Prosody in Speech Interaction
Expression of the Speaker and Appeal to the Listener

Klaus J. Kohler
Institute of Phonetics and Digital Speech Processing

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. Paralinguistics thus moves from the periphery of linguistics to its center. In particular, prosodic manifestations are discussed of argumentation structure, of categories of emphasis, and of attitudes towards the listener in questions.

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 [1], and has even seen a renaissance in the field of prosody [2,3,4,5,6]. 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. A proper definition of paralinguistics has never been given. The boundaries between the two fields as they are set up by linguists are opaque. The expression of sender-recipient relationships, respect, for example, is to a large extent signalled by prosodic means in many European languages, and would be regarded as paralinguistic. In Japanese, on the other hand, the expression of social relationships between speakers is formalised at the morphological level, and is therefore classified as linguistic [7].

Similarly, the categories of ‘early’, ‘medial’, and ‘late f0 peak’ contour synchronizations for the expression of different kinds of argumentation structure in languages like German and English (cf. 2.1.2, 2.3.1) are considered pragmatic rather than categories of formal semantics, and ‘medial’ vs. ‘late’ seems to be manifested by gradience rather than discreteness. But they are undoubtedly categories of speech communication and therefore important in speech research.

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 [8, 9]. 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’ [10], 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 dialects, 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 [11, 12, 13, 14, 15, 16]. 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 accent-vowel onset (‘early’), after the accented-vowel onset (‘medial’) or ‘late’ in the accented syllable (or in a subsequent unaccented syllable), as illustrated in Figure 2.

![Figure 2: The ‘early’, ‘medial’, and ‘late’ f0 peak contours in the German sentence “Sie hat ja gelogen.” (“She’s been lying.”), synchronized with the onset (a), middle (b), or end (c) of the vowel in the accented syllable -lo-, as represented in speech wave and spectrogram.](image)

These prosodic categories were established by perception experiments with series of f0 peak shifts generated across the natural German utterance “Sie hat ja gelogen.” (“She’s been lying.”), -lo- being the only accented syllable in the sentence) [11, 12, 16]. 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 be-
tween ‘early’ and ‘late’, and there is no categorical perception [17].

Niebuhr [18, 19] 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 ‘medial’ categories, respectively. For the ‘late’ category, the waxing pitch and prominence trajectory is shifted to the end of the accented syllable and beyond into an unaccented one if there is one.

2.2. Methodological prerequisites

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

- We need to start from communicative function [20], going beyond linguistic functions as they surface in formal categories, such as sentence mode and focus, but 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 [5, 21], 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 need 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. Procedures, 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 argument structure – ‘conclusion’, ‘opening’, ‘unexpectedness’.

- The ‘early’ peak in Fig. 2 can be contextualized by “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.

- The ‘medial’ peak in Fig. 2 can be contextualized by “Now I understand. ___.”. The speaker signals the beginning of an argumentation and expresses openness as to a new observation or experience.

- The ‘late’ peak in Fig. 2 can be contextualized by “Oh___.”. The speaker again signals the beginning of an argumentation, but expresses contrast to expectation and personal evaluation of this contrast as surprise.

These communicative categories of argumentation structure and their phonetic exponents can be transferred one-to-one to English [22] and are equally found in the English equivalent “She’s been lying.” In both languages, argumentation structure is grafted onto 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 [23, 24]: ‘weighted information selection’, cf. 2.3.2 and 2.3.3.

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 [16, 19]. Figures 3-6 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 look at old photos and come across one of a good friend.

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

Er war mal [l a ɲ k h]

Figure 3: Spectrum, f0, energy in German “Er war mal schlank.” “He used to be slim.” – ‘early peak’; male speaker kk (the author).
(2) In an opening argument, the speaker indicates that s/he has observed, and become aware of, something.

Er war mal schlank.

Figure 4: Spectrum, f0, energy in German “Er war mal schlank.” “He used to be slim.” – ‘medial peak’; sp. kk.

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

Er war mal schlank.

Figure 5: Spectrum, f0, energy in German “Er war mal schlank.” “He used to be slim.” – ‘late-medial peak’ sp. kk.

(4) In an opening argument with overlays of contrast and expressive evaluation, the speaker shows his/her feelings about the unexpected observation.

Er war mal schlank.

Figure 6: Spectrum, f0, energy in German “Er war mal schlank.” “He used to be slim.” – ‘late peak’; sp. kk.

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 argumentative and expressive functions as well as the prosodic exponencies of the 4 peak patterns can also be transferred to English and are found in the English equivalent sentence “He used to be slim.” 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 [25].

