

# Neglected Categories in the Modelling of Prosody

## Pitch Timing and Non-Pitch Accents

Klaus J. Kohler

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

### ABSTRACT

With the *Kiel Intonation Model* (KIM) as a point of departure, data from German and other languages are presented showing the importance of two timing categories in the production and perception of pitch patterns: (a) synchronization of holistic F0 contours with articulation, (b) pattern-internal faster/slower F0 trajectories, e.g. in rising-falling peak patterns. On the accent level, data are discussed that point to the need for a force accent category besides pitch accent.

## 1 The Research Questions

Distinctive pitch patterns, i.e. intonation, and distinctive prominence relations between words in an utterance, i.e. sentence accents, as well as their interrelations, are central prosodic features in speech communication as expressions of syntactic, semantic and pragmatic functions.

### 1.1 Timing as a determinant of pitch categories

Distinctive intonation patterns are based on F0 configurations, their internal timing and their synchronization with the vocal tract time course. In the intonation analysis of European languages, English in particular, perceived pitch contours have traditionally been associated with syllables or words or just accented words, depending on the degree of abstraction. This structural link provides some coarse timing information with respect to the synchronization of pitch and articulation. But different temporal alignments of pitch with syllables are categorised as separate patterns, such as the rise-fall vs. the fall in British intonation analysis, or L+H\* and L\*+H vs. H\* in autosegmental-metrical phonology and ToBI. These patterns can, however, be viewed as the same contour variously synchronized with the articulation of syllables or words.

As regards contour-internal timing (slower and faster falls and rises), they have largely been ignored, but there are some examples of structural descriptions of such cases in the literature, e.g. in Pike [8], p. 74: "In a phrase which is grammatically close knit, an EARLY DROP in pitch increases the effectiveness of the contrastive-pointing meaning ... Compare the follow-

ing pairs:

I	am	tired.		I	am	tired.
3-	°2-	-4		3-	°2-4-	-4 ..."

When intonation research entered speech synthesis, the synchronization of F0 contours with articulation became crucial. Thus within the ToBI framework, a host of speech production studies have been carried out on the temporal alignment of certain points in F0 patterns (F0 peak, beginning of F0 rise, end of F0 fall) in relation to segmental structure.

It has rightly been pointed out by van Santen [9] that these types of measurement at the acoustic surface level of speech production cannot capture the features that are involved in alignment and that stay invariant for a given phonological category. He proposes a model in which alignment is viewed as a mapping between an underlying global F0 trajectory and a temporal segmental structure. This abstract mapping relationship is invariant for a particular distinctive intonation pattern in a particular 'speech state' (affective states, speech rate, paralinguistic factors). The output from this mapping of abstract dynamic patterns generates variability of pitch - segment synchronization and segmental interference in surface F0 patterning.

Such an abstract, global contour approach to the synchronization of F0 with articulation has its complement in perceptual experiments with systematic temporal shifts of holistic patterns, e.g. entire peak contours. D'Imperio [2] introduced this perceptual perspective of alignment into the ToBI framework, but the decomposition of global F0 patterns into sections associated with sequential pitch levels L and H is still prevalent, in spite of 'the whole rise-fall contour' being considered a 'gestalt' configuration.

### 1.2 Phonetic properties of sentence accents

The phoneticians of the early 1900s (Sievers, Sweet) associated accent primarily with physiological force, hence the terms *expiratorischer Accent* and *stress*, culminating in Stetson's *stress (reinforced chest) pulses*.

Perceptual research established the cuing power of F0 and duration in accent perception, at the cost of the acoustic transform of physiological force, i.e. intensity. The autosegmental-metrical approach of Pierre-

humbert and her followers, as well as the ToBI annotation system, abandoned a separate level for accent besides pitch in prosodic modelling, reduced distinctive pitch to two levels, relegated degrees of accent to a variable dependent on, and deducible from, the metrical grid, and linked accent exclusively to pitch in the category *pitch accent*.

However, the prosodic analysis of natural spontaneous speech shows very quickly that for an adequate account of all the data, especially emphatic and emotional speech, it is necessary to complement the accent - pitch link by a force accent category, and thus to partially return to the former accent concept. It is related to greater physiological and articulatory effort translating into increased duration of syllable-initial consonants (instead of nucleus and coda lengthening under pitch accents) and higher intensity.

