Prosody Revisited
FUNCTION, TIME , and the LISTENER in Intonational Phonology

Klaus J. Kohler

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

Abstract

A new look at intonational phonology introduces FUNCTION, TIME, and the LISTENER as essential theoretical categories of prosody with reference to a wide array of language data.

1. Introduction

Mainstream intonation analysis lacks concern for FUNCTION in a wide sense, because it concentrates on linguistic function, if it considers function at all. In the traditional British School, unsystematic glosses are used to refer, often anecdotally, to functions of phonetic pitch distinctions. In Halliday's later systematization of this type of intonation analysis, the categories are established independently of function and meaning, beyond the general prerequisite that they must differentiate meaning. Once set up, their relationships with syntactic structures are worked out, in both directions, from syntax to prosodic categories, and vice versa. In the Dutch intonation model, too, meaning or function plays no role in the analysis: there the distinctive elements of intonation are based on perceptual differentiation through phonetic properties. After the distinctive units have been defined, questions may be asked as to their associations with functions, linguistic ones in the first instance. This also applies to a great deal of research into prosody over the past 20 years, from the Autosegmental-Metrical theory of Pierrehumbert and Beckman to Ladd, and finally to the labelling tool ToBI.

Mainstream intonation analysis also lacks concern for TIME because, with the exception of the Dutch School, it defines intonation contours independently of time. TIME may be introduced post hoc in AM phonology under the heading of 'alignment' (for instance 'peak alignment'), but it is conceptualized as phonetic manifestation within a-temporally defined phonological intonation categories, under various contextual conditions, such as different segmental chains. This differs from the approach in KIM [10, 11], to be discussed in Section 3, which distinguishes three phonological peak contour patterns that are differently synchronized with the vocal tract time course. Adherents of the alignment paradigm often postulate the same phonological intonation categories across languages and dialects and interpret auditory and acoustic differences as different phonetic alignments. A case in point is Grabe's comparative intonational phonology of English and German [7], which concludes that the phonetic realization of H*+L shows later peak alignment in the accented syllable in German than in English. It is based, among others, on experimental production data in situational frames which, at first sight, look identical in the two languages, viz.:

Anna and Peter are watching TV. A photograph of this week's National Lottery winner appears. Anna says: Look, Peter! It's Mr. Sheafer/Shift/Sheaf! Our new neighbour!

Anna und Peter sehen fern. Ein Lottogewinner wird vorgestellt. Anna sagt: Na sowas! Das ist doch Herr Schiefer/Schiff/ Schief! Unser neuer Nachbar!

But closer examination reveals that they are situationally quite different. The German context *Na sowas! (Well, I never!)* introduces the attitudinal feature of surprise, whereas the English context *Look, Peter!* stays neutral in this respect. An observation that is contrary to the speaker's expectation, which is further highlighted by the modal particle *doch* in the test sentences, provides the context appropriate for a phonological late peak in German. It is thus highly probable that the German speakers used a different peak category from the English speakers - late vs. medial peak, and the alignment differences Grabe found are not phonetic but phonological timing distinctions in different behavioural contexts. So not only is TIME to be introduced into the definition of the intonational categories, experimental data collection must also pay much more attention to the functional aspect of speech behaviour in order to interpret language variability adequately.

Finally, mainstream intonation analysis lacks concern for the LISTENER as a processor of acoustic messages in speech communication, because, again with the exception of the Dutch School, it focuses on production.

This paper illustrates the need for the systematic introduction of FUNCTION, TIME, and the LISTENER into intonational phonology with reference to research at IPDS Kiel over almost 25 years, drawing on experimental work and descriptive analyses of large corpora of unscripted speech in German dialogue.

