I

Klaus J. Kohler

Form and Function of Intonation Peaks in German: A Research Project

1. Origin and orientation of intonation research at Kiel

In 1983, the German Research Council (DFG) instituted a new programme "Forms and Functions of Intonation" to encourage interdisciplinary research into prosody, combining experimental and instrumental analysis with linguistic description and explanation. This framework imposes a number of constraints on the direction in which work has to proceed.

1.1 Methodological constraints

1.1.1 Separation of macro- and microprosody

If measurements are an essential component of the scientific investigation, it must be known what is to be measured and what the instrumental data represent. In the case of fundamental frequency curves it must, therefore, be ascertained what are intended contrasts, and what is contextually conditioned variation. This leads to the separation of global, meaning-related macro- and local, production-related microprosody (Hirst, 1981), and to the systematic analysis of purely phonetic, articulatory modification as a prerequisite to the study of prosodic systems and their linguistic functions. So, instead of trying to eliminate microprosody by various procedures (e.g. in the Dutch school: Cohen & 't Hart, 1967; Scheffers, 1988), it must be integrated into intonation research as an essential component.

1.1.2 Macroprosodic constancy in microprosodic analysis

The investigation of microprosody necessitates the constancy of macroprosody across different segmental chains, which only trained speakers, preferably trained phoneticians, can approximate. This determines the method of data collection: carefully constructed sentence material read by trained subjects under laboratory conditions, rather than continuous texts or even spontaneous speech. The data recording controls specific, constant macroprosodies in various segmental contexts with regard to the following question:

What changes have to take place in an F0 contour to guarantee the identity of a linguistic stress or intonation pattern across *different* segmental strings (= micro F0)?

This defines one of the research areas in the Kiel Intonation Project, following logically from the conceptualization of the DFG programme.

1.1.3 Production data heuristics and perceptual test procedures

The constancy of macroprosody across different articulatory strings remains a problem even with trained phoneticians. Although they can easily avoid changes from rising to falling patterns, or listing intonations, they are nevertheless not immune to slight changes in the timing and extension of peak and valley contours, as well as in downstepping and resetting, all of which may not only change the phonological categories of the intonational elements, but also interfere with the manifestation of microprosody, i.e. the object of investigation. It is, therefore, to be expected that the results vary from one analysis of microprosody in speech production to another, all the more so, the fewer constraints are imposed on the data collection. The discrepancies of the values for intrinsic and cointrinsic micro F0 modification found in the literature have their explanation in this methodological problem (see also di Cristo, 1985; di Cristo & Hirst, 1986). This means that production data can only have a heuristic function stimulating questions in the perceptual domain, which provides an obligatory complement. The systematic and controlled parameter manipulation in computer signal processing, on the basis of hypotheses derived from preliminary production data analysis, is thus a further logical development of the research method geared towards interactive perceptual evaluation and/or formal listening tests.

1.1.4 Macroprosodic analysis on a perceptual basis

The ultimate aim of the separation of micro- and macroprosody and of the detailed analysis of the former is to determine the structural elements of an intonational phonology that are able to code linguistic functions, i.e. to find answers to the question

When do changes in the F0 contour (e.g. shifts of F0 peaks or valleys) across the *same* segmental string effect changes in linguistic patterning (= macro F0)?

This defines the second research area in the Kiel intonation project, again following logically from the conceptualization of the DFG programme.

The macroprosodic question cannot be answered by a production data collection at all, because the definition of a syntactic or semantic function in the instructions to subjects or in simulated contexts does not necessarily elicit one determined phonological intonation category and even less so well

delimited parameter values. Moreover, informants' imagination plays an enormous role in what they produce. Prosody does not only code syntactic structures and semantic features of a linguistic system in the narrow sense, but also an expressive component and attitudes to the communicative partners, the task, the situation etc., i.e. paralinguistic attributes. Thus the great variety of phonetic patterns obtained in controlled productions of a particular syntactic or semantic function, rather than constituting one large phonetic scatter of a specific linguistic category, may represent the phonetic manifestations of different phonological units, which can, in turn, have one-many and many-one relationships to elements of meaning. That is why in this field perception analyses are essential for reaching the scientific goal, and, consequently, the second research question in the Kiel Intonation Project also makes systematic signal parameter manipulation on the computer mandatory.

