 
    interacts with the end point of the rise.
4 Results

4.1 Descriptive
For each of the 20 pitch patterns, averages of the scores of the 43 subjects are calculated on each scale, with the left section getting the negative, the right section the positive sign, i.e. the scales are treated as interval scales, which twists reality somewhat but is a prerequisite for subsequent parametric statistics as there seem to be no non-parametric tests available for this type of data and questions. Figures 3 and 4 provide the data for peak and valley contours, respectively, in each case separated into the scales (I) "concluding" – "continuing", "questioning" – "asserting", "emphatic" – "unemphatic", "involved" – "indifferent", and (II) "final" – "ready to compromise", "accepted" – "contrastive", "known" – "unexpected", "matter-of-fact" – "surprised".

4.1.1 Peaks
For each peak position p01 – p10 of the shift series (cf. Figure1), Figure 3 gives the average scores across the 43 listeners on each of the semantic scales in groups (I) and (II), respectively. Position 1 of the F0 maximum, at 60 ms before the accented vowel onset, yields the largest value for "concluding" (early), positions 8 – 10, in the last 80 ms of the accented vowel, the smallest values, around 0 (late). On the "accepted" – "contrastive" scale, position 1 shows the largest negative value, the 3 late positions positive values; the preceding positions 5 – 7, at 90 – 150 ms into the accented vowel, have the smallest negative values (late medial). There is thus a partitioning into 4 domains along this scale. On the "known" – "unexpected" scale, all values are negative, but they decrease in the late-median and late domains, and are smallest for the last two positions. On the "matter-of-fact" – "surprised" scale, the early position has the largest negative value, close to –3, the 3 late positions the smallest, around 0.

The "emphatic" – "unemphatic" scale shows no clear picture, with all values in a very small range around 0 (-1 < r < +1). The problems with this scale have already been pointed out in 2.2. On the remaining three scales, the data provide no category formation; all peak positions are highly "assertive", moderately "final", and in a very small range around 0 for "involved" – "indifferent".

4.1.2 Valleys
For each valley position/shape 1 – 5 (cf. Figure 2), Figure 4 gives the average scores across the 43 listeners on each of the semantic scales in groups (I) and (II), respectively. At each valley position/shape, the scores for medium and high rises are juxtaposed in that order. Each high-ending valley contour has a substantially higher value than its medium counterpart on every single scale; in the case of "involved" – "indifferent", there is a change of sign, medium being positive, high being negative. Within each of these patterns, the concave valley shift (positions 2 – 4, F0 minimum at 0, 50, 100 ms into the accented vowel, followed by a low F0 trough of 100 ms) shows quite small effects. The more prominent ones are a decrease of "continuing", of "questioning" and of "unexpected" in the medium contours. There are no comparable effects for the corresponding high contours. This points to a strong interaction of the high end point on the position factor. The shape of the rise also shows small effects. The linear medium rises (positions 1 and 5, F0 minimum at accented vowel onset) are more "surprised" than the concave rises. Again this does not apply to the corresponding high rises. The linear medium rise from 110 Hz is most "indifferent", the linear high rise from 110 Hz least "involved", in their respective valley series.
4.2  **Inferential statistics**
Repeated-measures analyses of variance are carried out separately for hypotheses I (peaks), II (valleys), and III (valley positions), in each case first a multivariate MANOVA with the 8 scales as dependent variables, then univariate ANOVAs for each scale. To test for peak series internal contrasts, paired-samples t tests of successive factor value means are applied to each scale. Hypothesis IV (valley shape) is also tested by paired-samples t tests. Significance thresholds are set at p=0.01 for all tests, and means must differ by at least one scale point. The t tests are 2-tailed. The SPSS software is used.

4.2.1  Hypothesis I: semantic contrasts among the peak series
The analyses of variance use one factor "peak position" with 10 alternatives. The MANOVA is highly significant: Wilks-Lambda 0.313, F= 6.627, df= 72, p=0.000, pointing to significant contrasts among the 10 peak positions on the dependent variables. It therefore makes sense to test the factor on each dependent variable in separate ANOVAs, with application of the Greenhouse-Geisser correction for non-homogeneity of variances. The differences between peak positions are highly significant on all scales, except scale 3 "final" – "ready to compromise" (Table 1).