## 2 Categories of pitch timing and sentence accent in the *Kiel Intonation Model* ( KIM)

### 2.1 Development of KIM

Model development based on experimental pitch production and perception analysis started in 1985 as part of a German Research Council (DFG) funded project *Form and Function of Intonation Peaks*. First results were presented at the Acoustics Congress in Toronto 1986, at ICPHs in Tallinn 1987, and at the First Conference in Laboratory Phonology at Columbus/Ohio 1988. This basic research was subsequently fed into R&D for an improved prosodic module for a German TTS system in cooperation with Infovox/KTH Stockholm [6]. A comprehensive account of the model development was given in [5].

The prosodic model was then applied to devising a notation system ( PROLAB) for prosodic labelling of spontaneous speech. KIM and PROLAB were presented at the ATR Workshop *Computational Approaches to Processing the Prosody of Spontaneous Speech* in 1995 [7]. A substantial data base of German *The Kiel Corpus of Read/Spontaneous Speech* has since been collected and labelled segmentally as well as prosodically, using PROLAB. So far, 4 CDROMs have been issued, constituting the largest segmentally and prosodically labelled database available for any language.

### 2.2 Central aspects of KIM

The following aspects are central to prosodic modelling in KIM and differentiate it from other approaches, e.g. AM phonology and ToBI.

(a) The category of sentence accent is a separate prosodic level from intonation, controllable independently from rhythmic, syllabic and segmental structuring, on a scale from 1 to 3. Although it shares F0 as a physical property with intonation, it is not en-

tirely determined by it, but also depends on syllable and segment duration, intensity, and possibly other features. Perception experiments [5] have proved that shifts of rising-falling F0 contours from an originally accented syllable to another potential accent position only change accent perception if the syllable duration falls within the expected range; if it is too short the accent continues to be perceived on the original syllable in spite of it no longer having the dynamic F0 pattern.

(b) Sentence accents refer to accented words, and in a language with lexical stress, such as German or English, they manifest themselves by default at the lexical stress position within the word.

(c) Contrastive pitch is conceptualized as global patterns, i.e. peak, valley, combined (falling-rising), and level contours, associated with sentence accents.

(d) Peak, valley and combined contours are variously synchronized, as whole patterns, with articulatory timing. In German and other languages, early, medial and late peak, as well as early and late valley positions in accented syllables signal different semantic and pragmatic functions. The signal - function relations of peak patterns were established experimentally in the perceptual domain by changing F0 maximum synchronization with accented-vowel onset through temporal shift of whole peak contours.

(e) Perception of intonation and accent interact in that, e.g., a late peak on a potential accent syllable can at the same time be an early peak on a following potential accent syllable. In such cases, accent perception may oscillate between the two positions.

(f) Perception of accent and of intonation - accent interaction is not just determined by the synchronization of an F0 peak maximum with articulation but also by the shape of the whole peak contour, i.e. by its internal timing of, e.g. fast vs. slow falling pitch.

## 3 Expanding timing categories in KIM

### 3.1 Internal timing of peak contours affecting the perception of peak category

The effects of internal timing of pitch patterns, in addition to the effects of different pitch - articulation synchronizations, was demonstrated for accent perception in German [5]. Applying the peak shift paradigm [5], Niebuhr [‘Perceptual study of timing variables in F0 peaks’, paper at 15th ICPHs] has shown for German that the timing of rise and fall in a peak contour influences the perception of early vs. medial peak categories. A fast rise followed by a slow fall produces a significantly earlier category change from early to medial compared with a slow rise + fast fall. These findings suggest that the perception of an early peak category is based on a contrast between a low pitch range in the consonant - vowel transition of the ac-

cented syllable and a high pitch range before it, and that for a medial peak the opposite holds.

The consonant - vowel transition is crucial here because it is characterised by an increase in intensity, which heightens the pitch change. The pitch change from a high to a low or from a low to a high range in this transition period cues an early vs. a medial peak. The deferment of the low-high pitch change till later inside the vowel cues a late peak. It is to be expected that the shift to a late peak occurs earlier if there is a fast rise + slow fall as against the reverse. In all these cases, the contrastive strengthening of high or low pitch trajectories at initial or internal accented vowel positions cues different intonation categories. They in turn code different semantics. In German, the high-low F0 trajectory into the accented vowel of an early peak is associated with 'finality', the low-high F0 trajectory at the same syllable position in a medial peak with 'openness', and the deferment of the late peak high trajectory further into the syllable adds 'unexpectedness'. So 'high' and 'low', variously synchronized with accented vowel timing, are important, but in a contour sense, which is different from the sequences of HIGH and LOW pitch level points of AM phonology and ToBI. This is another aspect of Ohala's *Frequency Code*.