1.1.5 Perceptual differentiation and phonological categorization

This kind of investigation at first only shows up the perceptual relevance of particular parameter values and does not say much about their function. The perceptibility is, however, a prerequisite for the signalling of functions, and the analysis of perceptual differentiation thus leads, first and foremost, to phonological categories of intonation in a particular language. But it also goes beyond the individual prosodic system by asking the question as to human speech perception in general, i.e. the universality or at least wide distribution of certain 'acoustic stimulus/auditory percept' relationships in the tonal features of languages.

Up to this point, the research strategy consists in the transformation of phonetic substance into phonological form via perceptual categorization on the basis of hypotheses from a heuristic production data analysis. Ranges of measureable properties (F0, segment or syllabe duration, spectrum and intensity) are associated whith phonological elements of a macroprosodic system. Within this system, stress and intonation must be differentiated, and among the latter three F0 peak positions relative to the stressed vowel onset - early, medial, late - have to be recognized. Changes in F0 contours may alter the melodic pattern without affecting the stresses (e.g. rising vs. falling or early vs. late peak), or they may change the stress but keep the melodic pattern (e.g. early peak on a different stressed syllable within a

single-accent utterance), or they may change both simultaneously (e.g. early vs. late peak on two different stressed syllables). The stress and intonation categorizations are abstractions at different levels: lexical stress, sentence stress, and intonation configurations, such as peaks, valleys, hat patterns, which may all have a variety of measureable realisations in different phonetic strings, in male vs. female voices, in lively vs. dull expression etc., without necessarily changing their phonological status.

1.1.6 Phonological form and linguistic function

The question now arises as to what linguistic functions (syntactic, semantic, pragmatic) are coded by this phonological form. The investigation should in particular unravel the *underlying* semantic categories associated with elements of the prosodic system, which, combined with the syntactic structures and lexical semantic categories, result in a *surface* semantics, which is observable as such, and can be paraphrased by informants or experimenters. In direct questioning of informants as to the meaning of intonationally different utterances, the elicited open metalinguistic judgements will always be influenced by the surface semantics and largely depend on the subjects' extremely variable powers of imagination; they can therefore only provide indications - some better than others - about the underlying semantic features of a linguistic analysis, but they do not necessarily capture the latter. A factor analysis over many response scales according to the principle of the semantic differential is undoubtedly a better procedure but puts high demands on the subjects and is highly disturbed by the lexical semantics of the utterances to be judged. Successful factor analyses of this type have been applied to rather limited, semantically neutral single-word utterances, such as "yes/no" (Richter, 1967). A further paradigm for the analysis of semantic functions of intonation is the contextualization of an utterance to be assessed and the elicitation of "matching/not matching" responses. This is, of course, only possible when the experimenter has developed hypotheses about the semantics of prosodic patterns.

The setting up of the underlying semantic categories can make use of all these procedures in order to gather hints from informants in empirical tests as to the components of meaning, but these hints must not be mistaken for the semantic categories themselves. The linguistic intuition of the

native-speaker experimenter also plays a fundamental role in the abstraction of the relevant semantic categories from the surface utterance meaning. So the analysis cannot be inductive, but must proceed deductively, by postulating semantic structures on the basis of certain preliminary observations and of intuitions, and by pursuing and evaluating the consequences resulting from such postulates for the description of a wide array of data. This means that the results obtained from the study of isolated sentences will also have to be projected onto corpora of natural continuous speech production and judged for their applicability and appropriateness.