2. Function

2.1. Language vs. paralanguage and lab speech vs. spontaneous speech

Ladd [12] singles out the *'Linguist's Theory of Intonational Meaning'*, the central idea being that "the elements of intonation have meaning" (which is trivial), and that "these meanings are very general", but are part of a system that "has a categorical linguistic structure". The definition is rather vague and rests on a basic separation between language and paralanguage, between linguistic and paralinguistic, which is anything but clear and straightforward. Paralinguistic elements have to be filtered out before linguists can get at their linguistic type of intonational meaning. This filtering process is achieved by typical linguistic methodology: linguists work out language examples, usually of sentence length (and often of doubtful semantics) at their desks, check them with colleagues and students and finally fill them with phonetic substance under lab conditions. In these circumstances paralanguage is unlikely to interfere, so data collection itself provides the filter. Moreover, the construction of sentence

types prejudges the phonological categories to be investigated.

The situation changes drastically when real life interactions, or at least close approximations, become the centre of attention. Then linguistic and paralinguistic phenomena, the output from all functions served by speech, are closely interwoven and have to be dealt with. In this diversity of communicative reality, searching for the *'Linguist's Theory of Intonational Meaning'* is like the unreal retreat to a desert island. Today there is a lot of talk about spontaneous speech, but what linguists often move to when they transgress the sentence level is spontaneous only in the sense that it is unscripted, e.g. the MAP Task paradigm has very restricted ingredients of spontaneity and of functions beyond the linguistic one, and it is not even totally unscripted. The Verbmobil corpus is better, but not ideal. The break-through in semi-controlled spontaneous speech data collection is Peters' 'Daily Soap Scenario' (<http://www.ipds.uni-kiel.de/pub_exx/bp2001_1/Linda21.html>), although it needs further refinement and elaboration, too.

2.2. Direct link between phonetic substance and behavioural function

To deal with prosodic structure of spontaneous speech adequately, we have to turn to the other functions, which have always been given preferential treatment outside linguistics, in life sciences, such as psychology, sociology, anthropology, psychiatry. In addition to the functions of sentence modality, of focus, of syntactic phrasing we need to include the attitudinal and expressive function, the interactive function (i.e. turn-taking between speakers in dialogue) and the guide function (i.e. prosodic continuity and prosodic structure of the preceding context guiding perception in real speech situations).

As long ago as 1981, Barry [2] proposed a progression from the phonetic base to communicative functions and stressed two points: (a) to consider function systematically in the very analysis of prosody, and (b) to go beyond the traditionally overemphasized linguistic function. Ohala's *'Frequency Code'* is an attempt, within phonetics and linguistics, to proceed from the phonetic base to the communicative function by relating phonetic substance, namely high vs. low F0, to social behaviour, viz. subordination vs. dominance, irrespective of linguistic form. With reference to such external relationship between phonetic substance and social function, we can explain linguistic form, for example the use of high or rising F0 in questions in the languages of the world. This way an explanans principle - functions of F0 - is independent of the explanandum - the use of F0 in linguistic structures, e.g. questions, and the explanation avoids circularity. But this also means that we have to give up the *'Linguist's Theory of Intonational Meaning'*.

Instead, the recognition of different functions and the expansion of their scope, becomes the prerequisite for the prosodic analysis of spontaneous speech, where the whole spectrum of communicative features of meaning interact and produce prosodic exponents that would be forced into a Procrustean bed, or cannot be handled adequately at all, by prosodic categories that are established entirely within linguistic structures. There have been other attempts within linguistics to overcome the dominating focus on intonational phonology in a narrow linguistic sense, and to consider other functions in connection with discourse and pragmatics. Gussenhoven, particularly in his contribution to "Speech Prosody 2002" in Aix transcends the realm of the *'Linguist's Theory of Intonational Meaning'* by linking the expression of universal meanings to biological codes.

2.3. The attitudinal function in questions and emphasis

Uldall's study of 1960 [13] demonstrated the importance of the attitudinal function of pitch in conveying meaning. Unfortunately, her semantic differential paradigm was lost in mainstream intonation analysis. But it was revived in a paper by Ernst Dombrowski at the "15th ICPhS".