In view of these significant test results the ensuing question is where along the peak shift series contrasts between peak positions become significant. Paired-samples t tests of successive factor means are to provide the answer for each scale.

- **Scale 3**
  Significance was already ruled out by the ANOVA.

- **Scale 2"questioning" – "asserting"
  There is either no significance, or the difference of means is <1 scale point. All means are above 2.1. So scale 2 does not contribute to the semanticization of peak positions.

- **Scale 7 "emphatic" – "unemphatic"
  There is either no significance, or the difference of means is <1 scale point. All means are above −1 and below +1. The problem connected with the naming of this scale has already been pointed out in 4.1.1. So scale 7 does not contribute either to the semanticization of peak positions.

- **Scale 1 "concluding" – "continuing"
  There is significance for the comparison of positions p08_p09 (t=-2.977, df=42, p=0.005), but the difference of means is slightly below 1 scale point. The total range of means between peak positions is considerable, from −2.05 to 0.19. Thus the *early* peak p01 is clearly "concluding", *late* peaks from p08 move towards "continuing".

- **Scale 5 "known" – "unexpected"
  There is no significant contrast, but the total range of means between peak positions ranges from −2.19 to −0.81. Thus the *early* peak p01 conveys the feature "known", whereas late peaks point more in the direction of "unexpected".

- **Scale 6 "matter-of-fact" – "surprised"
  Contrasts p01_p02 and p08_p09 are significant (t=-3.884; 2.946, df=42, p=0.000; 0.005), but with mean differences below 1 scale point; contrast p07_p08 is significant
(t= -3.656, df=42, p=0.001) with means of −1.19 and 0.14. Thus the early peak p01 conveys the feature "matter-of-fact", whereas late peaks point in the direction of "surprised".

- Scale 8 "involved" – "indifferent"
  Contrast p06_p07 is significant (t=−2.898, df=42, p=0.006), but below 1 scale point; contrast p03_p04 is significant (t=3.625, df=42, p=0.001) with means of 0.86 and -0.30. p01 has the largest positive value of 1.14. Thus the early peak p01 tends towards the feature "indifferent"; from p04 there is more of a tendency in the direction of "involved".

- Scale 4 "accepted" – "contrastive"
  Significant contrasts are: p01_p02 (t=−4.606, df=42, p=0.000), means of -2.12 and -0.65; p04_p05 (t=−3.732, df=42, p=0.001), means of −1.51 and −0.56; p07_p08 (t=−3.884, df=42, p=0.000), means of −0.53 and 0.63. The total range of means is from p01=−2.12 to p10=1.30. So the early peak conveys the semantic feature "accepted", for medial peaks the strength of this feature decreases, late-medial peaks tend towards neutral, and late peaks are clearly "contrastive".

Except for the "emphatic"/"unemphatic" scale (see also 2.2, 4.1.1), descriptive and inferential statistics have thus confirmed hypothesis I: there is semantic category formation corresponding to the ranges of early vs medial vs late-medial vs late peak positions through semantic feature bundles of the scales "concluding" – "continuing", "accepted" – "contrastive", "known" – "unexpected", "matter-of-fact" – "surprised".

4.2.2 Hypothesis II: semantic contrasts between the two valley series
The analyses of variance use two factors "height of end point" with 2 alternatives and "valley position/shape" with 5 alternatives. The MANOVA is highly significant on both factors and their interaction: HEIGHT Wilks-Lambda 0.045, F= 92.814, df= 8, p=0.000; CONTOUR Wilks-Lambda 0.339, F= 6.347, df= 32, p=0.000; HEIGHT*CONTOUR Wilks-Lambda 0.453, F= 4.460, df= 32, p=0.000. These results point to significant contrasts between the 2 final pitch heights and between the 5 valley positions/shapes, but also to strong interactions of height on position/shape in the dependent variables. Therefore the two factors and their interaction are tested on each dependent variable in separate ANOVAs, with application of the Greenhouse-Geisser correction. The HEIGHT differences are highly significant on all scales. This also applies to CONTOUR differences and the HEIGHT*CONTOUR interactions. An exception is scale 3 "final" – "ready to compromise", for which all contours within the medium or the high class convey an equally low or high degree of "ready to compromise", respectively. (Table 2)