1.2 Characterization of the Kiel Intonation Project

1.2.1 Phonetic substance - phonological form - linguistic function

The Kiel project may be briefly characterized as "from phonetic substance to phonological form to linguistic (syntactic, semantic, pragmatic) function". The important points are that the phonetic-semantic relationship is not direct in the sense that the measured values themselves represent semantic categories, but that the link operates via formal elements that, on the one hand, are related to features of meaning, but are, on the other hand, defined by phonetic ranges. This phonetic substantiation of phonological categories is just as essential as the recognition of structure in phonetic substance. Both phonetic substance and phonetic structure (or signal measures and phonological form) are required for an adequate description of the phonetic-semantic relationship.

If phonetic structure is ignored and parametric values are directly related to linguistic function, the separation of distinctive contrast from random variation is impossible, and the statements about the link become false or at least misleading. For example, the measurements of all the possible F0 patterns produced in response to the request for a certain sentence mode (e.g. wh-question) will say very little about the expression of such a syntactic structure, even if classes are formed post hoc according to parametric properties and an elaborate statistics is carried out. What should be established first are the prosodic categories and subcategories that are involved (peaks vs. valleys, syllable alignment etc.), then the phonetic variation should be established within these formal elements, finally the question should be asked as to how the different phonological categories are

related to syntactic structures and further semantic differentiation.

On the other hand, if phonetic substance is ignored, the linguistic analysis is dissociated from speaking and hearing, the bases of speech communication, and any solutions can become possible on paper without being falsifiable by empirical data. Moreover, it is precisely the phonetic manifestations that allow generalisations beyond the individual language to the use of prosody in human speech on the whole. Thus the theoretical stand advocated for intonation research again stresses the unity of phonetics and phonology, as was done with regard to segments in Kohler, 1991e. At the same time, this is a plea for turning phonology into an experimental discipline and taking it into the laboratory (Ohala & Jaeger, 1986; Kingston & Beckman, 1990).

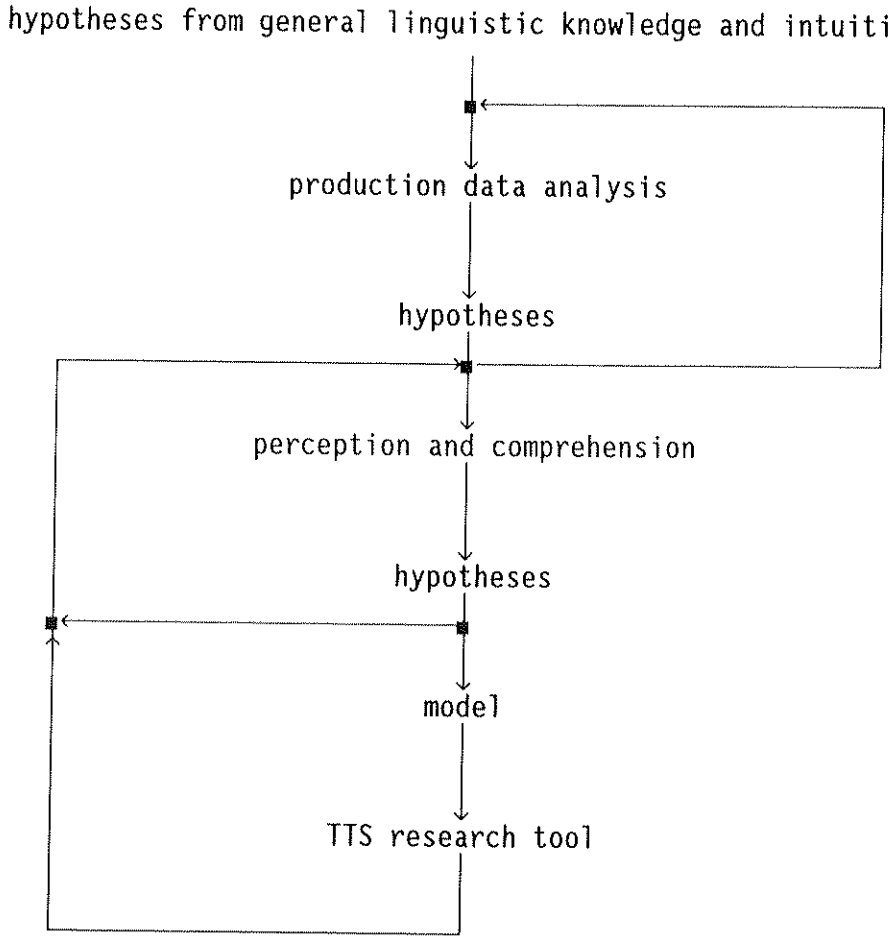
1.2.2 Points of departure and steps of progression

