ABSTRACT
On the basis of Karl Bühler’s Organon Model and the Kiel Intonation Model (KIM), this paper develops components of a paradigm of Speech Communication, which integrates the functions of Expression of the Speaker and Appeal to the Listener with the function of Representation of the Factual World into a comprehensive framework of language. In particular, prosodic manifestations are discussed of argumentation structure, of categories of emphasis, and of attitudes towards the listener in questions.

Keywords: information/argumentation structure, emphasis, questions, prosody

1. INTRODUCTION
Modern linguistics delimits its territory by many frontiers, creating such dichotomies as phonetics and phonology, linguistics and paralinguistics. The frontier between the former two fields has become permeable in Laboratory Phonology and has been broken down in Experimental Phonology. However, the currently prevalent paradigm locates paralinguistics at the linguistic frontier and centres on that function of the linguistic sign that has been called the representation of objects and factual relationships. In this wake, phonetic science is preoccupied, at the lexical level, with the differentiation of words by segments, stress and tone, and, at the utterance level, with the coding of phrasal chunking, sentence mode, focus and information structure by prosody. In tone languages, such as Chinese, the superposition of phrasal prosodies onto lexical tones has been studied very extensively, again largely limited to the linguistic function of representation. In such a framework, the expressive function of the linguistic sign, related to the sender, and its appeal function to the receiver are, therefore, excluded from the core domain of linguistics and relegated to its frontiers.

Since de Saussure, two main questions have dominated this linguistic perspective:
(1) What is the relationship between formal categories and their variation, more particularly between phonological units, such as the phoneme, and their phonetic substance?
(2) What is the relationship between formal categories and meaning?

In the pursuit of these questions, the formal categories are determined negatively by abstraction from observable phenomena according to the principle of otherness. For example, the phonemes /b/ and /p/ or the pitch accents H* and HL* must be different, and the phonetic measures, VOT or f0 alignment, must be such that they form two disjunctive sets with clear boundaries between them. This categoriality was also generalised to perception, resulting in the Haskins principle of discrete category identification coinciding with a maximum of discriminability across the category boundaries and a minimum inside them. The paradigm persists to this day in spite of the well-founded objections right from the moment it was put forward [28], and has even seen a renaissance in the field of prosody [36,3,7,35,38]. It reflects the assumption that speech communication is
based on these formal linguistic categories and their *discrete differences* in manifestation, rather than on *substantive identification*.

Phenomena linguists find difficult to reduce to such a frame of discrete otherness are relegated to *paralinguistics* and thus differentiated clearly from the formal level of linguistics. As long as the dichotomies of linguistics vs. paralinguistics and of phonology vs. phonetics are heuristic devices for making the phenomenology of speech accessible to scientific investigation they are useful in advancing our knowledge. Eminent phoneticians like Peter Ladefoged have dedicated a great part of their academic career to the study of sounds in the world’s languages, as used to differentiate words [25, 26]. Since to a large extent they work on unwritten languages that have not been studied before, they have no option but to approach the language by a very simple formal framework of sound units distinguishing words. Phrase-level and paralinguistic sound phenomena enter this research procedure as distorting noise.

However, if the central focus on formal linguistic categories is reified it makes researchers ask questions as to whether a particular phenomenon IS linguistic or paralinguistic, IS phonological or phonetic. Such questions give no insight into the functioning of speech communication, which should be the ultimate goal in speech science, because they result in futile ontological disputes and at the same time exclude a large part of speech interaction from scientific phonetic and linguistic investigation. Speech modelling on such a basis will always remain a torso, unable to handle everyday spontaneous speech communication fully and satisfactorily. The point has come when, in languages whose linguistic structures have been well-described, such as Arabic, Chinese, Danish, Dutch, English, French, German, Hindi, Italian, Japanese, Russian, Spanish, Swedish, we should pull down the formal scaffoldings that helped to build them, and develop new concepts that can bridge the dichotomies of phonology and phonetics, of segments and prosodies, and of linguistics and paralinguistics, and provide a more adequate theoretical basis for the insightful analysis of speech communication.