Hypothesis II has thus been confirmed by descriptive and inferential statistics: there are different category formations for the two valley series on each semantic scale, with greater degrees of "continuing", "questioning", "ready to compromise", "contrastive", "unexpected", "surprised", "involved" for the high-ending valley, and vice versa for the medium-ending one. This list should also include "emphatic", but see 2.2 for problems connected with this scale.

4.2.3 Hypothesis III: semantic contrasts in the concave valley series
The significant CONTOUR effects and HEIGHT*CONTOUR interactions of the ANOVAs in 4.2.2 refer to valley position as well as valley shape. The subsequent analyses concern the
respective contributions to these effects by position and shape separately. Therefore two-
factor ANOVAs were run for each scale, with 2 values for HEIGHT and 3 values for POSI-
TION in the concave valley shift series (Table 3).

The factor POSITION only shows significant effects on scales 1 "concluding" – "continuing", 2 "questioning" – "asserting", 5 "known" – "unexpected" (also scale 7 "emphatic" – "unem-
phatic", the naming of which has, however, proved problematic, see 2.2). POSITION interacts highly significantly with HEIGHT on scales 2 and 5.

Hypothesis III has thus been confirmed by descriptive and inferential statistics: the position factor is weak in valleys and interacts with the end point of the rise. The later position of a medium-rise concave valley reduces the semantic features "continuing", "questioning" and "unexpected".

4.2.4 Hypothesis IV: semantic contrasts between valley shapes
In the series of both medium and high rises there are two pitch contrasts of linear and of concave shape, rising from 85 Hz or 110 Hz (med_linear_85 vs med_concave_85, med_linear_110 vs med_concave_110; high_linear_85 vs high_concave_85, high_linear_110 vs high_concave_110). It has already been pointed out in the descriptive assessment of the data (4.1.2) that these shape contrasts yield different results for the medium, but not the high rise on the scale "matter-of-fact" – "surprised", and that the linear medium rise from 110 Hz has the largest positive value on the "involved" – "indifferent" scale. These data point to a weak shape effect and to strong interactions of shape with height. To test for these effects t tests are applied to each of the 4 comparisons on each of the 8 scales.

There are only very few shape contrasts that are significant with >1 scale point difference of means. The med_*_85 contrast is significant on scales 2 "questioning" – "asserting" (t=3.932, df=42, p=0.000) and 6 "matter-of-fact" – "surprised" (t=4.670, df=42, p=0.000), med_linear_85 being less "questioning" and more "surprised". The med_*_110 contrast is significant on scale 8 "involved" – "indifferent" (t=-4.103, df=42, p=0.000). It just misses the set significance level on scale 6 "matter-of-fact" – "surprised" (t=-2.554, df=42, p=0.014), with slightly <1 scale point difference. There are no significances for high_*_*.

Hypothesis IV has thus been confirmed: there is a weak shape effect in valleys, which inter-
acts with the end point of the rise. Taking the descriptive and the inferential statistics together we can say that the linear medium rise strengthens the feature "surprised", whereas the concave medium rise points more towards "matter-of-fact". The linear medium rise from 110 Hz is more "indifferent" than any of the other medium rises, especially the concave ones; this contrasts with the highly weighted "involved" feature of the high rises. However, there is the complementary tendency in the linear high rise from 110 Hz towards greater indifference. Starting the rise at a high level seems to increase this semantic feature.