### 3.2 Semantic and pragmatic changes associated with different peak-internal timing

In a semantic differential analysis, Dombrowski ['Semantic features of accent contours: Effects of F0 peak position and F0 time shape', paper at 15th ICPHS] showed influences from both timing factors. Besides the well-known semantic differences between medial and late peaks, the strengthening of the high F0 range around the medial peak produced a significant effect towards the question end of the *+questioning* scale. In the case of the late peak position, there was a significant effect on the *+emotional* scale towards an expressive component. These data demonstrate the importance of two F0 timing factors in the semantics and pragmatics of speech communication. Synchronization of F0 and articulation is linked to semantic dichotomies, at least partially based on categorical perception of the pitch patterns. Internal timing seems to add gradual semantic shades of expressiveness.

Grice and Baumann [4] refer to two early peak patterns in German, which they classify as H+!H\* and H+L\* in GToBI notation. They associate the downstepped H+!H\* with Kohler's early peak [5] and the meaning 'summarizing, concluding argumentation'. The other pattern is characterised as expressing a fatalistic attitude. What is essential for the prosodic specification of these two patterns is twofold:

(1) They both contain high-low trajectories in the consonant - vowel transition of the accented syllable, for the same type of synchronization with articulation.

(2) The first-mentioned pattern has a slower descent than the second, and through this different internal timing intensifies the high pitch range in its trajectory.

As regards the semantics of the two patterns, they are both characterised by finality as against openness, but in the more slowly falling one this finality is weakened, in keeping with the *Frequency Code*. In an expression of inevitability and resignation this pattern would probably not be used. GToBI notation, being an abstraction from the time scale, does not only not capture the two timing dimensions involved here, but it does not bring out either what the patterns have in common and what separates them in a universal perspective of human speech communication.

### 3.3 The use of F0 timing in peak contours of different languages and dialects

Gili Fivela [3] analysed peak patterns in Pisa Italian and found that peaks for broad focus and narrow contrast, respectively, differed by the former reaching F0 maximum later and trailing off more slowly afterwards. That means that a combination of earlier/later synchronization of a peak contour coupled with a faster/slower timing of the fall has the function of differentiating broad and narrow focus.

On the other hand, D'Imperio [2] found that in Neapolitan Italian later synchronization of F0 peaks differentiates questions from statements, and that raising the post-peak F0 range in a plateau or in a slower descent increases question responses in perception experiments. This again points to a distinctive function of the strengthening of the high F0 range later in the accented syllable, but this time to code questions.

Similarly, Tamara Khromovskikh [ongoing study at IPDS Kiel] has established, in a series of perception experiments, that statements and questions in Russian are differentiated by a slow rise + fast fall in an early peak vs. a fast rise + slow fall in a late peak. In Bulgarian, the same peak-internal timing difference holds between contrastive focus and questions [1]. The pitch timing contrasts in the rising and falling trajectories highlight the low or the high range inside the accented vowel, which, in turn, code statement vs. question in accordance with the *Frequency Code*.

Peak-internal F0 timing also differentiates between dialects. In Southwest German dialects, for example, falls are generally slower than in North German ones. The abruptness usually observed in (Standard or North) German speakers of English by native English speakers is, at least partly, connected with F0 falls in peak contours also being slower in English.

## 4 Expanding accent categories in KIM

In KIM, the distinctive peak, valley, combined and level intonation patterns are associated with sentence

accents, which are separately and independently controlled in 3 degrees. In the description and modelling of spontaneous speech another accent category is required that further emphasises the independence of accent from intonation in that it is primarily based on exponents of physiological and articulatory effort, and may not even be linked to a distinctive local pitch pattern at all.

