ON THE INTERDEPENDENCE OF SOUNDS AND PROSODIES IN COMMUNICATIVE FUNCTIONS

Klaus J. Kohler

University of Kiel, Germany
kjk@ipds.uni-kiel.de

ABSTRACT

Sound segments have traditionally occupied a central place in phonetic science. Other sound aspects have been conceptualized in a broader, suprasegmental frame as prosodies, especially pitch, but also energy, voice quality, rhythm. This has resulted in the current dichotomous research paradigm of sounds and prosodies. This paper takes a new look at this division, as a useful initial heuristics separating threads that subsequently have to be interwoven in a tapestry of speech communication, which incorporates all forms of meaning, propositional, attitudinal and expressive. The central question to which this paper proposes a few answers is: "How are sounds and prosodies intertwined, mutually shaping each other, as a reflection of different communicative functions in speech interaction?" The paper proceeds from communicative function to phonetic form and presents English and German data from production as well as perception analysis in all three fields of meaning.

Keywords: sounds and prosodies, communicative functions

1. THE DICHOTOMY OF SOUNDS AND PROSODIES

The concept of *sound segments* has traditionally played a central role in the phonetic representation of words. It underlies the development of alphabetic writing systems, of phonetic transcription and of phonemic theory. Other sound aspects, especially pitch, but also energy, voice quality, rhythm have been conceptualized as being superpositioned on segments in a broader frame of syllables and utterances. These *suprasegmentals* or *prosodies* may be linked to the level of lexical phonology, too, as lexical tone, lexical voice register or lexical stress, adding to the phonetic differentiation of lexical units, but their main domain of operation is above the lexical level.

The segment is associated with the short-time window of opening closing gestures of the vocal tract, and simultaneously, with the differentiation of propositional lexical meaning, whereas prosodies are generally associated with long-time windows of pitch, energy and voice-quality control, and predominantly with attitudinal and expressive utterance meaning. This different substance – meaning duality in sound segments and prosodies accounts for their study having developed in two separate fields largely independently of each other, resulting in the current dichotomous research paradigm of sounds and prosodies. The link of the segment to the more elementary units of opening-closing gestures in articulation and of words in linguistic structure can also explain the preferential status accorded to segments: nobody, perhaps with the exception of Port [19], has turned the relationship around and coined the concept of infraprosodies. The sound – prosody dichotomy has, however, been repeatedly called into question, particularly in the study of long articulatory components of, e.g., palatalization, velarization, nasalization in the linguistic function of distinguishing words and morphologic-al constructions [4], and has always been bridged in the analysis of lexical stress, where segmental aspects of vowel duration and vowel spectrum, and prosodic aspects of fundamental frequency and energy have jointly been taken into account [9].

But by and large, the two fields continue dealing with differently defined formal properties separately in their own right. This reliance on linguistic form and phonetic substance in the analysis of sound segments and prosodies also reflects the tenets of 20th century structural linguistics, as it relegates the functional aspect of speech communication to a post hoc level. Such a dichotomous formal approach is a useful heuristics to come to grips with the enormous complexity of speech, especially in the initial stages in the investigation of a language. Yet, the formal manifestations in the guise of sounds and prosodies jointly result from functions in speech communication. If these functions are taken as the superordinate control variable, the axiomatic

formal dichotomy of sounds and prosodies fades away because they interact, with varying weights, in the coding of specific communicative functions. This paper takes a new look at the interweaving of the heuristic threads of sounds and prosodies in a tapestry of speech communication in languages, whch incorporates all forms of meaning, propositional, attitudinal and expressive. The central question to which this paper proposes a few answers is: "How are sounds and prosodies intertwined, mutually shaping each other, as a reflection of different communicative functions in speech interaction?" So, contrary to linguistic practice, the paper proceeds from communicative function to phonetic form, presenting English and German data from product-ion as well as perception analysis in all three fields of meaning. It starts with a discussion of the segment-prosody interplay along a scale from high to low-key expressive and attitudinal meaning, leading to the exponents of different types of emphasis. It then moves on to the interrelation of segments and prosodies in question and statement perception, and finally looks at function words in propositional meaning transmission, examining how segmental sound units are reduced to distinctive long articulatory components, *articulatory prosdies*, which spread across syllables and words.

2. HIGH-KEY AND LOW-KEY IN FORM AND FUNCTION

Hawkins [8] has shown conclusively that the interplay of prosodies with fine segmental detail of vocal-tract dynamics can only be disregarded at the expense of an insightful explication of communicative functions in speech interaction. For example, in the realization of the English utterance "*I do not know*", a wide spectrum of rhythm, intonation, loudness, voice quality and articulation features converge to convey specific meanings beyond the propositional meaning of lack of knowledge. The weakly contracted form [aɪ dəʊn̰ ˈnəʊ], with a medial f0 peak contour [11] as well as an energy peak in the only accented syllable "*know*", provides the listener with new, neutral information that the speaker cannot give the requested answer. More strongly elaborated or reduced forms add expressive and attitudinal connotations.

