

Communicative Functions Integrate Segments in Prosodies and Prosodies in Segments¹

Klaus J. Kohler

Institute of Phonetics and Digital Speech Processing, University of Kiel, Kiel, Germany

Abstract

This paper takes a new look at the traditionally established divide between sounds and prosodies, viewing it as a useful heuristics in language descriptions that focus on the segmental make-up of words. It pleads for a new approach that bridges this reified compartmentalization of speech in a more global communicative perspective. Data are presented from a German perception experiment in the framework of the Semantic Differential that shows interdependence of f₀ contours and the spectral characteristics of a following fricative segment, for the expression of semantic functions along the scales questioning – asserting, excited – calm, forceful – not forceful, contrary – agreeable. The results lead to the conclusion that segments shape prosodies and are shaped by them in varying ways in the coding of semantic functions. This implies that the analysis of sentence prosodies needs to integrate the manifestation of segments, just as the analysis of segments needs to consider their prosodic embedding. In communicative interaction, speakers set broad prosodic time windows of varying sizes, and listeners respond to them. So, future phonetic research needs to concentrate on speech analysis in such windows.

Copyright © 2011 S. Karger AG, Basel

1. The Dichotomy of Segments and Prosodies

This paper takes a renewed look at the established dichotomy between the fields of segments and prosodies. The segmental concept and the potential dichotomy pre-date scientific concern with sound systems. It is several thousand years old, as old as the invention of alphabetic writing in the Semitic language environment, which resulted in the graphic representation of words by means of characters that refer to segments. Crucially, the link in Semitic languages between triconsonantal sequences and the lexico-semantic field was the essential communicative frame which allowed the concept to become established in consciousness. In the late 19th and the 20th century,

¹ Audio examples, linked to figures, are made available in wav files as online supplements at www.karger.com/doi/10.1159/000328433.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2011 S. Karger AG, Basel
0031–8388/11/0682–0026
\$38.00/0
Accessible online at:
www.karger.com/pho

Prof. Dr. Klaus J. Kohler
Holm 4
D–24113 Molfsee (Germany)
Tel. +49 431 651 453, E-Mail kjk@ipds.uni-kiel.de

this culminated in the development of phonemic theory and word phonology of vowel and consonant systems.

The long tradition of the human mind dealing with the sounds of speech raises the question as to whether there may be something intrinsic to speech production that favours a segmental approach at the level of the lexical word, as against the global prosodies of the utterance. Such an intrinsic attribute may be related to different time windows in the vocal-tract, in the glottal and in the subglottal dynamics of the production process, and can then be offered as an alternative explanation to a graphemically led segmental consciousness, which, in most languages, is divorced from the consonantal regularities that gave birth to the segment in Semitic writing.

Vocal-tract articulation is typically characterized by short-term opening-closing movements, which are the basis of the syllable concept, and which embody different degrees of stricture giving rise to further segmentation of consonantal elements in close, and vocalic elements in open, strictures [MacNeilage, 2008]. Glottal phonation is also short-term, inasmuch as it can be switched on and off – or, more rarely, adjusted for other states of the glottis – and linked locally, as voiced/voiceless, to the segments of the opening-closing movements. Voicing is favoured in vowels and sonorants, and voicelessness in obstruents, the opposite combinations occurring more rarely in the languages of the world [Lindblom and Maddieson, 1988]. So, this synchronization of the short-term time courses of vocal-tract articulation and phonation buttresses the concept of a segment, which is in turn best suited to the communication of a message at the level of the short communicative-unit window, namely of the morpheme and word. In tone and tonal-accent languages frequency of vocal fold vibration in the articulatory open phase is used as a further segmental attribute for lexical differentiation, and in some languages (e.g. in South-East Asia) mode of vibration is added as yet another more local, lexical differentiator [Abramson and L-Thongkum, 2004].

However, there are also communicatively important long-term phonation settings, as in individual, regional, social or language voice qualities [Laver, 1980], but also in global falling(-rising), or rising(-falling), or level pitch contours [Kohler, 2006b], as well as in pitch registers across whole utterances and utterance sections. Likewise, vocal-tract dynamics have a long-term setting in the ‘basis of articulation’ [Honikman, 1964], or in long-term articulatory prosodies of nasalization, labialization, palatalization, velarization for linguistic function [Firth, 1948], or in an appellative function of, e.g., motherese, where global lip rounding may be used in soothing an infant. Subglottal dynamics are primarily long-term in breathing and force [Ladefoged, 1967], and are therefore typically associated with whole utterances. However, both tonal and subglottal dynamic features also operate relationally within an utterance, modifying the long-term setting for shorter-term communicative purposes, for example in focal or contrastive pitch accents [Ladd, 1996; Kohler, 2006a], or in a force accent for negative intensification of a word or syllable, as against the long-term expression in the emotion of anger [Kohler and Niebuhr, 2007; Niebuhr, 2010].