The communicative categories that have been discussed 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’.

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 [26, 27, 28], 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 the sentence accent that was not coded by the pitch patterns of Kt 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 phona tion, even voicelessness, in the accented vowel. This accent was termed 'force accent' [29].

Figure 7 provides an example to illustrate the phonetic exponents of 'negative intensification'.

- There are pitch accents on “Valerie” and “Treppe”, the arrows mark the two falling peak contours.
- “runterkickt” is integrated into the second fall and does not have a separate pitch accent in this object+adverbial+verb construction, only a microprosodic f0 increase.
- A ‘force accent’ intensifies the negative verb meaning.
- The initial voiceless plosive and its aspiration are considerably lengthened, the vowel is very short, absorbed in the surrounding non-sonority.

Figure 7: Speech wave, spectrogram, f0 in German "(Wie Boris) Valerie die Treppe runterkickt." "(When Boris) kicks Valerie down the stairs.", 'negative intensification' on kickt, from LINDENSTRASSE corpus, male speaker mpi; plain lines=word, dotted=vowel, arrows= peak contours.

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 [30]. As initial consonant lengthening is also characteristic of speech dysfluency, the question arose as to whether and how emphasis and dysfluency are differentiated by the listener. So the analysis of production was supplemented by a perception experiment with systematically manipulated stimuli from the corpus. 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, and that they are equally relevant in perception, albeit only duration was formally tested, the relevance of the other two being deduced from the results. Further investigation of new data thus became necessary to complete the picture.

Reference to the literature [31, 32] 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 for ‘emphasis’ 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 [33] 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 8 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 8: Spectrum and f0 in German “das stinkt!” with ‘negative’ (above) and ‘positive intensification’ (below) from systematic data collection of ‘emphasis’, female speaker kl.

These differentiations between ‘negative’ and ‘positive intensification’ can again be transferred to English [24] and are
found with the same exponents in the corresponding English utterances in Figure 9.

Besides ‘positive’ and ‘negative intensification’, another type of ‘emphasis’ has to be distinguished with regard to function as well as phonetic exponency. Its function is ‘reinforcement’ of a statement or a point of view. It is more than rational ‘weighting of information selection’, it is insisting selection and thus comes in the domain of the Expression function, just like ‘positive’ and ‘negative intensification’. But the latter express feelings of (un)pleasantness. As to the phonetic manifestation, it shows long initial voiced or voiceless consonants in relation to the accented vowel, whereas in ‘positive intensification’ the accented vowel is long in relation to the initial consonant. Furthermore, f0 and energy time courses create a pitch-prominence profile with a plateau configuration for ‘positive intensification’, whereas for ‘reinforcement’ it is a peak configuration. In spite of accented-syllable initial consonant strengthening, the exponency of ‘reinforcement’ differs from ‘negative intensification’ by modal voice quality and pitch pattern.

Figure 10 provides an example of each in the German utterance “Es hat sich enorm verbessert.” (“It has improved enormously.”) from a systematic data collection of ‘emphasis’: ‘positive intensification’ (above, male speaker sk), ‘reinforcement’ (below, female speaker al); plain lines = word “enorm” in IPA transcription, broken line = initial consonant and vowel.

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
- ‘insisting information selection’ by reinforcement of the speaker’s point of view, and by correcting or contradicting discourse partners, signalled by initial consonant strengthening in addition to 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, the opposite case of positive lexical semantics being 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 and dialects

Since the categories of ‘argumentation structure’ and ‘emphasis’ are defined functionally and since these functions represent basic interactions in human behaviour they may be assumed to apply to any language and should therefore be part of a general model of human speech communication. The question then is as to how they are implemented at the array of linguistic levels of phonetics, morphology, syntax, and the lexicon in individual languages and dialects. Here are a few more data that show convergences and differences.

‘Argumentation structure’ and ‘emphasis’ in SW-German dialects.

South Rhine-Frankish, Swabian and Alemannic dialects, including Swiss German, exhibit a prosodic peculiarity among German dialects: the function associated with the ‘medial peak’ in Standard German (which has been the basis for Kim) and in other German dialects is manifested by a rising-falling peak contour that is synchronized late with the accented vowel. Figure 11 provides an example from the Freiburg corpus of the German Research Council project Intonation of Regional Varieties of German carried out by Peter Auer, Margret Selting, Peter Gilles, Jörg Peters [34, 35, 36].