The Kiel Phonetics Institute joined the DFG research programme on intonation as from 1 January, 1985. Taking previous studies on microprosody (Kohler, 1982, 1985) and on the semantics of peak alignment (Kohler, 1978) as points of departure, the two research questions formulated in 1.1.2 and 1.1.4 above became the central issues under the aspects of phonetic substance, phonological form and linguistic, especially semantic and pragmatic, function, and with the methodological constraints outlined in 1.1. The investigation was based on data from German and at first restricted to terminal contours.

The goal of the Kiel Intonation Project has been to establish the phonologically relevant prosodic elements and to separate them from their contextual modifications, i.e. to develop a functional model of German intonation on the basis of human pitch production and perception in speech communication, linking phonetic output with meaning. Based on a number of hypotheses, production data were analysed and micro- and macroprosodic patterns subjected to perceptual evaluation in formal listening tests as well as in interactive experimentation at the computer. From the results of these empirical investigations a formalised model was derived, which was implemented in a text-to-speech system within the RULSYS framework (Carlson, Granström & Hunnicutt, 1989). This implementation was then used interactively as a research tool and perceptual test instrument for reiteratively improving the model and expanding it to include non-terminal patterns.

1.2.3 Research methodology

Work on the Kiel Intonation Project adopted the following hypothesis-driven loop procedure of scientific empirical investigation towards theoretical model construction and linguistic explanation. Starting from initial hypotheses, based on general linguistic knowledge and intuition, systematic production data are collected and analysed, resulting in the modification and expansion of hypotheses, which are in turn translated into perception and comprehension experiments leading to new hypotheses. They may be tested and modified in further loops of perception and comprehension experiments and/or of production data collection according to very specific constraints imposed by the hypotheses. The result of each hypothesis testing is a partial, but ever more comprehensive model, which is implemented in a TTS research tool and made to generate data to be evaluated. The final, but still distant goal is a fully comprehensive generative model that outputs intonation contours from symbolic input of semantic and syntactic elements and relations, and which may be tested perceptually and evaluated against the comparison with natural productions. This research methodology is outlined in the following diagram:



1.2.4 Theoretical demands on a model of intonation

1.2.4.1 Generative vs. taxonomic models

In order to be an adequate and economic representation of the production and perception of intonation contours in the speech communication process the model cannot work on the basis of an inventory of tunes or melodic elements. Such a taxonomic approach misses important generalizations and has to change the model whenever new elements are needed to cope with empirical findings under a variety of conditions (speech rate, semantic, pragmatic and expressive aspects). It has problems factoring out the interaction of microprosody, stress, and intonation; it needs to add new patterns all the time to cope with the three phenomena. Intonation models should therefore be developed in a generative framework, which postulates underlying elements to which (at least partly) ordered rules are applied at a number of different levels to yield an empirically testable F0 output. The syntactic, semantic, pragmatic and expressive functions of F0 are then not linked to the output contours directly, but to the underlying elements and to the various levels of the rule system. This way the demand for a functional model of intonation can be met, allowing the transformation of basic patterns by the introduction of several layers of meaning-induced components, i.e. successive modification according to superimposed functional factors.

1.2.4.2 Pitch contours vs. pitch levels

The Kiel Intonation Project has followed the British tradition of intonation research (e.g. Halliday, 1967), rejecting the atomization of pitch contours to pitch level points upon the rationale that speakers control contour shapes starting at certain structurally determined points in time, rather than a succession of levels. Intonation peaks, for example, cannot be decomposed into the elements of rise and fall, which is clearly shown by medial peaks - and even more so by late ones -, where both movements are needed for the patterns to be perceived as such. The fall may be deferred to a later accent in a hat pattern, but it must occur, otherwise the pattern changes. Furthermore, rises in hat patterns are fast in the same way as in single-accent medial or late peaks.