**2. A PARADIGM OF SPEECH COMMUNICATION**

**2.1 Theoretical bases**

**2.1.1 The Organon Model**

This paper illustrates a paradigm of speech communication which moves the expression of the speaker and the appeal to the listener into the centre of phonetic investigation, complementing the narrow linguistic functions. Its theoretical point of departure is Karl Bühler’s ‘Organon Model’ [4], which is schematically portrayed in Fig. 1. It relates the linguistic sign to the SENDER, the RECEIVER, and the FACTUAL WORLD of objects and factual relationships. This threefold link establishes the functions of Expression, Appeal, and Representation by *symptoms*, *signals*, and *symbols*, respectively. This three-faceted concept of the linguistic sign integrates paralinguistics and linguistics into a comprehensive field of speech communication. The speaker’s Expressions are attitudes towards the listener and the factual world in communicative settings, or emphatic evaluations, or emotions; the Appeal to the listener is carried by commands, requests and questions. Within this frame of Expression and Appeal, this paper discusses the prosodic manifestations of

- argumentation structure: concluding, opening, contrasting or expressively opposing argumentation, showing different kinds of speakers’ reactions in the unfolding of communicative interchange
- categories of negative and positive expressive intensification, as in “It stinks!” vs. “It’s marvellous!”.
- attitudes towards the listener in questions, prejudging answers or leaving them open to the listener, or expressing surprise.
The data will be mainly from German and English, but suggestions are advanced as to how this communicative paradigm can be introduced into the study of other languages, and how the phonetic exponents may be related to deep-seated features in human behaviour.

2.1.2 The Kiel Intonation Model (KIM)
To investigate the prosodic exponents of the Expression and Appeal functions, the categories of The Kiel Intonation Model (KIM) are applied to the data [12, 13, 14, 15, 16, 21]. KIM is a superpositional model of intonation, which parameterizes global patterns that are either (rising-) falling ‘peak contours’, (falling-)rising ‘valley contours’, (rising-) falling-rising ‘combined contours’ or ‘flat contours’, for the differentiation of meaning. This pitch – meaning relationship implies that the listener is the ultimate judge for categories in a prosodic phonology; patterns are established auditorily, and acoustic features are then related to the auditory categorization.

The movement patterns are further differentiated by different synchronizations with vocal tract dynamics: peaks may have their f0 maximum before the accented-vowel onset (‘early’), after the accented-vowel onset (‘medial’), or ‘late’ in the accented syllable (or in a subsequent unaccented syllable), see [21].

These prosodic categories were established by perception experiments with series of f0 peak shifts generated across the natural German utterance “Sie hat ja ge…gen.” (“She’s been lying.”, -lo- being the only accented syllable in the sentence) [12, 13, 21]. The transition from ‘early’ to ‘medial’ is such that identification in the series switches abruptly as the peak maximum enters the accented vowel, and concomitantly, pair discrimination shows maximal sensitivity across this point. So there is not only category identification in the stimulus series, but also categorical perception in the Haskins sense. However, transition from ‘medial’ to ‘late’ shows gradient perception, albeit change of category. These results are extremely robust and have been found over and over again with different groups of German listeners from different parts of the country. A similar valley shift in relation to the accented vowel onset also produces a switch in identification, but only between ‘early’ and ‘late’, and there is no categorical perception [31].

Niebuhr [29, 30] has refined the phonetic detail of peak category manifestation by adding shape and height of f0 peaks as well as duration and energy of the accented vowel to the parameter of synchronization, and by analysing the interplay between these physical properties in the coding of the associated functional categories. All parameters converge in focussing a waning high-low vs. a waxing low-high pitch and prominence trajectory into the accented vowel for the ‘early’ and
2.2 Methodological prerequisites

There are six prerequisites to a successful elucidation of speech communication:

- We need to start from communicative function [39], and go beyond linguistic functions as they surface in formal categories, such as sentence mode and focus, including categories of expressiveness (‘emphasis’, emotions), of speaker-hearer relations, and of argumentation structure, which is developed by the communicators in ongoing discourse, parallel to, and different from, propositional information structure.
- We need to pay attention to fine phonetic detail in auditory and instrumental investigation as carriers of such functions.
- We need to give equal weight to speaker and listener, and supplement production by perception studies.
- For this new orientation we need a new methodology of data acquisition by dialogue contextualization
  - in spontaneous corpora of various scenarios
  - as well as in systematically stylized experimental discourse frames.
- Any modelling of speech needs to be based on data sets appropriate for the scientific goal. If the model is to account for paralinguistic structures the database must be collected in such a way that it can be considered an adequate representation of paralinguistic phenomena. It will, therefore, no longer do to generate databases from isolated sentences out of context which are constructed according to formal linguistic criteria and are often of doubtful semantics, and which are read in a rote fashion or reproduced after acoustic prompts by speakers selected at random without proper screening as to their aptitude for the task. This is standard practice in studies of peak alignment [35, 2], which either ignore communicative function or handle it rather poorly. In another type of data acquisition, sentences are enacted with different emotions by actors. This begs the question because it presupposes that actors realise the phonetic manifestations of different emotions in such a reading task in the same way as speakers in natural communicative settings.
- Inferential statistics needs to be applied to data sets in accordance with the design of the research question and interpreted with regard to plausibility of productive and perceptual differentiation in speech communication. Tests, such as ANOVAs, should not be used mechanistically nor to establish the discreteness of data sets for formal linguistic categorization.

2.3 Expression of the speaker

2.3.1 Argumentation structure

In German, the categories of ‘early’, ‘medial’, and ‘late f0 peak’ contour synchronizations signal the expression of different kinds of argumentation structure – ‘conclusion’, ‘opening’, ‘unexpectedness’. Contextualization of the ‘early’ peak in the example given in 2.1.2 may be “Once a liar, always a liar. ___”: the speaker signals the end of an argumentation, summarizes, and expresses finality as (s)he knows or sees it. For the ‘medial’ peak it may be “Now I understand. ___”: the speaker signals the beginning of an argumentation and expresses openness as to a new observation or experience. For the ‘late’ peak it may be “Oh. ___”: the speaker again signals the beginning of an argumentation, but expresses contrast to expectation, and personal evaluation of this contrast as surprise.

On the basis of further data observation, the categories of argumentation structure have been expanded by introducing an additional peak function ‘late-medial’ between ‘medial’ and ‘late’. It differs from ‘medial’ by having a more pronounced f0 upward glide and by adding contrast to information selection in an opening argument; and it differs from ‘late’, which has a low plateau
before the rising part of the contour and superimposes an affective component on contrast [21, 30]. Figures 2-5 illustrate the 4 peak positions in the German utterance “Er war mal schlank.” (“He used to be slim.”), for instance in a context of situation where two people are looking at old photos and come across one of an old friend.

(1) In a concluding argument, the speaker describes his/her own apperception as something obvious.

(2) In an opening argument, the speaker indicates that s/he has observed, and become aware of, something.

(3) In an opening argument with an overlay of contrast, the speaker indicates that the observation contradicts the expectation.

(4) In an opening argument with overlays of contrast and expressive evaluation, the speaker shows feeling over an unexpected observation.

The f0 and energy time courses show parallelism.

- For the ‘early’ peak, they both descend into the vowel. This waning profile accentuates the low-falling pitch, which becomes an associate of finality in argumentation.

- For the ‘medial’ peak, both contours rise into the vowel. This waxing profile accentuates the high-rising pitch in the first half of the sonorous syllable, which becomes an associate of openness in argumentation. ‘early’ and ‘medial peaks’ are thus opposites in pitch and prominence courses from the pre-accent to the accent syllable, and this may be the reason why the perceptual change in an f0-shift series is so clear-cut and robust.
- In the ‘late-medial’ peak, high pitch is strengthened in a more extensive low-high movement contrast in the vowel, and by energy staying high longer. This is where a gradient pitch-prominence feature enters and becomes associated with degrees of functional contrast.
- This continues right into the ‘late’ peak, where the waxing f0 and energy profile is not only shifted to late in the vowel but is preceded, concomitantly, by a low level in both parameters thus making the contrast between low and high pitch and prominence in the accented syllable even greater. The intensified low-high contrast can become associated with expressive evaluation of a functional contrast.