In the German spontaneous speech example *wie ... Boris ... Valerie die Treppe runterkickt* ‘when Boris kicks Valerie down the stairs’, l061a018 of the LINDENSTRASSE database) both *Valerie* and *Treppe* have peak contours associated with default accent level 2, and the F0 fall continues to a very low level on *runterkickt*. In spite of this absence of a separate intonation pattern on the final word, the second part of this compound verb, *kickt*, bears a strong accent (level 3). It is related to articulatory force and signalled by a very long and intensely aspirated initial plosive “swallowing up” a good part of the nucleus [j], which is very short, thus heightening the duration of the initial consonant. So particular types of phonation (breathiness, glottalization), ultimately also relatable to physiological effort, are further exponents of the force accent.

In this syntactic pattern of object - verb, the object is focussed and the verb is subordinated: this is achieved by a single F0 peak pattern, which is aligned on the object and spans both. Beside this focussing of a syntactic structure by an accent - intonation link, a force accent on the verbal part of the compound verb (which is not lexically stressed) adds an expressive component, which emotionally intensifies the meaning of the verb. The force accent is completely independent of pitch in this case. There are, however, other cases where the two accent types are superimposed: a force accent occurs inside an accent - intonation structure, the latter coding syntactic and semantic relations, the former adding expressive emphasis.

A force accent may be regarded as a language universal, whereas a pitch accent, although wide-spread in accent languages, is not. In French, for example, pitch patterns are primarily used to signal global phrasing, rather than local accent relations. Thus narrow contrastive focus is not coded by pitch but by syntactic means. Given this different status of accents in French, it is not surprising that French linguists have always recognised the force accent category under the name of *accent d’insistance*. This typological difference in language prosody is also the reason why ToBI has not been so readily applied to French. The category of pitch accents in AM prosodic phonology and ToBI presupposes a certain type of prosodic system, exemplified by English, which is not generalizable. And even in the languages that behave similarly to English the complete category fusion of accent and pitch prevents deeper insight into the working of prosodic systems and

into prosodic universals of speech and language.

## 5 Outlook

Future prosody research needs to pay more attention to pitch timing and non-pitch accents and to model them systematically in both production and perception. The modelling should attend to individual language structures as well as to general aspects of human speech communication from both descriptive and explanatory points of view. It may, for example, be hypothesized that different internal timing patterns can be used in all three German peak types to produce various combinations of slow or fast rises and falls at the early, medial or late synchronization in order to add significant expressive differentiations. This is an interesting area for future research to link paralinguistic aspects to linguistic patterns. Existing prosodic models will have to be adapted to the task. The peak and valley synchronization categories and the separate accent level in KIM can form a nucleus for this research.

## REFERENCES

- [1] B. Andreeva, T. Avgustinova, W. J. Barry. “Link-associated and focus-associated accent patterns in Bulgarian”, in *Current Issues in Formal Slavic Linguistics*, G. Zybatow, U. Junghanns, G. Mehlhorn, L. Szucsich (Eds.), pp. 353-364. Frankfurt/M.: Peter Lang, 2001.
- [2] M. D’Imperio. *The Role of Perception in Defining Tonal Targets and their Alignment*, Ohio State University: PhD thesis, 2000.
- [3] B. Gili Fivela. “Tonal alignment in two Pisa Italian peak accents”, *Proc. Speech Prosody 2002*. Aix-en-Provence: 2002.
- [4] M. Grice, S. Baumann. “Deutsche Intonation und GToBI”, *Linguistische Berichte*, vol. 191, pp. 267-298, 2002.
- [5] K. J. Kohler. *Studies in German Intonation*. AIPUK, vol. 25, Kiel: IPDS, 1991.
- [6] K. J. Kohler. “Parametric control of prosodic variables by symbolic input in TTS synthesis”, in *Progress in Speech Synthesis*, J. P. H. van Santen, R. W. Sproat, J. P. Olive, J. Hirschberg (Eds.), pp. 459-475. New York: Springer, 1997.
- [7] K. J. Kohler. “Modelling prosody in spontaneous speech”, in *Computing Prosody*, Y. Sagisaka, N. Campbell, N. Higuchi (Eds.), pp. 187-210. New York: Springer, 1997.
- [8] K. L. Pike. *The Intonation of American English*, Ann Arbor: U of Michigan Press, 1945.
- [9] J. P. H. van Santen. “Quantitative modelling of pitch accent alignment”, *Proc. Speech Prosody 2002*. Aix-en-Provence, 2002.