The attitudinal function may be demonstrated by the use of F0 in 'yes-no' questions. For English and German, text books attribute rising F0 to them, as against falling F0 in question word questions. Although this distribution is confirmed in the analysis of large data bases as statistical tendencies, both F0 patterns occur with each of these grammatical structures, but they signal different attitudes. In *würde Ihnen das passen?* ("would that suit you"; g091a013 of the "Kiel Corpus", female speaker ANS), a final rising contour occurs and conveys consideration for the listener: the speaker leaves the decision 'yes' or 'no' open and does not indicate her own expectation. The same utterance resynthesized with a final fall suggests that the speaker prejudges the answer 'yes'; a *late* fall adds the feature of 'irritation' to the attitude signalled by the fall. In *haben Sie denn einen Termin noch im Mai frei?* ("have you got a free date for an appointment in May still"; g092a000 of the "Kiel Corpus"), the same female speaker has a final fall to signal her confidence in finding a mutually suitable date in May, since she has just thanked her dialogue partner for the general invitation to come and see her. The resynthesis with a final rise produces a real request to suggest a date. In the case of an early rise, which signals casualness, the request sounds matter-of-fact. The late rise, on the other hand, introduces friendly concern for the addressee.

Attitudinal function is also essential in the use of a force accent beside a pitch accent, as in *wie ... Boris ... Valerie die Treppe runterkickt* ("when Boris kicks Valerie down the stairs"; l061a018 of the Daily Soap Corpus "Lindenstraße"). Both *Valerie* and *Treppe* have peak contours associated with sentence accents, 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 an emphatic accent. 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 /ɪ/, 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 the 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.

2.4. The guide function

The guide function has several facets.

- Prosodic continuity enables a listener to separate two speech streams. Darwin [3] demonstrated that in ear switching experiments of dichotically presented sen-

tences listeners followed the integral prosodic pattern into the other ear in a shadowing task in spite of syntactico-semantic ruptures.

- Temporal regularity of rhythm in opening and closing gestures of the vocal tract, and of sequences of pitch accent patterns in pitch accent languages, such as German and English, provides listeners with expectations of upcoming speech events and thus guides them in understanding messages.
- The guide function is further operative in production across false starts or hesitations and subsequent pauses where intonation continues after the interruptions as if they had not occurred. The pitch pattern across the articulatory dysfluency forms a global unit that glues the utterance together in spite of internal breaks. It is this holistic pitch contour that gives the listener the perceptual cue that local disturbances of articulatory timing are to be interpreted as paralinguistic dysfluencies, rather than planned prosodic-syntactic-semantic phrase boundaries.

Thus the guide function brings in the LISTENER as an essential category in the modelling of prosody. But the listener is also important at the metalinguistic level of data processing (e.g. labelling), where modern lab phonology approaches tend to ignore the ear and to rely on measuring acoustic parameters, which are then coded, e.g., in binary terms H/L, as sequences of local points. Phonology sets up linguistic units without explicit reference to speaker and hearer, so prosodic categories within a strictly linguistic framework lack essential aspects of the communicative situation. The importance of the category of the listener for intonational phonology will be further developed in the next section of this paper.

3. Timing of F0 contours and the listener

3.1. The peak shift experiments: synchronization of pitch contours with articulation

In 1982/3, W. J. Barry and I gave an advanced seminar series on "Functions of intonation and their phonetic manifestations" at IPDS Kiel, where we elaborated the association of pitch with different functions. In particular, we studied the semantics and pragmatics of different falling F0 patterns in the German sentence *Sie hat ja gelogen* ("She's been lying"). An early fall on *gelogen* may be paraphrased as 'knowing', a later fall as 'recognising/realising/learning', a very late fall as 'surprise'. In more general terms, the distinctions could be defined as 'opening' or 'concluding an argument' or 'assessing the situation in contrast to one's own expectation'.