The validity of this expanded system of peak contour functions in German has been strengthened in a perception experiment using the semantic differential technique in a frame of contextualization [20]. The communicative categories discussed so far form a system of argumentation structure. They are set by the partners engaging in communication and are not identical with an external information structure of ‘given’ and ‘new’. The photo provides the externally given fact “he was once slim”. The speaker decides on the argumentative weight of this fact and puts it into the frame of “this is what it is” or “this is how I see it”. And accordingly the speaker focusses f0 and energy trajectories differently for pitch – prominence signalling. Thus, argumentation structure needs to be clearly differentiated conceptually from information structure to avoid serious misunderstanding.

Givenness and argumentation may even go against each other. The following examples may illustrate this.

(1) After a long discussion at the beginning of term about finding a suitable alternative time and day for a series of seminars to accommodate all who want to attend, the tutor says “We are going to move the seminars to Thursday.” He says it with an early peak on the last word to indicate to the audience that this is final and the discussion is now closed, although this is ‘new’ information.

(2) The session then continues to discuss other course matters, and at the end, before departing, the tutor reminds the audience “Please remember. We have moved the seminars to Thursday.” He says it with a medial peak to insist, although the information is now ‘given’.

The communicative categories of argumentation structure and their phonetic exponents can be transferred one-to-one to English [11] and are equally found in the English equivalent sentences “She’s been lying.” and “He used to be slim.” In both languages, argumentation structure is grafted onto ‘weighted information selection’ by accentuation, which is achieved by deaccentuating the environment of the highlighted element (‘narrow focus’) and by scaling its salience, in particular through f0 height in peak contours [27, 22].

2.3.2 ‘Negative’ and ‘positive’ intensification
A new method was devised for the eliciting of genuine spontaneous speech. The result is the German LINDENSTRASSE corpus in the VIDEO TASK SCENARIO [34, 24, 10], where two speakers talked about differences between two sets of video clips, which were presented to them separately, each excerpted and spliced together from the German TV series. The speakers were quite familiar with the series and knew each other very well, so they achieved a high degree of spontaneity and naturalness.

This data acquisition design produced a great deal of expressive speech. In the labelling of these data, a problem arose with a sentence accent that was not coded by the pitch patterns of KIM and did not fit into the functional categories associated with them. The communicative function is expressive intensification (‘negative intensification’), and is manifested prosodically by the strengthening of non-sonority, the lengthening of initial consonants, especially voiceless ones, at the expense of the accented vowel, and by non-modal phonation, even voicelessness, in the accented vowel. This accent was termed ‘force accent’ [17].

There were 41 ‘force accents’ in the corpus, which were contrasted with 35 pitch accents in segmentally comparable words and analysed in the signal parameters of duration and energy, as well as by auditory assessment of pitch pattern and voice quality, all pointing to a strengthening of
non-sonority features for an emphatic negative intensification [19]. It can be concluded from these analyses that ‘force accents’ constitute a separate prosodic category with at least three phonetic features in speech production – onset duration, energy, and voice quality.

Reference to the literature [1, 5] and informal data observation prompted the hypothesis that a function of ‘positive intensification’ needed to be distinguished both from ‘negative intensification’ and from ‘weighted information selection’. The functional difference seems to be coded by raised pitch level and sonority in the accented syllable rhyme. To test this hypothesis in German, a new database was required. It had to be obtained in a systematic data acquisition of the three functional categories with control of segmental and prosodic structures in corresponding utterances, and it needed to be as natural as possible. Utterances were designed and arranged in written texts to provide a linguistic and situational context frame that provokes the elicitation of the respective function on a selected key word, helped by pictures of facial expressions. The texts were read by pairs of speakers (one male, one female) in face-to-face communication. The speakers were known to be extrovert, and they knew each other very well.

The data analysis [23] showed a clear difference between the strengthening of non-sonority in ‘negative’ and of sonority in ‘positive intensification’ in relation to ‘weighted information selection’. Figure 6 compares the realizations of the utterance “das stinkt!” (“it stinks!”) in the contexts “Sag mal, hast du in der Klärgrube gebadet? boa! __ zum Kotzen!” (“tell me, did you bathe in the sewers? boa! __ it makes you vomit!”) and “Ich liebe diesen alten Limburger. Wie __ Herrlich!” (“I love this old Limburg cheese. __ Wonderful!”).

**Figure 6:** Spectrum and f0 in German “das stinkt!” with ‘negative’ (left) and ‘positive intensification’ (right) from systematic data collection of ‘emphasis’, female speaker kl.