5 Discussion and explanation of findings

5.1 Peak synchronization
Prosodic features define four ranges within the synchronization series of F0 peak contours.

- In early peaks, there is a high-low F0 trajectory into the accented vowel. If the ac-
cented syllable is preceded by another syllable the latter carries the high focal point, as in the test sentence, otherwise F0 drops quickly from voice onset of the accented syl-
lable. A low F0 level is reached early in the accented syllable and continues to drop in subsequent unstressed syllables. The overall auditory impression of the early peak contour is low-pitch falling.

- In medial peaks, there is a low-high F0 trajectory into the accented vowel. The high focal point is reached shortly after vowel onset, how late in relation to vowel offset depends on the duration of the vowel. Microprosodic influence from syllable-initial voiceless obstruents raises F0 and creates a continuously falling F0 pattern in the accented vowel. The perceived pitch pattern is, however, of the same type as the one associated with the rise-fall in other syllable contexts. A low F0 level is reached later in the accented syllable and subsequent unstressed syllables, depending on the segmental timing. The overall auditory impression of the medial peak contour is high-pitch falling.

- As the low-high F0 trajectory is moved further into the accented syllable, the pitch rise becomes prominent and creates the overall auditory impression of a rising-falling pattern for a late-medial peak, distributed across the accented and following unstressed syllables.

- In late peaks the rise of the rising-falling pattern is preceded by a low F0 trough in the accented vowel. Where the high focal F0 point is located depends on the segmental timing: it may be at the end of a syllable containing a long vowel (and sonorant consonants) or in a following unstressed syllable, particularly with intervening voiceless obstruents, and in the case of utterance-final accented monosyllables, vowels, especially short ones, need to be lengthened to accommodate the complex low + rising-falling pattern. The overall auditory impression of the late peak contour is a low-high pitch contrast late in the accented vowel before the even later fall.

The partitioning of the peak synchronization continuum into four F0-defined pattern ranges is mapped onto perceptual categories of high-low or low-high pitch contrasts in different positions around the accented syllable as part of a terminal fall. The sharpest separation is between the reversal from a high-low to a low-high contrast at accented-vowel onset, i.e. between early and medial because of the support by the increase of acoustic energy at the consonant – vowel landmark. The remaining three ranges are less sharply separated in signal parameters as well as in perception because they lack this referential landmark. But they nevertheless constitute perceptual categories. In the experiment of this paper, the early category is only represented by one F0 peak position since extensive research has well established the sharp division between early and medial. The other three categories are represented by up to three positions.

Each of these prosodic categories is associated with different weightings in semantic feature bundles. The early peak signals high degrees of "matter-of-fact", "known", "accepted" and "concluding". For the medial peak, the values on all these features are reduced, most significantly so for "accepted". The late-medial peak differs from medial by the very low degree on the "accepted" feature, moving to neutral. The late peak is weighted on the "contrastive" feature and has the lowest values of "matter-of-fact" and "concluding", approaching neutral in both cases.

The auditory low-pitch falling pattern of the early peak may thus be linked to semantic and pragmatic finality. By contrast, the auditory high-pitch falling patterns of the medial and late-medial peaks are relatable to semantic and pragmatic openness. In the late-medial peak the
auditorily more prominent rising pattern in the accented syllable, before the subsequent fall, strengthens openness and overlays it with semantic and pragmatic contrast. In the late peak, the late low-high pitch contrast in the accented syllable strengthens contrast and adds surprise. All four peak patterns are equally high on the "asserting" and "final" features, which separate them from the valley patterns.

5.2 High and medium-rising valley patterns
All valleys are associated with the features "questioning" and "ready to compromise" and thus form a different semantic and pragmatic category from peaks. It is divided into two subcategories according to the medium or high end point of the rise. All high-ending valleys have high values on the features "continuing", "questioning", "contrastive", "unexpected", "surprised", moderate to high values on the feature "involved" and moderate values on the feature "ready to compromise". By contrast, all corresponding medium-ending valleys are low on the features "continuing", "questioning", "ready to compromise" (tending to neutral), "unexpected"; they are around 0 on the feature pairs "accepted/contrastive" and "matter-of-fact/surprised" and moderate to high on the feature "indifferent". Thus, irrespective of the shape of the rise and its synchronization with articulation, the high end point may be said to signal question, interest to continue the argument and surprise, whereas the medium end point signals statement, softened conclusiveness and disinterest. Utterances in declarative syntax require extra high F0 rises to signal question.