Prosody and articulation can be elaborated together in several degrees to insist on 'not knowing'. The number of accents and the f0 range are increased, and the articulatory movements are expanded in time and space (cf. also Lindblom's H & H theory [14]). In a first step, "*don't*" is given a second accent, with concomitant raising of the f0 peak in "*know*", and there is considerable lengthening of the complete vocal-tract opening-closing movement in "*don't*", as well as a plosive realization at its end: [aɪ ˈdəʊnt(ʰ) ˈnəʊ]. An overarticulated expansion [aɪ ˈduː ˈnɒtʰ ˈnəʊ] of the function words "*do not*", entailing additional accents on both, emphatically reinforces the truth value of the statement of 'not knowing'. This may be heightened still further by giving "*I*" yet another accent and by inserting pauses between the words, and even by accompanying each accent with a downward movement of the arm and pointing index finger, and of the hand banging a table or a lectern, depending on cultural and social codes. A famous example of this reinforcement of the truth value of a statement is former US President Bill Clinton's emphatic denial "*I did "not [pause] "have [pause] "sexual [pause] re"lations with that 'woman, [long pause] Miss Le'winsky*." in the televised public announcement to the American People on 26 January 1998, where these underscorings of accents can be heard and seen *http://www.youtube.com/watch?v=KiIP_KDQmXs*

But in the elaboration of the statement "*I do not know*", the speaker may also introduce an expressive rejection of a listener's repeated questioning. The utterance is then no longer a reinforcement of propositional components of a statement but expressive intensification of a negative contrastive attitude towards the listener. This is signalled by late peak contours [11] on the accented syllables, by tense breathy, instead of modal, phonation throughout the utterance, and by consid-erable lengthening of all vocal-tract opening-closing movements, thus expressing degrees of exasperation. Such a negatively intensified utterance is, again dependent on socio-cultural conventions, more likely to be accompanied by different body movements from the ones used in propositional reinforcement, viz. contraction in furrowed brows and clenched fists.

Moving in the other direction on an elaboration-reduction scale, the opening-closing movements of the vocal tract may be levelled, f0 narrowed and synchronized differently with articulation. This leads to [(aɪ) də ˈnəʊ], which is casual and informal and conveys to the listener that the speaker does not care about not being able to provide an answer. If the accent is marked by an early peak contour [11] the speaker signals finality in argumentation

structure – the closure of a communicative turn: there is no more to be said. If used in a formal situation this may be received as insolence.

Reduction may still go further to the maximally contracted form [ə̤ə̤ə̤], with very weak segmental articulation. It may be uttered in a relaxed communicative situation between family members. For example, A asks B, who is busy reading a book and does not want to be disturbed, where the newspaper is, and B's response is made as a sideline to her main activity at the time. Its function is to signal to A in an aside way not to expect help in looking for the newspaper because she is otherwise occupied. This communicative function determines the phonetic output. The three opening-closing gestures of the weakly reduced form "*I don't know*" of the neutral message remain in a rudimentary fashion as a movement from a more open and fronted through a central to a more closed and retracted vocoid shape of the vocal tract, reflecting the progression of the open phases in the fuller gestures, with superimposed nasality representing the negation. This progression could neither be reversed, nor could nasality be embodied in a nasal contoid, e.g. [n̥n̥n̥], and it must not be absent either. The rhythmic timing also reflects the fuller form, with the central section being shortest, the final one longest. A possible intonation pattern is falling on the first section and low rising on the last, conveying a friendly rebuttal "*Don't ask me.*" in the situational setting described above. The open-close progression of the vocoid quality, the nasalization through-out and the rhythmic timing of the vocoid sections together with the falling-rising f0 pattern form the phonetic essence [17] of this utterance in its situational speech function. If the f0 contour is resynthesized on an otherwise constant schwa hum it becomes noise and is no longer decodable as reduced speech.

The German equivalent of English "*I do not know*", in its various formal manifestations reflecting the semantic and pragmatic functions, is "*keine Ahnung*" [12]. Neutral [kʰa̤ënë ˈã̤ːnõŋ] compares with reinforced, or expressively intensified tense breathy, [ˈkʰaĩːnë ˈʔã̤ːnõŋ], and with maximally reduced [hä̤ë̤ ˈãɑ̃õ]. The neutral version has four clearly demarcated opening-closing vocal-tract movements, with glottalization marking the opening phase in the third, and an early peak pattern is associated with the prominent third gesture for the function of argumentative finality. A medial peak is also possible in the neutral version to signal argumentative openness. In the reinforced version, articulation and prosody are intensified, a glottal closure marks the juncture between the closing of the second and the opening of the third gesture, and both "*kei(ne)"* and "*Ah(nung)*" are accented. If, in addition, tense breathiness takes the place of modal phonation throughout the utterance, and late instead of medial peak contours occur with the accents, the effect of contrastive negative intensification, expressing exasperation, is introduced.