**Figure 7:** Spectrum and f0 in English “it stinks!” with ‘negative’ (left) and ‘positive intensification’ (right) from data collection of ‘emphasis’, female speaker rb.

These differentiations between ‘negative’ and ‘positive intensification’ can be transferred to English [22] and are found with the same exponents in the corresponding English utterances in Figure 7.
2.3.3 ‘Emphasis’ categories in the Expression function of the Organon Model and their integration into KIM

Through the corpus and the systematic data analysis the following ‘emphasis’ components are added to KIM:

- ‘weighted information selection’, signalled by f0 range
- ‘contrast to one’s expectation’ – degree of affective evaluation of a discrepancy between observed fact and expectation, signalled by ‘medial’ to ‘late f0 peak’ synchronization with the accented syllable
- ‘expressive intensification’, signalled by special prominence for amplifying the verbal meaning
  - ‘positive’ by strengthening sonority of the accented syllable
  - ‘negative’ by weakening sonority of the accented syllable.
- If lexical semantics and prosody go opposite ways in the expression of ‘positive’ or ‘negative intensification’, prosody wins. So, the negative meaning of “stink” is turned into a positive feeling with ‘positive’ prosodic intensification, for instance spoken by a gourmet who loves strong smelly cheeses. Or, in the opposite case, positive lexical semantics is overruled by ‘negative’ prosodic intensification, as in “You did that beautifully.”, to express that it was quite awful. This clash between the two semantic levels leads to irony and sarcasm.

The paralinguistic domain of ‘emphasis’ is thus part of a model of speech communication, not just a linguistic adjunct. To understand speech interaction the paralinguistic use of prosody is essential.

2.3.4 ‘Argumentation structure’ and ‘emphasis’ in other languages

It has already been pointed out in 2.3.1 and 2.3.2 that the same peak synchronizations occur as manifestations of the same argumentation structures in English as in German, and that the semantics and phonetics of positive and negative intensification are also comparable. On the hypothesis that ‘argumentation structure’ and ‘emphasis’ functions are basic in human communication, they will find their expression in languages generally, and thus also in tone languages, such as Mandarin Chinese. The question then is how. There are some indications that ‘finality’ vs. ‘opening argument’ are signalled in Mandarin Chinese by superimposing aspects of a waning or a waxing pitch – prominence pattern, namely pitch height, onto the word tones. Figure 8 compares the word “hao” “OK” (with the low tone) and the word ‘xing’ “OK” (with the rising tone), each spoken as ‘closing’ or ‘opening argument’, respectively. (I am very grateful to Yi Xu, London, for providing the recordings.)

![Figure 8: Spectrograms, f0 (plain) and energy (dotted) traces of Mandarin Chinese “hao” “OK” (above) and “xing” “OK” (below) with ‘closing argument’ (left) and ‘opening argument’ (right); male speaker yx.](image-url)
For 'closing argument', the low tone ends low and concomitantly acoustic energy decreases towards the end, whereas for 'opening argument', they both rise towards the end. Similarly, the high tone ends lower or higher and concomitantly the time course of acoustic energy is on a lower or higher level for ‘closing’ vs. ‘opening argument’. So, the waning or waxing pitch – prominence patterns seem to be at work again, adapted to the conditions set by the tone language.

These Mandarin Chinese data mirror the use of prosody for similarly contextualized yes in English. The following excerpts from the film The Queen with Helen Mirren in the role of Queen Elizabeth II are to illustrate actors’ awareness of these communicative potentials of prosody. The scene at the beginning of the film introduces Queen Elizabeth in full regal robe in a portrait painting session, watching a television report on the final stage of Tony Blair’s 1997 general election campaign on election day, and talking to the portrait artist, played by the Jamaican Earl Cameron:

P. “But it is [late peak] your government.
Q. “Yes. [early peak] Suppose that’s some consolation.

In a telephone conversation between the Queen and Tony Blair, played by Michael Sheen, the following dialogue takes place:
B. “Let’s keep in touch.”
Q. “Yes, [medial peak] let’s.”

Figure 9 compares the two prosodies of “Yes.”; with ‘early peak’ it expresses resigned acceptance of a political fact, with ‘medial peak’ it suggests an initiative for further contacts. Again f0 is lower initially in the former, and the energy trace follows the more gradual descent of f0 but is more sharply peaked for ‘medial’, which is not phrase-final and therefore also much shorter.