5.2.1 Valley synchronization and valley shape
The concave valley and its shift through the accented vowel introduce a progressively longer stretch of low F0 before the final rise. In the late position of the concave valley, 200 ms of the accented vowel, i.e. almost its entire duration, are occupied by a low F0 fall + trough, which is a perceptually very prominent low-pitch precursor to the rise. It has no significant effect on the very wide range of the high rise. But in the much narrower range of the medium rise it lowers the generally low values of the features "continuing", "questioning" and "unexpected" even further. Moreover, the comparison of the linear and concave medium-rise contours from 85 Hz shows a significant change on the "matter-of-fact" – "surprise" scale from "surprise" to "matter-of-fact". These features point to a strengthening of statement and conclusiveness.

On the other hand, the linear rise from 110 Hz raises the pitch percept of the valley contour. Comparison with the late concave valley shows a significant increase in the feature "indifferent" for the medium rise and a decrease of the complementary feature "involved" for the high rise. Furthermore, the comparison of the two linear medium rises, from 85 Hz and 110 Hz, shows up "surprise" for the former but not for the latter. This does not apply to the high rises. Thus a narrowing of the range of the rise at the higher level seems to convey reduced interest, especially in the medium rise, where the range is narrower and the interest conveyed is lower anyway.

5.3 Syntagmatic pitch contrast – auditory enhancement
The discussion of the experimental data in this paper has been based on global peak and valley contours, and it has been argued that the listener evaluates them perceptually as well as semantically and pragmatically with reference to syntagmatic pitch contrasts. In the case of peak contours it is the high-low or low-high pitch trajectories in varying synchronization with the articulatory timing of the accented syllable that determines pitch categorization. In valley contours the overriding feature is the rising trajectory to a terminal point of varying height that determines categorization. Internal syntagmatic contrast also plays some role in the categorization of valley contours, in the form of linear versus concave rise and its synchronization, or the height of the starting point of the rise. But these effects are dominated by the over-
all height of the rise, which prevents the same clear separation into synchronization ranges of valleys as for peaks. These syntagmatic pitch contrasts may be seen as another aspect of auditory enhancement (Diehl 1991).

5.4 The frequency code

Some aspects of the valley as well as the peak semanticizations of this paper may be explained with reference to Ohala's frequency code (1983). The expression of questions in declarative syntax and of interest in the continuation of an argument in dialogue involves the speaker's subordination to the listener, and is signalled by extra high rising pitch. Moving towards statement and conclusiveness brings in speaker orientation and lower rising pitch. If instead of readiness to compromise the speaker wants to express assertiveness the interchange becomes speaker dominated and the pitch changes to a peak pattern. With the low-pitch fall of the early peak this domination becomes final, whereas the high-pitch fall of the medial peak opens the door to new arguments.

6 Conclusion and outlook

This paper has attempted to show that for an insightful interpretation pitch needs to be analysed as global peak and valley contours (see also Kohler, 2006) in their timing with reference to articulation and to internal syntagmatic contrast, and that semantic and pragmatic functions are central and need to be considered right from the data collection stage by situational contextualization of stimuli, and that the listener plays a pivotal role for the assessment of pitch categories. The dissection of global contours into elements and their symbolization, as in the dominant framework of autosegmental-metrical phonology and ToBI, can be no more than heuristic devices to get symbolic records of pitch data, and they should not be reified into underlying phonological entities in cognitive speech processing by speakers and listeners, which the phonetician fills with phonetic measurement in the laboratory. To arrive at these cognitive entities the opposite path is necessary, viz. speech is investigated in an experimental framework to derive language categories that determine speech production and perception.