In the reduced version, the clear demarcation disappears, but the progression of closing-opening-closing front-back vocoid shaping of the vocal tract remains and nasalization marks the final two gestures instead of a nasal contoid separator, while the initial opening movement retains breathy phonation. The timing of the gestures is also kept, and an early peak pattern is again associated with the prominent third one to fit the conclusiveness of the communicative situation. The vocoid modulation, its rhythmic timing, the superimposed nasalization, together with the early peak synchronization, and the intial breathiness form the phonetic essence of this reduced German utterance: reduction does not go any further if it is to stay speech.

The discussion of form and function of corresponding English and German sets of utterances containing the same strings of words suggests that a speech-functional scale can be set up that extends from a neutral propositional pivot in two directions, to high-key emphasis with segmental and prosodic expansion, on one side, and low-key playing-down with segmental and prosodic levelling, on the other. Both high-key strengthening and low-key weakening introduce the speaker's expressiveness and attitudes towards the listener and the communicative situation. High-key emphasis may be reinforcement of the truth value of propositional meaning to drive home an argumentation point to the listener, or negative or positive intensification of expressive or attitudinal meaning (cf. section 3). Low-key playing-down may be a parenthetical aside in an argumentation, or it may signal indifference to the listener and disinterest in what is going on in the communicative situation. Different locations on this function-al scale activate distinctive bundlings of segmental and prosodic properties in speakers' productions of utterances in their language. Listeners decode the received phonetically rich signals by relating them to speech functions along this scale, thus understanding the propositional, attitudinal and

expressive aspects of meaning conveyed by speakers in specific communicative situations.

These principles of high-key and low-key in form and function can be assumed to be universal in human communication by speech in the languages of the world. But the formal manifestations of high and low-key functions vary greatly across languages and offer an exciting field of phonetic investigation for our scientific community. This new type of study will uncover interesting typological groupings that go well beyond traditional segmental or prosodic typologies. The comparison of English and German form – function data in this paper already shows that speech communication in the two languages proceeds along very similar strategies, and the other West-Germanic languages Dutch, Low German and Frisian no doubt follow the same patterns. Among the Romance languages, French may be expected to exhibit more different relations between form and function in high and low key, and tone languages will be even more distant.

3. SOUNDS AND PROSODIES IN EMPHASIS

3.1. In statements

In the English and German high-key illustrations, two types of emphasis were distinguished: (1) reinforcement of the truth value of propositional meaning, (2) negative intensification of expressive or attitudinal meaning. Neither of them is congruent with emphasis for contrast [1], or contrastive focus in recent studies [2], which refer to propositional meaning in information structure. Niebuhr [16] presented a detailed analysis of acoustic parameters of expressive intensification in German statements with falling pitch contours, collected in situational contexts that trigger either positive or negative reactions in the speaker. He termed these types of emphasis 'positive' and 'negative intensification', by the side of 'reinforcement', which he proposed as a third type. Reinforcement corresponds to (1), negative intensification to (2).

Negative intensification strengthens non-sonorous, positive intensification sonorous aspects of accented syllables. In detail, this means that, in the former, the vowel is not overly long, f0 forms a pointed peak with a steep fall that starts quite early in the vowel, phonation is tense breathy, and the initial consonant is considerably lengthened, which adds to the asonority when the consonant is voiceless. Positive intensification has the opposite characteristics: the vowel is substantially lengthened, the f0 rise levels out into a plateau followed by a shallow fall that starts around the end of the vowel, phonation is slightly breathy-voice, and the initial consonant is not lengthened. The body language accompanying positive intensification also differs from that used in negative intensification: dilation rather than contraction – raised eye brows and upward movement of arms and hands.