There are also indications that ‘negative intensification’ is implemented in basically the same way in Mandarin Chinese as in German or English. Much needed further research should tell.

It has been extensively discussed in the literature that French, which lacks lexical stress, has an ‘accent d’insistance’ that reinforces the statement being made by strengthening the beginning of a word, especially lengthening its initial consonant and raising the acoustic energy and f0 of this syllable. If the word begins with a vowel it may be strengthened by a glottal stop, or the ‘accent d’insistance’ occurs on the next syllable and affects its initial consonant. [8] The discussion seems to conflate two functions of this type of accent, ‘reinforcement’ and ‘intensification’, which most probably differ in fine phonetic detail, particularly as regards the relative weighting of C and V in the accented syllable and the type of voice quality.

Grammont takes the word ‘épouvantable’ to introduce the notion of accent d’insistance, which he describes as follows: “…il peut se faire qu’au lieu de dire cette phrase avec calme on la prononce avec une certaine émotion, que l’on éprouve le besoin de mettre en relief l’appréciation
que l’on formule.”[8, p. 140] This example, like many others he quotes, illustrates the function of ‘negative intensification’, i.e. expressive emphasis, not insisting ‘reinforcement’. But he also lists cases like “Cette conception n’est pas romaine, mais doit être punique.”, which clearly represent a different function, namely insistence and reinforcement. It is an empirical question as to whether the two uses of ‘accent d’insistance’ are carried by different phonetic exponents; I presume they are, in parallel with the German data. Future research will tell.

2.4 Appeal to the listener

2.4.1 Functional and formal definitions of questions

Questions are typical means of appealing to the listener. They fulfil at least three different functions:

- asking for specific factual information
- asking for a decision along a positive-negative polarity scale
- asking for repetition or further specification with an expression of surprise.

For languages like German, English, French and many others, questions have been defined formally with reference to syntax and lexical items. Thus, in German or English, question-word questions and word-order questions are differentiated. According to common textbook opinion, the former are associated with falling, the latter with rising intonation. Among the lexically marked questions, a subclass is defined prosodically by having rising pitch from the question word to the end of the sentence: these are the repeat questions insisting on the factual information focussed on by the question word.

2.4.2 Questions in the Kiel Corpus of Spontaneous Speech

Based on the Appointment-Making Scenario data of the Kiel Corpus of Spontaneous Speech for German [9], which are available in orthographic transliteration and phonetic as well as prosodic transcription, search operations were carried out for the four combinations of lexical/syntactic question type and falling/rising pitch. Table 1 gives the results.

Table 1: Distribution of frequencies of falling (f), high rising (hr), low rising (lr) and other (o) pitch patterns in word order and question-word questions.

<table>
<thead>
<tr>
<th></th>
<th>f</th>
<th>hr</th>
<th>lr</th>
<th>o</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>word-order</td>
<td>21%</td>
<td>39%</td>
<td>30%</td>
<td>10%</td>
<td>100%(121)</td>
</tr>
<tr>
<td>question-word</td>
<td>57%</td>
<td>10%</td>
<td>24%</td>
<td>9%</td>
<td>100%(172)</td>
</tr>
</tbody>
</table>

Falling and rising patterns occur in both interrogative structures. Word-order questions predominantly show rising patterns, question-word questions predominantly falling ones. There is a negligible proportion of high-rising contours in question-word questions, whereas they dominate in word-order questions. Subsequently, instances of each of the four pairings, syntactic/lexical and f/r, were semantically interpreted in their contextual settings; selected examples were resynthesized, changing the pitch pattern from falling to rising or vice versa in each case, and interpreted in the same naturally produced context with regard to contextual compatibility and paralinguistic change of meaning.