After preliminary production data analysis, a new paradigm was developed for perceptual experiments, in which a whole F0 peak contour was shifted in equal time steps through a segmentally constant utterance to test perceptual pitch changes in the stimulus series as well as their associated semantic features. It introduced two further concepts into intonation research:
(1) The synchronization of phonatory and articulatory time courses became a feature of intonational phonology (just as, several years later, articulatory phonology incorporated the temporal aspect into the segmental strand).
(2) The essential intonational components were assumed to be global contours (peaks, valleys), rather than local pitch levels.

The paradigm was later elaborated in a long-term German Research Council project "Form and function of intonation peaks". The perception results established three peak positions for the prosodic phonology of German, with a categorical change from early to medial peak, and a more gradual one to the late peak. The semantic changes are also clear-cut. On this basis the "Kiel Intonation Model (KIM)" was developed [10,11]. In this model, the listener assumed a central role.

3.2. Shapes of peaks: pitch-internal timing

Latterly, a second time dimension was introduced into the experimental paradigm in two MA dissertations at IPDS Kiel: the internal timing (shape) of the peak contour, i.e. the speeds of the peak rise and the peak fall (each slow vs. fast). Oliver Niebuhr tested the effects of this category with extensive listening experiments in German and found strong interactions with the position change from an early to a medial peak: a fast rise + slow fall produces a significantly earlier percept change from early to medial, compared with a slow rise + fast fall (see paper at "15th ICPhS"). Similarly, Tamara Khromovskikh tested the internal timing effects on the coding of statements vs. 'yes-no' questions (which lack syntactic markers) in Russian ("Perzeptionsuntersuchungen zur Frage im Russischen", 2003). Here the two sentence categories are coded by early vs. late peak positions, combined with slow rise + fast fall vs. fast rise + slow fall (and by lower vs. higher peak value).

These data may be supplemented by findings from a number of other languages that show effects of the same combination of synchronization and internal timing in pitch peak contours, but with varying functions: narrow-focus statement vs. question in Bulgarian [1]; narrow contrast vs. broad focus in Pisa Italian [6]; statement vs. question in Neapolitan Italian [5]; command vs. question in Bari Italian [8].

3.3. Explaining the data

We need to propose an explanation as to why the perceptual changes depend on synchronization and internal timing in the way they do, and why specific feature combinations in peaks may be used to differentiate between 'opening' and 'concluding argumentation', 'question' and 'statement/command', 'broad' and 'narrow focus'. This may be achieved with reference to two theoretical considerations: *'auditory contrast'* in contours at specific syllable points (*'auditory enhancement hypothesis'*, cf. Diehl [4]), and the *'Frequency Code'*.

The Kiel findings on German 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 accented 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 perceptual salience of a 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. In all these cases, the contrastive strengthening of high or low pitch 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 low-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 sequences of HIGH and LOW pitch level points of AM phonology and ToBI.

What seems to be essential for perception here are not the

low and high points by themselves, but the contrast of a LOW with a preceding or following HIGH in synchronization with different parts of the accented syllable. This contrast is first of all established by different peak positions in relation to the accented vowel onset, but is strengthened by a slow rise before the accented vowel in the early peak, or by a fast rise in the accented vowel in the later peak, respectively. If the former also has a fast fall and the latter a slow one, the contrast gets even sharper. This contrast leads to a strong focus on low or high pitch in the accented vowel, which may be heightened still further by increased peak point F0 in the case of the later peak. This focus on low or high pitch range in the accented vowel, respectively, is linked to the semantics of 'finality' vs. 'openness', which include 'dominance' vs. 'subordination'. We are thus dealing with another instance of Ohala's *'Frequency Code'*.

All the peak alignment data and the functions they serve in the different languages can be subsumed under the same two principles of *'auditory enhancement'* and the *'Frequency Code'*. This applies not only to the form-function relationship of the German data, already explained in this theoretical frame. The later, faster rising and higher F0 peak configuration also contains all the ingredients for a low-high pitch contrast in an accented vowel to mark the question function, and vice versa for the statement/command, in direct line with the *'Frequency Code'*. Narrow focus singles out an element of an utterance by a fast falling peak contour, sounding final. Broad focus leaves the scope open, signalled by a peak contour that stays high longer.