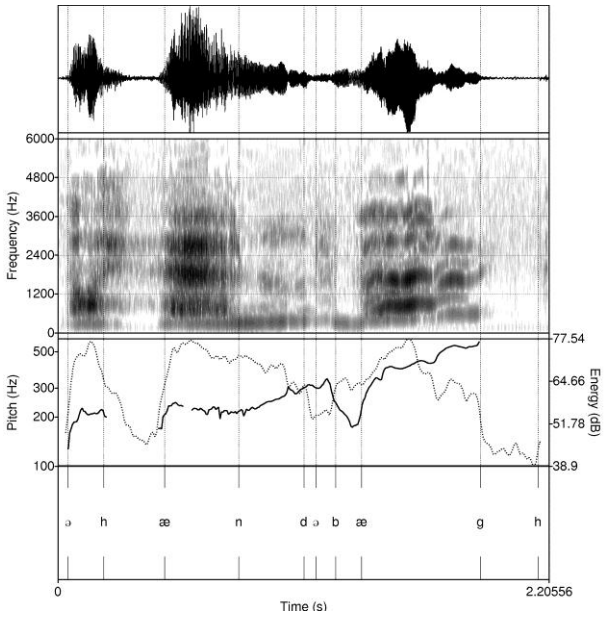
The manifestation of reinforcement combines characteristics of both intensifications. An extended duration of the initial consonant and a pointed f0 peak with a quickly falling slope resemble the pattern of negative intensification. However, the fall of the peak contour sets in later and in a longer vowel than for negative intensification (yet not as late, nor in as long a vowel, as for positive intensification), and phonation is not tense breathy. Reinforcement does not elaborate asonorous aspects in the accented syllable but expands the onset of the opening-closing vocal-tract movement, in unison with a pointed, quickly falling f0 contour. Such syllable-initial articulatory strengthening, which has also been described as a characteristic feature of the French *accent d'insistance* [6], attracts the listener's attention to an upcoming important speech event, even more strongly when it is also preceded by a pause. It is therefore well suited to highlight the truth value of a unit of propositional meaning. At the same time, reinforcement lacks the asonorous phonation or the f0 plateau expansion of negative or positive intensification and thus does not signal the speaker's expressiveness towards the listener.

3.2. In questions

High-key intensification also occurs in questions with high-rising f0 contours to express surprise with a negative or positive attitude towards the addressee. A famous example of negative intensification of surprise from the English stage is the line *"A handbag?"* produced like the blast of a foghorn by actress Edith Evans in productions of Oscar Wilde's *The Importance of Being Earnest.* In the role of Lady Bracknell, a socially domineering and haughty aristrocatic lady, she questions Mr Worthing, her daughter's suitor, about his credentials, and when he has to admit that he was found in *"a handbag"*, she repeats the two words as a surprise question, expressing indignation and utmost incredulity and horror at

this revelation of socially unacceptable origins. In Figure 1, from a performance together with John Gielgud as Mr Worthing, the acoustic analysis shows the following characteristics (audio at *http://www.youtube.com/watch?v=tiNVy5nfbcQ*):

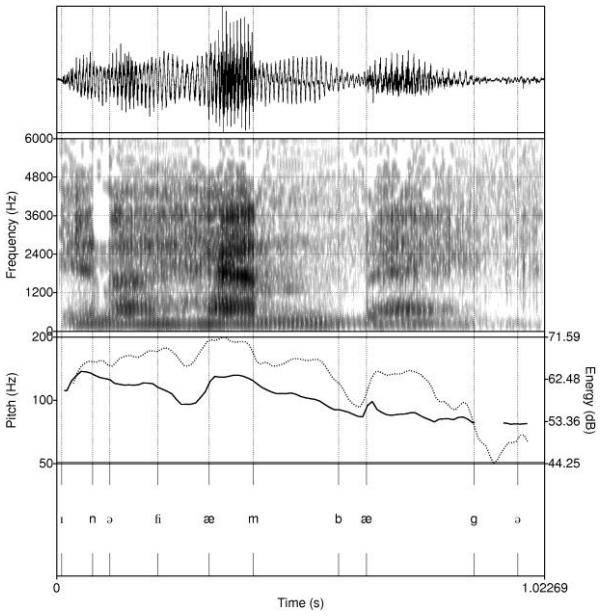
Figure 1: Speech wave, spectrogram, time courses of f0 (plain) and energy (dotted), and phonetic transcription of "*A handbag?"*, spoken by Edith Evans.



- Both opening-closing movements of "*handbag*" are substantially lengthened to about 1 sec and 0.7 sec, the initial [h] to about 250 ms.
- The accented vowel has irregular phonation in amplitude and periodicity, preventing a smooth f0 calculation. It is of the tense breathy type. This asonority fades away with rising f0.
- [d] has a vocoid release of 80 ms before [b].
- From the beginning of the nasal to the end of the word, f0 rises from a low of 215 Hz to a high of almost 600 Hz. Pitch forms a concave high-rising valley contour, whose lowest point is synchronized late in the accented vowel.
- The energy maxima in the three syllables are 76 dB, 77 dB, 77 dB, pointing to an overall high loudness and to strong equal accentuat ion on each of the opening-closing gestures.