One female speaker produced a set of all four combinations of linguistic form and prosody in the Appointment-Making Scenario corpus. Here are two of her examples to illustrate the functions of the prosodic patterns. For further details cf. [18]. With an original high-rising pitch pattern in the word-order question “Würde Ihnen das passen?,” “Would this suit you?” the speaker leaves the decision as to the polarity to the listener. In the resynthesis with a ‘medial peak’ pattern, the speaker expects the answer “yes”. In the resynthesis with a ‘late-medial peak’, the speaker expects the answer “yes” and sets it against a negative answer, thus indicating that a different answer would not be appropriate, giving the utterance a tone of irritation and impatience. With an original ‘early
peak’ pattern in the question-word question “Was würden Sie denn davon halten?” “What would you think of that?” the speaker terminates her turn by asking the dialogue partner to make a final suggestion. This is the ‘closing argument’. In the resynthesis with a high-rising pitch pattern, the speaker requests a comment from the dialogue partner on the date that has been proposed, and the Appeal to the listener takes over.

2.4.3 A function – form framework of questions

The data from the German spontaneous corpus and the systematic f0 manipulation and contextual interpretation show quite clearly (a) that both rising and falling pitch patterns occur with both formal sentence structures, (b) that there is a statistical link between word-order questions and rising pitch as well as between question-word questions and falling pitch, (c) that the opposite pitch pattern introduces different functional aspects in each case.

These results fit into a function – form framework which may be illustrated by the following contextualized question types in German.

(1) Speaker B:
   “Wo?” “Where?”
   in the dialogue context of
Speaker A:
   “Das Treffen findet in Hademarschen statt.”
   “The meeting will take place in Hademarschen.” (a small town in the north, not widely known in Germany).

(1.1) With a ‘medial peak’, B asks for more information about the location of the venue in the town. This introduces the function of ‘opening argument’ of a ‘medial peak’ contour into the question context.

(1.2) With a ‘late-medial peak’, B sets the need for more information about the venue against the insufficiency of the information so far given by A. This introduces the superimposed functions of ‘contrast’ and ‘opening argument’ of a ‘late-medial peak’ into the question context. The utterance has a tone of irritation and impatience.

(1.3) With a high-rising ‘late valley’, where the rise starts in the accented vowel, B still asks for more information about the venue, but appeals to the listener to give it. The utterance sounds less categorical and more friendly than with a peak pattern.

(1.4) With a high-rising ‘early valley’, where the rise starts before the accented vowel, Speaker B appeals to the listener to repeat the name of the place because s/he has not heard it properly or finds it strange.

Figure 10 provides these four pitch patterns.

![Figure 10: Speech waves and f0 traces of 4 pitch categories in the question-word question “Wo?” “Where?”](image)

From left to right: ‘medial peak’, ‘late-medial peak’, high-rising ‘late valley’, high-rising ‘early valley’; male speaker kk (the author).
(2) Speaker A2: “Ist er in Rome?” “Is he in Rome?”
in the dialogue context of
Speaker A1: “Wo ist er denn eigentlich?” “Where does he happen to be?”
Speaker B: “Er ist nach Italien gefahren.” “He has gone to Italy.”

(2.1) With a ‘medial peak’, A wants more information about the person’s whereabouts and suggests a place, expecting the answer to be “yes”. This introduces the function of ‘opening argument’ of a ‘medial peak’ contour into the question context.

(2.2) With a ‘late-medial peak’, A wants more information, as in (2.1), but sees his/her suggestion as being different from what one might have expected. This introduces the superimposed functions of ‘contrast’ and ‘opening argument’ of a ‘late-medial peak’ into the question context.

(2.3) With a high-rising ‘early valley’, where the rise starts before the accented vowel, A does not prejudge the answer but appeals to the listener for a polarity decision.

(2.4) With a high-rising ‘late valley’, where the rise starts in the accented vowel, A still appeals to the listener for a polarity decision, but this time with an expression of surprise at the person perhaps being in Rome.

Figure 11 provides these four pitch patterns. Such function–form relationships need to be investigated in other languages; they may be expected to be comparable in, e.g., English and Dutch.

Figure 11: Speech waves and f0 traces of 4 pitch categories in the word-order question “Ist er in Rom?” “Is he in Rome?” From top to bottom: left – ‘medial peak’, ‘late-medial peak’; right – high-rising ‘early valley’, high-rising ‘late valley’, in each case on “Rom”; male speaker kk (the author).