The paradigm of the semantic differential needs refinement, and a lot more thought should be invested into the naming of semantic scales to develop a semantic net for pitch analysis. The already available data also need to be analysed with regard to the interactions between falling and rising patterns, such as the possible groupings of concave valley shifts with late peaks. Also a more finely meshed valley shift has to be implemented, as well as the investigation of peak height semanticization within the whole system of pitch categories. Different contextualizations and segmental types of test stimuli should also be used. Gender effects of the listener, as well as the speaker are further research issues. And finally this type of study should be tackled in a variety of languages. However, to achieve this extensive research programme we need a new orientation towards function-oriented experimental phonetics, which overcomes the superficial alignment measurements in preestablished pitch categories of Laboratory Phonology.

Acknowledgements

The research reported in this paper received help from two people. Felicitas Kleber arranged and administered the numerous listening tests and compiled the questionnaire data in electronic files for descriptive and inferential statistics. Ernst Dombrowski gave valuable statistical advice. I am very grateful to both of them. I also greatly appreciated the very helpful suggestions from the three reviewers, Randy Diehl, Nina Grønnum and Ilse Lehiste.
References


Figure legends

Figure 1
Spectrogram of the original utterance Er war mal mager. (produced with falling intonation), and, synchronized with it, the series of 10 time-shifted f0 peak patterns used for psola synthesis in praat. The plain-line trace has its F0 maximum aligned with the accented vowel onset; the dashed traces are shifted in eight 30 ms steps to the right and in one 60 ms step to the left.

Figure 2
Spectrogram of the original utterance Er war mal mager. (produced with rising intonation), and, synchronized with it, the 5 f0 valley patterns used for psola synthesis in praat. The plain-line trace has its F0 minimum aligned with the accented vowel onset, followed by a linear rise. The 3 dashed traces form the series of time-shifted concave valleys – F0 minimum at accented vowel onset and two 50 ms steps to the right. The dotted line refers to the valley that is also aligned with the accented vowel onset, but has a raised starting point of the linear rise.

Figure 3
Average scores for each of the 10 peak positions on semantic scales 1, 2, 7, 8 (I) and 3, 4, 5, 6 (II) across 43 listeners.

Figure 4
Average scores for each of the 5 positions/shapes in medium (M) and high-rising (H) valleys on semantic scales 1, 2, 7, 8 (I) and 3, 4, 5, 6 (II) across 43 listeners. Results for each M and H pair are juxtaposed in that order for each scale. H is represented by a darker gray shade of the otherwise identical column pattern.
Table 1  
F values, adjusted degrees of freedom according to Greenhouse-Geisser and significance levels of repeated-measures ANOVAs with 10-valued factor "peak postion" on the 8 semantic scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8.695</td>
<td>6.010</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>5.016</td>
<td>5.825</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>1.432</td>
<td>5.865</td>
<td>0.204</td>
</tr>
<tr>
<td>4</td>
<td>21.055</td>
<td>6.678</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>5.154</td>
<td>6.510</td>
<td>0.000</td>
</tr>
<tr>
<td>6</td>
<td>20.310</td>
<td>6.032</td>
<td>0.000</td>
</tr>
<tr>
<td>7</td>
<td>2.888</td>
<td>6.847</td>
<td>0.007</td>
</tr>
<tr>
<td>8</td>
<td>6.767</td>
<td>7.550</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2  
F values, adjusted degrees of freedom according to Greenhouse-Geisser and significance levels of repeated-measures ANOVAs with 2-valued factor HEIGHT and 5-valued factor CONTOUR on the 8 semantic scales