Parallel to negative intensification of peak-contour statements, asonority is also a constituent of negatively intensified questions but is restricted to the low section of the high-rising f0 pattern, whereas in statements the early rapidly falling f0 favours continuation of breathy phonation. Both the late f0 synchronization and the extreme length ening of both parts of the compound noun contribute to the speaker's expressiveness. The slowing down of the two opening-closing movements reduces the cohesion between the compound elements, which become separately focused units with a stop release between them, and therefore lack consonant place assimilation [ndb] > [mb] at their juncture. This would have been very unlikely in "*handkerchief*" [ˈhæŋkətʃɪf], which is a highly cohesive word syntagma, as the contracted form "*hankie*" [ˈhæŋkɪ] indicates. On the other hand, the assimilated form would not fit in with the emphatic intensification of the whole compound, since in such a high-key context, [mb] would be decoded as referring to a non-existent lexical unit "*ham-bag*".

Figure 2: Speech wave, spectrogram, time courses of f0 (plain) and energy (dotted), and phonetic transcription of "*In a handbag."*, spoken by John Gielgud.



If the speaker had used modal or breathy-voice, instead of tense-breathy, phonation, lowered the over-all energy level and given the initial [h] normal duration, retaining the other features, she would still have conveyed intensified surprise, but no longer negative expressiveness, rather a positive lenient and sympathetic attitude towards the addressee. If the late concave valley synchronization of f0 is replaced by an early convex one and the extreme lengthening is removed, the question becomes a request for repetition or confirmation of a preceding statement, when for one reason or another, the speaker is not sure of having understood properly.

In this case stop consonant assimilation [ndb] > [mb] is highly likely.

Mr Worthing's socially incriminating admission "*In a handbag.*", spoken by John Gielgud (Figure 2), which preceded Edith Evans' outburst, was spoken low-key, with 67 dB, 71 dB, 64 dB in each of the 3 syllables, with normal consonant and vowel durations in the various structural positions, with a medial f0 peak of 131 Hz on the accented syllable, and with place assimilation [mb]. In the context of Edith Evans' over-all high-key, socially dominating speech, John Gielgud's low-key phonetic form signals social embarrassment and subordination. He uses "*handbag*" another four times, all similarly low-key with place assimilation [mb] or at least [n͡mb].

Edith Evans' line is so well known in the UK that people quote it frequently even when they do not know the play. Anybody can make the test and mention "a handbag" in the course of a casual conversation, and the chances are very high that someone will respond by quoting the line with a quite successful imitation of Evans' emphatic delivery. Journalist Helen Rumbelow referred to it in an article in the Times "The Queen's English. Go the America to hear it" on 6 April 2011. Writing about an interview with Jonnie Robinson, sociolinguistic curator in the British Library in London, she quotes his question "You see an attractive person. What do you call them, one slang word, beginning with P?", and his answer "Ask any teenager, they'll say 'peng'." Her comment is: "Peng? I repeat, Lady Bracknell style." It shows that the "historic" utterance is more than a literary stage line but illustrates a fundamental form – function relation in English speech communication, which the ordinary language user is aware of. This in turn means that the singular example and its acoustic manifestation can be taken as representative of negative intensification in surprise questions, even if the details vary with repetition.

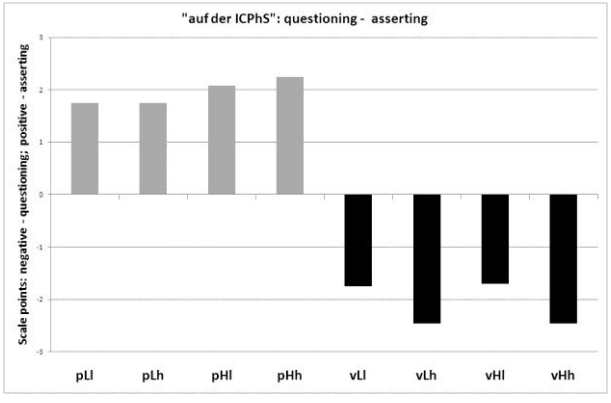
4. SOUNDS AND PROSODIES IN QUESTION AND STATEMENT PERCEPTION

Kohler [12] showed that German surprise questions and contrastive statements differ in several segmental and prosodic parameters. In a dialogue contextualization referring to the 1995 ICPhS in Stockholm, the surprise questions "*Bist du nicht in Stockholm? Auf der ICPhS*?" ("Are not you going to be in Stockholm? At the ICPhS?") have one pitch accent each with a concave high-rising valley contour synchronized late in the last syllable, reaching about the same high f0 value more than 1.5 octaves above the minimum. In addition, the preceding unaccented syllables form a high f0 prehead, and the final [s] of "*ICPhS*" is high-pitched with a sharp increase of spectral energy between 3 kHz and a spectral peak at 4 kHz, followed by a slow decline above, thus weighting the high spectrum.