2.4.4 Questions in a paradigm of speech communication

To lead the discussion of types of questions and their formal manifestations, from syntax to prosody, into a framework of Speech Communication it is essential to approach the phenomena from the functional point of view and to ask what bundles of formal features code various relationships between the SENDER, the RECEIVER, and the FACTUAL WORLD in the transmission of questions. In the case of requesting specific factual information by question-word questions, the Representation function figures prominently in the speaker’s intention. So, a matter-of-fact request message will have a prototypical falling intonation. Contrariwise, a polarity question is more prominently oriented towards the RECEIVER, if the speaker leaves the decision between the polarities entirely to the listener; so, the prototypical realization will be rising intonation for an Appeal to the listener.

However, if the Representation function in a factual information question is supplemented by a consideration for the addressee, requesting rather than matter-of-fact asking for information, this friendliness colouring will result in a rising intonation from the last sentence accent to the end. Or,
if the speaker asks for repetition or further specification of the focussed factual aspect, the Appeal
to the listener comes to the fore, and the intonation rises from the question word to the end of the
sentence. On the other hand, if the speaker prejudges the decision between polarities and is
therefore not primarily listener-oriented, the question loses some of its Appeal character and may
be realised with falling intonation.

In such a speech communication perspective, the prosodic exponency of questions is not
determined by the formal linguistic structures but by the way the speaker constructs his/her
relationship with the listener and with the external world in ongoing communication.

3. THE ETHOLOGICAL BASIS OF EXPRESSION AND APPEAL

In all languages, polarity questions are associated with high pitch, either high register, or rising or
expanding and strengthening the maximum of peak patterns. Why should this be so? High pitch is
commonly perceived as a signal of uncertainty and submissiveness, whereas low pitch reflects self-
confidence and dominance. Actors are chosen for roles according to their voice register. Richard
III and Macbeth, as against Hamlet, should not have a high pitch level. Radio and television news-
readers with low voices are preferred by the traditional channels because they sound more
authoritative and convincing. But free stations as well as call centres cultivate youthfulness and
customer friendliness and employ speakers with high pitch levels. Charles Darwin has already
given an explanation of this phenomenon [6] A dog makes his body big or small to demonstrate
strength or weakness; he growls or whimper with big or small vocal cords.

When we ask a question we request something from a dialogue partner, we submit to the
listener’s discretion. This is the basis for Ohala’s ‘frequency code’ [32, 33] as an ethological
explanation for high pitch in the manifestation of polarity questions in the languages of the world.
It can also explain the divergence from high pitch when the Appeal function to the listener recedes
and the Representation function comes to the fore, in spite of the formal structures used for this
type of question. The principle also explains the frequent use of falling pitch patterns in question-
word questions, where obtaining representational information is central and the listener orientation
moves into the background, but when it becomes the focus high pitch patterns reappear in this type
of question as well.

Even peak contour synchronizations can be related to the ‘frequency code’. Their decisive
differentiating features are syntagmatically contrastive waxing or waning pitch – prominence
patterns synchronized around an accented syllable. All the acoustic properties converge in
establishing such low-high or high-low trajectories for the listener.

Thus, argumentation structures as part of the Expression function, and questions as part of the
Appeal function can be seen as being rooted in the ethological basis of human behaviour, and
captured by the ‘frequency code’.

As for the emphasis categories of the Expression function, ‘negative emphasis’ may be assumed
to be a universal function expressing anything from dislike to disgust, and as such may be
considered an adaptation of the biological function of vomiting [37] to human speech
communication. Like its vegetative root, this communicative stylization would be characterized by
vocal tract, more particularly pharyngeal, narrowing, and raising of the larynx. The phonetic
properties of heightening non-sonority by lengthening initial voiceless consonants, by fricativizing
nuclear vowels, and by using non-modal phonation would result naturally from such a production
basis. ‘Positive emphasis’ has the opposite articulation basis, i.e. vocal tract widening, with the
observed phonetic properties following from it, and, of course, the differences go together with
different facial expressions.
4. OUTLOOK

What we need now is a comprehensive investigation into the function – form – substance relationships of the Expression and Appeal functions across a variety of structurally different languages in the world. This is a large-scale project, encompassing different accent, tone and rhythm types. It presupposes a different attitude towards paralinguistics and its relationship to linguistics. When they are treated as integral parts in a paradigm of Speech Communication our understanding of how human interaction by language works and why it operates the way it does will be greatly advanced.

5. REFERENCES