Figure 11: Spectrogram, f0, energy in the Freiburg version of “Da ist ja der Erste am Sonntag.” (“That’s when the first is on Sunday.”). Rising-falling contour encased, maximum f0 at beginning of unaccented vowel marked by dashed line; dotted lines raised f0 peak contour. Female speaker fr01a.

The phonetic exponents of ‘opening argument’ in this dialect are

- late synchronization of the f0 peak maximum
- a slow f0 fall following the maximum, spread out over unaccented syllables, forming a plateau before the last one if there are several
- none of these syllables following the f0 maximum perceived as stressed
- energy trace in parallel to the f0 descent

This exponency is different from the ‘late peak’ in the contrastive and expressive function of Standard German, as regards the timing of the f0 descent and the energy decline across the unaccented syllables. To signal degrees of contrast and expressive evaluation the f0 maximum of the peak contour is raised in a scalar fashion. The exponents of ‘closing argument’, on the other hand, are the same as in Standard German, illustrated in Figure 12, where the same speaker as in Figure 11 talks about people’s allegiance to their dialects in other parts of Germany, and then summarizes by saying “Why shouldn’t we show allegiance to the alemannic dialect.”.

Figure 12: Spectrogram, f0, energy in Freiburg “(Warum solle mir nit dazu steh,) zum Alemannische.”. “(Why shouldn’t we show allegiance to the dialect,) to Alemannic.”. ‘Closing argument’ with ‘early peak’ synchronized at the CV boundary of the accented syllable “-mann-”. F0 maximum marked by plain line. Female speaker fr01a.

Although the phonetic detail for signalling ‘closing’ and ‘opening argument’ in Southwest German, more specifically the Freiburg Alemannic dialect, are different from Standard German there is convergence in a more basic way: ‘finality’ is coupled with a waning pitch-prominence pattern into the accented vowel, ‘opening’ with a waxing pitch-prominence pattern into the accented vowel or out of it, and contrast and expressiveness are superimposed by increasing the pitch-prominence range and maximum of this waxing pattern.

So far I have not come across any examples of ‘positive’ and ‘negative intensification’ in the analysis of the Freiburg corpus, but my familiarity with the dialect tells me that the exponents are exactly comparable to the ones found in Standard German and English. What has come to light, however, are examples of ‘reinforcement’. Figure 13 illustrates it with an example from a wine grower, who praises the dryness of his wine as against the sweetish stuff that is offered in inns. He first states, in an ‘opening argument’, “This one’s dry.” and gives the word “trocke” (“dry”) the late synchronization of a waxing pitch-energy pattern. He then reinforces the statement by saying “As dry as a fart, it is.” This ‘reinforcement’ is to drive the argument home, but now he has a medial peak synchronization with a waxing pitch-prominence pattern into the vowel, a much higher f0 maximum, and waning pitch-prominence following. The dynamics of this waxing-waning pitch-prominence is similar (apart from the f0 height) to the ‘medial peak’ in Standard German, but its function is different.
The ‘reinforcement’ is further heightened by syntactic fronting of the adjective, which has a semantic intensifier. Without ‘reinforcement’ the word would have double stress, on the stem and on the intensifier, each realised with late synchronisation in the Freiburg dialect. But a single, initial accent is a third aspect of the ‘reinforcement’ function. ‘Reinforcement’ can thus be carried by a bundle of fronting features: fronting in the syntactic construction, fronting in the accentuation of a compound, and fronting in the strengthening of syllable initials, with the latter already being sufficient to signal the function, the other features heightening it.

The following excerpts from the film The Queen with Helen Mirren in the role of Queen Elizabeth II 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. She regrets that she has not a vote; then the following dialogue develops:

Q. “The sheer joy of being partial.”
P. “Yes. [late peak] Of course one forgets that as sovereign, you are not entitled [late peak] to vote.”
N. “No.” [early peak]

In the same scene, there is also the following interchange:

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

In two later telephone conversations between the Queen and Tony Blair, the following dialogues take place:

B. “Let’s keep in touch.”
Q. “Yes. [medial peak] let’s [medial peak].”
B. “Is it your intention to make some kind of appearance or statement?”
Q. “No. [late peak] No. [late peak] Certainly not.”