On the other hand, the contrastive statements "Wir *treffen uns doch in Stockholm. Auf der ICPhS*." ("I thought we are going to meet in Stockholm. At the ICPhS.") have a falling peak contour synchronized late in the accented syllable. It reaches a low value in the speaker's pitch range in the fully voiced final syllable of "*Stockholm*", but descends by less than 2 st from the peak maximum before the voiceless fricative in the last syllable [ɛs] of "*ICPhS*", the late f0 fall being truncated in the voiceless environment. In addition, the preceding unaccented syllables form a low f0 prehead, and the final [s] of "ICPhS" is low-pitched with energy concentrated in the lower part of the spectrum and a faster decline in the higher part, compared with the high-pitched [s] in the question.

A perception experiment using the Semantic Differential technique [18] provided answers as to how these parametric differences in production are mapped onto the semantic decoding of questions and statements by listeners in a systematic 2 (f0 direction) x 2 (f0 prehead) x 2 (final fricative) design of the utterance "*Auf der ICPhS*". Along the semantic scale 'questioning/asserting' (Figure 3), pitch direction has the greatest influence, clearly separating valleys and peaks, irrespective of the just rudimentary f0 drop in the late peak pattern.

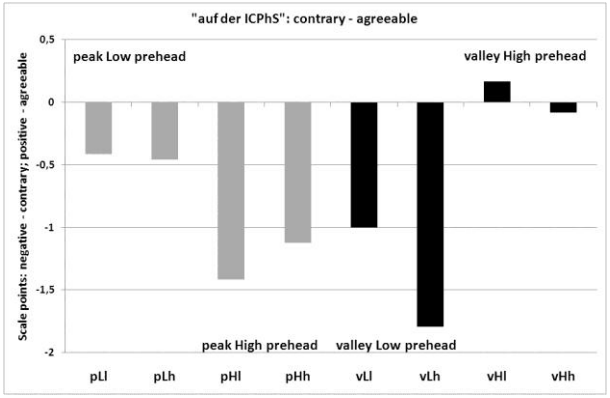
Figure 3: Means of judgments by 24 subjects for each of the stimuli on the Scale *questioning – asserting.* p/v=peak/valley contour, H/L=high/low prehead, h/l=high/low-pitched [s].



However, the low-pitched final fricative in the valley pattern (vHl, vLl) reduces the degree of questioning, compared with the high-pitched one. The greater acoustic weight of the higher fricative pitch functions as the prototypical manifestation of the fricative in a prototypical high-rising question, where tonal pitch is thus continued in the pitch of the final noise.

On the scale 'contrary/agreeable' (Figure 4), there is an interdependence between pitch direction and prehead: falling pitch is judged more contrary when it is introduced by a high prehead than by a low one, whereas for rising pitch it is the other way round. In both cases the pitch contrast between the beginning and the end of the utterances is widened. The wide range intensifies the contrastive assertion of the late peak and the surprise question of the late concave high-rising valley, in both cases conveying an attitude of contrariness in the speaker – listener relationship. This attitudinal interpretation of the complete utterance contour necessitates recognition of a wide processing window in speech perception, beyond local pitch accents (cf. also [5]).

Figure 4: Means of judgments by 24 subjects for each of the stimuli on the Scale *contrary – agreeable.*



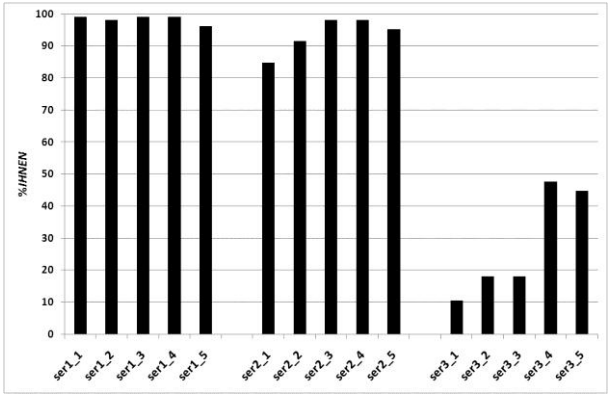
5. ARTICULATORY PROSODIES AND PHONETIC ESSENCE

It has been pointed out in section 2 that it may no longer be possible, in the analysis of low-key reduction, to bundle articulatory and acoustic properties in segmentable units corresponding to sounds and phonemes but that such properties manifest themselves in wider contexts as articulatory prosodies [10] of, e.g., nasalization, palatalization, glottalization. The impact of such long components on perception was investigated in [13], on the basis of a German spontaneous speech sample "*ich kann Ihnen das ja mal sagen*" ("I can mention this to you"), spoken as an aside in an appointment-making dialogue. The sequence of function words *kann Ihnen das*, whose canonical phonemic representation will be /kan iːn(ə)n das/, is realized as [k̟ʰɛ̈nʲnʲəs] without a segment [i], but with palatalization spread across the long nasal environment. This palatalization differentiates the utterance from non-palatalized *kann das* [kʰannas], which is equally possible in the same context "*ich __ ja mal sagen*" but does not contain the lexical information "*Ihnen*".