3.4. ToBI accounts of the data

Grice and Baumann [9] refer to two early peak patterns in German, which they classify as H+!H* and H+L* in GToBI notation. Their empirical observation is new and accurate, but a dialectal feature has to be taken into consideration as well. In South German dialects all peak contours are characterised by slower descents than in Northern German, e.g. late peak in *'hasch a blaues Wohnmobil?* ("have you got a blue caravan") or early peak in *mer stellt sich einfach dahinner* ("you simply join the queue"). So Grice and Baumann may actually be dealing with dialectal traces.

They associate the downstepped H+!H* with my early peak, which is actually H+L* in their notation, and with 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 patterns, both are characterised by 'finality' vs. 'openness', but in the more slowly falling one this finality is weakened, again 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, but does not bring out either what the patterns have in common and what separates them in a universal perspective of human speech communication.

The suggested AM/ToBI representation for the Italian question - command data is H* and L+H* (L-L%), i.e. presence/absence of a low leading tone. This solution does not capture the relationships, neither with regard to the pitch contrast, nor with regard to the low vs. high pitch focus on the accented vowel. I would maintain that it is not the presence or absence of an L point that is essential, but the opposition of a prototypical early vs. a prototypical later peak contour: the former is characterized by a high-low trajectory through the vowel-consonant transition, after a slow rise, the latter by a fast low-high trajectory through the same critical syllable position, followed by a fall; the fall may also be faster vs. slower, if there is enough articulation time available.

4. Outlook

I wanted to show that our analysis of prosody and intonation is not just a metalinguistic academic pursuit but actually aims at describing and explaining language behaviour in realistic communicative situations, and therefore the prevalent paradigm of the *'Linguist's Theory of Intonational Meaning'* needs to be amended through the incorporation of such central concepts as FUNCTION, TIME and the LISTENER. There has been a fair amount of rumbling, for instance at "Speech Prosody 2002", announcing growing unease with mainstream prosodic theory and practice, with ToBI among others, and this campaign will no doubt gather momentum in the years to come.

5. References

[1] Andreeva, B.; Avgustinova, T.; Barry,W. J., 2001. 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.). Frankfurt/M.: Peter Lang, 353-364.

[2] Barry, W. J., 1981. Prosodic functions revisited again! *Phonetica*, 38, 120-134.

[3] Darwin, C. J., 1975. On the dynamic use of prosody in speech perception. In *Structure and Process in Speech Perception*, A. Cohen, S. Nooteboom (Eds.). New York: Springer, 178-193.

[4] Diehl, R., 1991. The role of phonetics within the study of language. *Phonetica*, 48, 320-340.

[5] D'Imperio, M., 2000. *The Role of Perception in Defining Tonal Targets and their Alignment*. Ohio State University: PhD thesis.

[6] Gili Fivela, B., 2002. Tonal alignment in two Pisa Italian peak accents, *Proc. Speech Prosody 2002*. Aix-en-Provence.

[7] Grabe, E, 1998. *Comparative Intonational Phonology: English and German*. Nijmegen University: PhD thesis.

[8] Grice, M.; Savino, M., 1995. Low tone versus 'sag'in Bari Italian intonation, *Proc. 13th ICPhS*, vol. 4, 658-661.

[9] Grice, M.; Baumann, S., 2002. Deutsche Intonation und GToBI, *Linguistische Berichte*, 191, 267-298.

[10] Kohler, K. J., 1991. *Studies in German Intonation*. AIPUK, vol. 25. Kiel: IPDS.

[11] Kohler, K. J., 1997. Modelling prosody in spontaneous speech. In *Computing Prosody*, Y. Sagisaka, N. Campbell, N. Higuchi (Eds.). New York: Springer, 187-210.

[12] Ladd, D. R., 1996. *Intonational Phonology*. CUP: Cambridge.

[13] Uldall, E.T., 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech* 3, 223-234.