<table>
<thead>
<tr>
<th>Height</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale 1</td>
<td>47.283</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 2</td>
<td>199.659</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 3</td>
<td>50.123</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 4</td>
<td>92.497</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 5</td>
<td>188.359</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 6</td>
<td>192.996</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 7</td>
<td>128.103</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 8</td>
<td>373.931</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contour</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale 1</td>
<td>3.996</td>
<td>3.067</td>
<td>0.009</td>
</tr>
<tr>
<td>Scale 2</td>
<td>9.561</td>
<td>3.513</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 3</td>
<td>1.160</td>
<td>3.383</td>
<td>0.329</td>
</tr>
<tr>
<td>Scale 4</td>
<td>5.918</td>
<td>3.749</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 5</td>
<td>8.633</td>
<td>3.158</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 6</td>
<td>13.126</td>
<td>3.304</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 7</td>
<td>4.908</td>
<td>3.506</td>
<td>0.002</td>
</tr>
<tr>
<td>Scale 8</td>
<td>9.527</td>
<td>2.903</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Height*Contour</th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale 1</td>
<td>5.349</td>
<td>3.445</td>
<td>0.001</td>
</tr>
<tr>
<td>Scale 2</td>
<td>6.676</td>
<td>3.299</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 3</td>
<td>2.744</td>
<td>3.442</td>
<td>0.038</td>
</tr>
<tr>
<td>Scale 4</td>
<td>5.790</td>
<td>3.705</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 5</td>
<td>5.073</td>
<td>3.486</td>
<td>0.001</td>
</tr>
<tr>
<td>Scale 6</td>
<td>9.769</td>
<td>3.396</td>
<td>0.000</td>
</tr>
<tr>
<td>Scale 7</td>
<td>4.367</td>
<td>3.618</td>
<td>0.003</td>
</tr>
<tr>
<td>Scale 8</td>
<td>5.022</td>
<td>3.421</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Table 3
F values, adjusted degrees of freedom according to Greenhouse-Geisser and significance levels of repeated-measures ANOVAs with 2-valued factor HEIGHT and 3-valued factor POSITION on the 8 semantic scales

<table>
<thead>
<tr>
<th></th>
<th>F</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEIGHT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale 1</td>
<td>42.478</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 2</td>
<td>152.139</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 3</td>
<td>25.442</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 4</td>
<td>57.155</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 5</td>
<td>157.654</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 6</td>
<td>186.673</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 7</td>
<td>94.949</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 8</td>
<td>287.309</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>POSITION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale 1</td>
<td>8.532</td>
<td>1.979</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 2</td>
<td>10.863</td>
<td>1.962</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 3</td>
<td>0.359</td>
<td>1.683</td>
<td>0.699</td>
</tr>
<tr>
<td>scale 4</td>
<td>3.312</td>
<td>1.924</td>
<td>0.043</td>
</tr>
<tr>
<td>scale 5</td>
<td>12.745</td>
<td>1.671</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 6</td>
<td>1.919</td>
<td>1.794</td>
<td>0.158</td>
</tr>
<tr>
<td>scale 7</td>
<td>8.638</td>
<td>1.763</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 8</td>
<td>2.253</td>
<td>2.000</td>
<td>0.111</td>
</tr>
<tr>
<td>HEIGHT*POSITION</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>scale 1</td>
<td>2.676</td>
<td>1.807</td>
<td>0.081</td>
</tr>
<tr>
<td>scale 2</td>
<td>8.888</td>
<td>1.860</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 3</td>
<td>2.257</td>
<td>1.835</td>
<td>0.116</td>
</tr>
<tr>
<td>scale 4</td>
<td>8.491</td>
<td>1.954</td>
<td>0.000</td>
</tr>
<tr>
<td>scale 5</td>
<td>6.300</td>
<td>1.943</td>
<td>0.003</td>
</tr>
<tr>
<td>scale 6</td>
<td>3.904</td>
<td>1.951</td>
<td>0.024</td>
</tr>
<tr>
<td>scale 7</td>
<td>5.838</td>
<td>1.842</td>
<td>0.004</td>
</tr>
<tr>
<td>scale 8</td>
<td>5.147</td>
<td>1.916</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Figures 1 & 2
Figure 3

Peak series: Semantic features I

- Concl-cont
- Quest-assert
- Invol-indiff
- Emph-unemph

Peak series: Semantic features II

- Final-comprom
- Accept-contrast
- Known-unexpect
- Fact-surprise
Figure 4

Valley series: Semantic features I

Valley series: Semantic features II