From the original spontaneous utterance another two were derived by splicing: (1) the central nasal section with the strongest palatalization, visible in the spectrogam, was excised, (2) the entire section [k̟ʰɛ̈nʲnʲ] was replaced by [kʰann] from another low-key utterance, not containing "*Ihnen*", by the same speaker. In each of these three stimuli, the duration of the nasal section was manipulated in 5 steps between the extremes set by the original [k̟ʰɛ̈nʲnʲ], the longest, and [kʰann], the shortest. The resulting 3 series entered a perception test in which listeners had to decide whether they perceived "*ich kann Ihnen das ja mal sagen*" or "*ich kann das ja mal sagen*".

The results in Figure 5 show a very clear break in the perceptual profiles between ser1-2 and ser3, in relation to the more drastic removal of palatalization in the nasal of ser3. There is also an influence of nasal duration on *Ihnen* identification in ser3. If there is strong palatalization, irrespective of the presence or absence of the most palatalized central section, identification of "*Ihnen*" is practically perfect and independent of duration in the range used in the test. If palatalization is absent, duration is only partially able to trigger "*Ihnen*".

Figure 5: Percentages of *Ihnen* judgements obtained for the 3(series)*5(durations), 21 subjects and 5 repetitions.



The results of this perceptual experiment demonstrate a further aspect of the segment – prosody interplay. Segments not only vary together with prosodies according to the functions their

combined phonetic effects are to fulfil in speech communication, which may sometimes appear as segments being shaped by prosodies, at other times as prosodies being shaped by segments. Over and above such interaction, segments are themselves malleable into becoming articulatory prosodies.

It is these long articulatory components that retain the phonetic essence of words in their low-key variability [17]: segments qua segmentable property bundles may disappear but the essential properties must be kept, as is illustrated by the phonetic manifestation of the extreme low-key reduction of "*I don't know*", and as shown again in the retention of palatality in spite of the disappear-ance of a high front vowel segment in "*kann Ihnen das*". These segment-prosody interplays and these articulatory prosodies are produced by the speaker and perceived and cognitively processed by the listener as carriers of communicative functions in speech interaction.

6. CONCLUSION AND OUTLOOK

When phonetic substance is studied from the perspective of communicative function in speech interaction, sounds and prosodies lose their established status of separate research fields in their own right. They are to be taken for what they are: heuristic tools in metalinguistic analysis, which provide a useful scaffolding for building up language descriptions without becoming parts of the object of study in its daily use in communication. We need to analyse phenomena in either field with constant awareness of the interrelation between these levels of observation within a communicative frame. When we analyse sounds we need to systematically vary their prosodic setting, because they vary with it. When we analyse prosodies, we need to look at how they are synchronized with vocal-tract articulation, and how segmental manifestations support and can even take over prosodic functions. So, prosodies also depend on sounds.

The dichotomy reflects the double focus in linguistic analysis that has been practised for centuries, on words and their distinctive sound markers, especially segmental phonemes, and on supralexical and suprasegmental structures in utterances. This focus needs to be adjusted to the functional goal, which requires new methodologies of data acquisition. It will no longer be sufficient to use sentence frames of the type *Say X again*, where X varies lexical items, often logatomes, along postulated phonemic distinctions in the language under investigation [15] because the metalinguistic contextualization highlights the citation form whose information value for communicative usage is low. Nor should experiments in speech under-standing continue to be modelled on the assumpt-ion of segmental phoneme perception and restorat-ion [3], since there is now sufficient evidence that listeners decode utterances by direct reference to fine phonetic detail that goes well beyond feature bundles in phonemic segmentation [8, 13, 17].

In prosodic data collection for the investig-ation of communicative functions, sentences must be meaningful, which rules out data samples of the type "*Die Nonne und der Lehrer wollen der Lola in Murnau eine Warnung geben, und die Hanne will im November ein Lama malen*."("The nun and the teacher want to give a warning to Lola from Murnau, and Hanna wants to paint a lama in November.") [20]. Question-answer paradigms of the type [21] used in the study of contrastive focus also need refinement to tease apart propositional, attitudinal and expressive meanings.

Prompt:	Target:
Who may know your niece?	Lee may know my niece.
What may Lee do to your niece?	Lee may lure my niece.
Who may Lee know?	Lee may know my niece.
What did you say?	Lee may know my niece.

Future research into the parallel, but differentially weighted segmental and prosodic contributions to the signalling of specific speech functions should draw on two data sources. First, corpus data, collected outside the individual research question they are to be analysed for, and in a variety of scenarios from reading to different forms of spontaneous speech, provide a rich source of segmental and prosodic variability in words and utterances. Admittedly, they lack the systematicity of constructed laboratory speech designs, but nevertheless allow the formulation of tentative hypotheses for further systematic data collection. To develop its full potential, such corpus analysis must, however, go beyond auditory, articulatory or acoustic descriptive accounts of singled out phonemic segments contained in specifically selected words, as e.g. in [7], and include the segmental and prosodic environmental setting above the word. In the next step, hypotheses may be tested with a second data source from constructed scenarios in which speakers enact communicative interchanges fulfilling functions that are the goal of the analysis, e.g. negative or positive intensification [16]. Such situational

contextualization also needs to be applied to perception experiments.