It is only the early peak pattern that does not need to have a rise, but a high F0 must be indicated before the accented vowel, either by a high-pitched unstressed syllable or voiced initial consonant, or by way of extrapolation

from a very low fall in the accented vowel after an utterance-initial voiceless consonant. This is the reason why only the early peak can follow the other two peaks in a hat pattern. In the case of a medial or a late peak at the end of such a configuration the peak has to be raised above the top level of the hat to guarantee the contour identity through a rise-fall.

But even early peaks can have a rise in syllable concatenation, e.g. in single-accent utterances with several initial unstressed syllables, or utterance-finally in a dipped peak sequence, where the rise is neither an element belonging to a preceding peak nor is it independent because it is fast rather than being the separate slowly rising element of continuation. Furthermore, in order to capture the systemic relation between all peak types, e.g. in peak shift experiments, it is best to regard the early fall as an early peak.

1.2.4.3 Downstep vs. declination

The Kiel Intonation Project has held the view that declination, i.e. the temporally fixed decline of F0, is not a feature of natural speech production. The F0 decline from stress to stress position does not occur on a time basis, but on a structural one, i.e. we are dealing with a constant downstepping from stress to stress, independent of the time that elapses between them, and the downstepping constant can vary according to focus, utterance type etc., and can at any moment, be changed - re-set - by the speaker in compliance with the meaning and intentions to be transmitted.

1.2.4.4 Prosodic units vs. prosodic features

Ladd's (1983a) feature approach to intonation has been further developed, and independent serial prosodic *units* have been eliminated altogether, prosodic *features* for the specification of peak and valley configurations (e.g. ±FSTRESS, ±EARLY, ±LATE, ±TERMINAL) being attached to vowels at the segmental level, triggered by syntactic, semantic and pragmatic categories. The prosodic features, in turn, control parametric time/F0 value pairs at specific points.

1.2.4.5 Production vs. perception models

The Kiel Intonation Model (KIM) is basically a production model, i.e. it generates pitch contours from symbolic input, although it has very strong perceptual components as a result of the weight of perceptual evaluation in

the model development. It will, however, have to be expanded to include a complete perception module to make it a real model of speech communication, but there is no doubt that this is the more difficult task to achieve. Questions such as the symbolic decoding of stress and intonation peaks from incoming signals are very thorny indeed.

2. Hypotheses in the Kiel Intonation Project

2.1 Hypotheses on F0 peak configurations

The empirical investigations on the Kiel Intonation Project were triggered by four logically ordered and temporally sequenced hypotheses on

(1) the microprosodic modifications of basic macroprosodic F0 peaks,
(2) the phonology and semantics of F0 peak alignment in single-accent utterances,
(3) the changes and interactions of stress and intonation brought about by shifting a single F0 peak through an utterance with more than one potential accent postion,
(4) the implementation by F0 peaks of more than one accent in an utterance.

Each hypothesis will be the subject of at least one contribution to this AIPUK volume.

2.1.1 Hypothesis (1): Microprosodic modifications

The same intended global macroprosodic terminal pitch contour can be associated with any sound segment string by postulating a basic underlying continuous F0 peak configuration and by subsequently adjusting it according to the following principles:

(a) temporal alignment as a function of stressed vowel type,
(b) expansion or compression of the F0 peak configuration in the time domain as a function of the number of syllables and their complexity,
(c) intrinsic F0 modifications in vowels as a function of their tongue height,
(d) co-intrinsic F0 modifications as a function of the voiced/voiceless or lenis/fortis dichotomy in preceding and following consonants,
(e) masking of F0 in voiceless signal stretches.