‘Argumentation structure’ and ‘emphasis’ in English

The following excerpts from the film The Queen with Helen Mirren in the role of Queen Elizabeth II 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. She regrets that she has not a vote; then the following dialogue develops:

Q. “The sheer joy of being partial.”
P. “Yes. [late peak] Of course one forgets that as sovereign, you are not entitled [late peak] to vote.”
N. “No.” [early peak]

In the same scene, there is also the following interchange:

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

In two later telephone conversations between the Queen and Tony Blair, the following dialogues take place:

B. “Let’s keep in touch.”
Q. “Yes. [medial peak] let’s [medial peak].”
B. “Is it your intention to make some kind of appearance or statement?”
Q. “No. [late peak] No. [late peak] Certainly not.”

‘Argumentation structure’ and ‘emphasis’ in English

The following excerpts from the film The Queen with Helen Mirren in the role of Queen Elizabeth II 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. She regrets that she has not a vote; then the following dialogue develops:

Q. “The sheer joy of being partial.”
P. “Yes. [late peak] Of course one forgets that as sovereign, you are not entitled [late peak] to vote.”

In the same scene, there is also the following interchange:

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

In two later telephone conversations between the Queen and Tony Blair, the following dialogues take place:

B. “Let’s keep in touch.”
Q. “Yes. [medial peak] let’s [medial peak].”
B. “Is it your intention to make some kind of appearance or statement?”
Q. “No. [late peak] No. [late peak] Certainly not.”

‘Argumentation structure’ and ‘emphasis’ in English

The following excerpts from the film The Queen with Helen Mirren in the role of Queen Elizabeth II 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. She regrets that she has not a vote; then the following dialogue develops:

Q. “The sheer joy of being partial.”
P. “Yes. [late peak] Of course one forgets that as sovereign, you are not entitled [late peak] to vote.”

In the same scene, there is also the following interchange:

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

In two later telephone conversations between the Queen and Tony Blair, the following dialogues take place:

B. “Let’s keep in touch.”
Q. “Yes. [medial peak] let’s [medial peak].”
B. “Is it your intention to make some kind of appearance or statement?”
Q. “No. [late peak] No. [late peak] Certainly not.”
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.

An example of ‘negative intensification’ is given in Figure 16. As in the German example of Figure 7, the negative meaning of “tordu” “crazy” is intensified by extreme lengthening of non-sonority in the initial voiceless consonant of the syllable tor; at the same time the preceding unaccented vowel is considerably shortened. The main difference between the examples from the two languages is that the French one is embedded in rising pitch contours, the German one in peak patterns.

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.”[37, 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 intensity, as in the German examples of ‘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.

‘Argumentation structure’ and ‘emphasis’ in tone languages.

On the hypothesis that ‘argumentation structure’ and ‘emphasis’ functions are basic in human communication, they will also find their expression in tone languages, such as Mandarin Chinese. The question then is how. There are some indications that ‘finality’ vs. ‘opening argument’ are signalled by superimposing aspects of a waning or a waxing pitch-prominence pattern, namely pitch height, onto the word tones. Figures 17 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.)

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.

There are also indications that ‘negative intensification’ is implemented in basically the same way as in German or English. Much needed further research should 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 fulfill 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 [38], 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/hr, 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. Two examples will be illustrated here, one from each formal question type; for further details cf. [39].

Figure 18 provides an example of 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.

Figure 19 gives an example of an original ‘early peak’ pattern in the question-word question “Was würden Sie denn davon halten?” “What would think of that?” After a general discussion about a possible date for an appointment 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.

The same speaker produced examples of all four combinations of linguistic form and prosody in the Appointment-Making Scenario.

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 can be put in a function – form framework with reference to 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; “Auchtermuchty” in Scotland would be comparable).

(1.1) With a ‘medial peak’ B asks for more information about the location of the venue in the town. This introduces the functional meaning of ‘opening argument’, associated with 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 superimposes the functional meaning of ‘contrast’ on the ‘opening argument’, associated with a ‘late-medial peak’ in German, and introduces it 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 rather strange.

Figure 20 provides these four pitch patterns.

(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, indicating that he expects the answer to be “yes”. This again introduces the functional meaning of ‘opening argument’, associated with a ‘medial peak’ contour in German, into the question context.

(2.2) With a ‘late-medial peak’ A wants more information, as in (2.1), but sees his suggestion as being different from what one might have expected. This superimposes the functional meaning of ‘contrast’ on the ‘opening argument’, associated with a ‘late-medial peak’ in German, and introduces it 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 21 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.

2.4.4. Questions in a paradigm of speech communication

To get 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 focused 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 pre-judges 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 the Expression and Appeal functions in Speech Communication

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 [40] A dog makes his body big or small to demonstrate strength or weakness; he growls or whimpers 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’ [41, 42] 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.’frequency code’ 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 [43] 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. 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