It follows from the preceding argumentation that future phonetic research would greatly increase its fidelity to its object of investigation, and its scientific impact if it were to make the following points its guiding principles:

- It recognises communicative function in all fields of meaning – propositional, attitudinal and expressive – as the frame of reference for speech analysis.
- It abandons the dichotomy of autonomous fields of sounds and prosodies in speech communication, although it still recognises its heuristic value in the progression of scientific knowledge acquisition.
- It abandons the phoneme as a cognitive unit in speech production and perception, and thus also gives up the phonetics – phonology division in speech research: phonetics deals with all oral and aural patternings in speech communication.
- It develops a methodology of data acquisition and elicitation to solve these function-oriented phonetic research issues. There is no one standard technique, specific techniques depend on the goals to be achieved, but they all need to be united in the ultimate goal to elucidate speech communication in human language. Word and sentence orientation in lab speech will still be a necessary part of this phonetic methodology, for example in the modelling of motor control, but it should not lose sight of the over-all communicative approach into which it needs to be integrated, and should always ask the question "How natural can one make the elicitation process, and how can one best guarantee that level of naturalness?"

7. ACKNOWLEDGEMENTS

Many fruitful discussions with W.J. Barry on the issues of sounds and prosodies, as well as of methodologies in data acquisition, and his very helpful comments on a previous draft of this paper are gratefully acknowledged.

8. REFERENCES

[1] Armstrong, L.E., Ward, I.C. 1926. *A Handbook of English Intonation*. Cambridge: Heffer.

[2] Baumann, S., Becker, J., Grice, M., Mücke, D. 2007. Tonal and articulatory marking of focus in German. *Proc. 16th ICPhS* Saarbrücken, 1029-1103.

[3] Ernestus, M., Baayen, H.R., Schreuder, R. 2002. The recognition of reduced word forms. *Brain and Language* 81, 162-173.

[4] Firth, J.R. 1948. Sounds and prosodies. *Transactions of the Philological Society*, 127-152.

[5] Grabe, E., Gussenhoven, C., Haan, J., Marsi, E., Post, B. 1997. Preaccentual pitch and speaker attitude in Dutch. *Language and Speech* 41, 63-85.

[6] Grammont, M. 1934. *La Prononciation Française* (8th ed.). Paris: Librairie Delagrave.

[7] Harrington, J. 2006. An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34, 439-457.

[8] Hawkins, S. 2003. Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics* 31, 373-405.

[9] van Heuven, V.J., Sluijter, A.M.C. 1996. Notes on the phonetics of word prosody. In Goedemans, R., van der Hulst, H., Visch, E. *Stress Patterns of the World, Part 1: Background*. The Hague: Holland Academic Graphics, 233-269.

[10] Kohler, K.J. 1999. Articulatory prosodies in German reduced speech. *Proc. 14th ICPhS* San Francisco, 89-92.

[11] Kohler, K.J. 2006. Paradigms in experimental prosodic analysis: from measurement to function. In Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schließer, J. (eds.), *Methods in Empirical Prosody Research*. Berlin, New York: de Gruyter, 123-152.

[12] Kohler, K.J. 2011. Communicative functions integrate segments in prosodies and prosodies in segments. *Phonetica* 68.

[13] Kohler, K.J., Niebuhr, O. 2011. On the role of articulatory prosodies in German message decoding. *Phonetica* 68.

[14] Lindblom, B.1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

[15] Mooshammer, C., Geng, C. 2008. Acoustic and articulatory manifeatations of vowel reduction in German. *J. Intern. Phon. Ass.* 38, 117-136.

[16] Niebuhr, O. 2010. On the phonetics of intensifying emphasis in German. *Phonetica* 67, 170-198.

[17] Niebuhr, O., Kohler, K.J. 2011. Perception of phonetic detail in the identification of highly reduced words. *Journal of Phonetics* 39.

[18] Osgood, C.E., Suci, G.J., Tannenbaum, P.H. 1957. *The Measurement of Meaning.* Urbana, Chicago, London: University of Illinois Press.

[19] Port, R.F. 2008. All is prosody: Phones and phonemes are the ghosts of letters. *Proc. 4th International Conference of Speech Prosody* Campinas, Brazil.

[20] Truckenbrodt, H. 2002. Upstep and embedded register r levels. *Phonology* 19: 77-120.

[21] Xu, Y., Xu, C.X. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33, 159-197.