The first step in working on hypothesis (1) was to be a heuristic one, taking a continuously voiced, naturally produced utterance token - consisting only of vowels and sonorants - as a point of departure for a transfer of F0

contour sections to other segment chains and for subsequent modifications according to (a) - (e) in an interactive graphic-auditory procedure at the computer terminal to establish perceptual congruence between the mapped and adjusted patterns, on the one hand, and the originally produced patterns in the same sound sequences, on the other. Microprosodic rules were to be developed covering the widest possible array of data and therefore representing the greatest generality. In a later step, significant points were to be abstracted from the empirical data to define the basic peak patterns by a minimum number of parametric values as input into the microprosodic rules. Contribution II (Gartenberg & Panzlaff-Reuter, 1991) gives a detailed account of the work and results in production data analysis and formal perceptual testing connected with hypothesis (1), up to the point of the model construction and interactive evaluation. That means that the description of the model in the last contribution (VII) will differ with regard to some parametric value specifications, which became necessary in the application of the TTS research tool to data generation and assessment. It was nevertheless thought necessary to present the empirical data analyses, not only to give a faithful account of the findings, but also in order to point out their heuristic nature in a hypothesis-driven theory construction and refinement.

2.1.2 Hypothesis (2): F0 peak alignment

If an F0 peak in a single-accent terminal utterance is shifted left from a medial position in the stressed syllable nucleus a categorical change occurs in perception, which is correlated with a semantic change along the dimension 'new/established'. The corresponding realignment to the right produces a gradual auditory change correlated with a semantic continuum expressing degrees of emphasis.

This hypothesis addresses the question of F0 cuing the phonological variation of *intonation* for constant *stress* with reference to peak patterns. Contribution III (Kohler, 1991c) presents the experimental data related to hypothesis (2) and their perceptual, phonological and linguistic interpretation.

2.1.3 Hypothesis (3): Stress and intonation

If there is more than one potential accent position in a single-accent

terminal utterance - either at the lexical or at the sentence stress level - hypothesis (2) holds at each accent position. But now an F0 peak shift does not only alter the phonological categories of *intonation*, but also that of *stress*, moving its position to different syllables. From this double cuing power of F0 peaks for *intonation* and for *stress* three corollaries are derived:

(a) *Stress* and *intonation* may interfere with each other.

(b) Other cues beside F0 (segment and syllable duration, intensity, vowel spectrum) have to be taken into consideration when F0 is not sufficient for the signalling of *stress*.

(c) The height of an F0 peak cues the prominence of a stressed syllable, but at the same time contributes to the identity of an *intonation* category.

Hypothesis (3) and corollaries (a) and (b) are dealt with in Contribution III (Kohler, 1991c). Corollary (a) is further treated in Contribution IV (Hertrich, 1991a), and corollary (c) is the subject of Contribution V (Kohler & Gartenberg, 1991).

2.1.4 Hypothesis (4): Peak accent sequences

If an utterance is to have more than one accent the series arises from a concatenation of peaks with hypotheses (1) - (3) applying to each one. But additonal contours have to be recognized as further derivations beyond the simple concatenation:

(a) The peaks are downstepped if they are all to have the same perceptual prominence and semantic weighting.

(b) The dips between the peaks are reduced, or even eliminated in a hat pattern, if the cohesion of the utterance is to increase and the prominence of the peaks to be lowered.

Contribution VI (Hertrich, 1991b) looks at the auditory relevance of a large variety of F0 patterns for two-accent utterances. The theme will then be picked up again in Contribution VII (Kohler, 1991d).

2.2 Further hypotheses and the development of an intonation model

The research on the basis of the four hypotheses outlined in 2.1 prepared the ground for the construction of a peak model of German intonation (KIM), which will be presented in Contribution VII (Kohler, 1991d). It has been extended

to include valley patterns as well, derived from the following hypotheses.

Hypothesis (5): F0 valleys and their alignment

The principle of aligning peaks to accented syllables in phonologically different ways can also be applied to two types of valleys, i.e. to low or high rises in continuation and question structures, which may start before the accented syllable nucleus or well inside it, with a clear perceptual differentiation and a semantic classification along the scale 'casual/interested'.

Hypothesis (6): Fall-rises

From the three terminal patterns - early, medial, late peaks - further phonological elements can be derived by adding a final rise to each one of them, with similar perceptual and semantic distinctions.

3. Other approaches to the study of prosody: similarities and differences

The model developed within the Kiel Intonation Project being generative has a great deal in common with the models proposed by Gårding and Bruce (Gårding, 1979, 1982; Bruce, 1977, 1982), Thorsen (1979), Pierrehumbert (1987), Hirst (1981, 1983) and Ladd (1983a,b). The similarities are greatest with Ladd in view of the common feature approach (see 1.2.4.4), but KIM does not adopt Ladd's (as well as Pierrehumbert's and Hirst's) tone sequence model, but advocates the concatenation of peak and valley contours (see 1.2.4.2), and the separation of intonation and accent (see 2.1.3), as is done in the models of Gårding, Bruce and Thorsen. Pierrehumbert & Steele (1989), although still using the tone sequence symbolization, also come very close to the conceptualization of peak patterns as whole units of intonation with different syllable alignments.

Whereas KIM makes use of downstepping (see 1.2.4.3), the Lund model relies on the declination grid as an underlying setting for different utterance types without, on the one hand, being able to determine convincingly the grid and its pivots for change of direction from F0 records of empirical corpus data, and without, on the other hand, showing that the pitch contours generated from different grid settings within a comprehensive rule system closely match natural productions. Because of the problems in production data collection and analysis (see 1.1.4), KIM does not accept the association of sentence

type and underlying grids, but this does not exclude the possibility of global (phonological) modifications of, e.g. peak sequences, related to sentence mode; on the contrary, Ladd's (1983a,b) criticism of the 'contour interaction theory' is not considered well founded. Similar considerations apply to Thorsen (1979, 1983), although in her case the objection is less strong, as Thorsen does not postulate a predetermined grid into which peaks and valleys are fitted.

But there is full agreement with Ladd as regards the negation of a direct phonetic, particularly signal encoding of information about sentence type (question vs. statement etc.) *without a formal (phonological) interface that links phonetic manifestation with syntactic and semantic categories*. Thus the approach of Altmann, Batliner & Oppenrieder (1989) - another project in the same DFG programme -, which may be briefly characterized as "from linguistic function (sentence mode and focus) to phonetic substance", i.e. the inverse of the strategy followed in the Kiel Intonation Project, is not considered an acceptable framework (see 1.2.1). What is regarded as essential in KIM is to postulate basic prosodic elements defined by phonological features, to link them with ranges of signal parameter values, on the one hand, and with linguistic functions, on the other, to introduce feature modifications in this intonation phonology by rule at a hierarchy of levels in accordance with syntactic, semantic and pragmatic structuring, and to derive the parametric consequences from these feature adjustment rules.

KIM differs categorically from the taxonomic models, especially the Dutch school (e.g. Cohen & 't Hart, 1967; Adriaens, 1984, 1991), in that it maintains that a taxonomy of concrete pitch contours cannot adequately capture the intricate relationship between sound and function. As a consequence of the phonological approach, basic, abstract, underlying peak and valley configurations, defined by significant points, are postulated and transformed into observable F0 patterns by a set of ordered rules introducing syntactically, semantically and pragmatically as well as segmentally and contextually conditioned modifications. This kind of generative approach also implies

(a) linking meaning functions (syntactic, semantic, pragmatic, expressive) with the generation of phonological form in the very construction of the model, rather than asking the functional question post hoc after the

setting up of intonational elements, and

(b) integrating microprosody as an essential component, instead of trying to eliminate it as a disturbance of the taxonomic pitch patterns.

Looking at it differently, we may also say that to arrive at the basic intonational elements observable F0 contours are to be stripped not only of the microprosodic interference, but also of the multitude of layers that determine, one after the other, the signal output. To cope with this conceptualization a generative framework that starts from postulated abstract patterns in relation to linguistic function and adjusts them in a series of rules triggered by functional as well as phonetic components, is considered the only adequate solution.