Looking at speech production from the typical and subsidiary operations of these elementary physical components can provide us with an understanding of why the segment-prosody divide arose and why it became the basis of phonetic and phonological research. There have always been research questions that have bridged the separated fields. The most prominent case are the evergrowing number of analyses of the physical exponents of stress in the languages of the world since Fry [1955, 1958, 1965], demonstrating that stress is signalled by f_0 change above a critical threshold [Fry, 1958, for

English; Isačenko and Schädlich, 1966, for German; Sluijter and van Heuven, 1995, for Dutch], by duration, specifically in vowels [Fry, 1955, for English; Sluijter et al., 1997, for Dutch], but also in consonants, particularly initial ones [Huggins, 1972, for English], by spectral expansion in vowels [Fry, 1965, for English; van Bergem, 1993, for Dutch], and by greater acoustic energy [Fry, 1955, for English; Sluijter and van Heuven, 1996, Sluijter et al., 1997, for Dutch], in this order of relevance. Huggins' [1972] observation that segment durations not only determine perceived speech timing via Just Noticeable Differences, but also influence the perception of stress in a much wider prosodic frame, is significant for this intertwining of segments and prosodies. When in the sentence *The hostel for paupers is more lawful than it was*, the duration of [l] in *lawful* was varied, primary stress shifted between *lawful* (long [l]) and *was* (short [l]). So, the prominence of *was* was affected by a segment that was four syllables away.

The investigation of speech rhythm provides a striking example of how an initial suprasegmental approach with such categories as 'stress timing' and 'syllable timing' as well as 'perceptual isochrony' and 'P centres' was changed to paradigms built on segments, such as the Pairwise Variability Indices of vocalic and intervocalic durations [Kohler, 2009a]. There has, however, lately been an overdue return to time windows that are expanded beyond the segment and also include f0 in rhythmic chunking [Arvaniti, 2009; Barry et al., 2009; Niebuhr, 2009a; Nolan and Asu, 2009; Cumming, 2010, 2011].

Recently, there has been an upsurge in the study of prosody, more particularly the study of intonation within the framework of Autosegmental Metrical Phonology, but the latter has, on the one hand, predominantly focussed on f0, while on the other hand introducing segmental concepts in the representation of pitch contours as sequences of local tones. The segmental orientation has been further strengthened by the notion of 'alignment', which relates specific prosodic points, e.g., an f0 peak value, to a segmental anchor, e.g., the vowel on- or offset [Arvaniti et al., 1998; Ladd et al., 1999, 2000], thus reinforcing a picture of (hypostasized) local tones being linked to (hypostasized) segments. Latterly, several investigations [e.g. Mücke et al., 2009] have supplemented the analysis of segmental anchoring in acoustic records by an analysis of articulatory anchors in articulo-graphic records, but the alignment of specific points in f0 and vocal-tract trajectories has remained.

Such point-to-point alignment is incapable of capturing the distinctive characteristics of different phonological categories of peak and valley contours that are temporally coordinated with vocal-tract opening-closing gestures. For example, the distinction between *early*, *medial* and *late peak* patterns [Kohler, 2005, 2006b] crucially depends on the temporal relation of the whole rising-falling f0 contour to the opening and closing of the vocal tract. In the *early peak*, it is the high-low descent, together with decreasing acoustic energy; in the *medial peak*, it is the low-high ascent, together with increasing acoustic energy, that coincides with the opening phase; in the *late peak* the low-high ascent, together with increasing acoustic energy, coincides with the closing phase [Kohler, 2009b]. To refer to this coordination of global vocal-tract and prosodic dynamics, Kohler [2006b] has proposed the concept of 'synchronization' to complement phonetic point 'alignment'. The holistic nature of the different timing of f0-peak contours is further emphasized by the weight the shape of a rise-fall has on the signalling of the intonational categories. For instance, even if the maximum peak point occurs before or at the beginning of the vocal-tract opening, a slow f0 descent that keeps the initial pitch level high results in decoding a medial peak [Niebuhr, 2007].

There is thus a broad range of phonetic point alignment within a phonological synchronization category.

Hawkins and Smith [2001] and Hawkins [2003] have shown conclusively in their Polysp framework that the interplay of prosodies with fine segmental vocal tract detail 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. Whereas the weakly contracted form *I don't know* provides the listener with neutral information, more strongly elaborated or reduced forms add expressive and appellative connotations. An overarticulated expansion to *I do not know* signals the speaker's unwillingness to accept the listener's insistence, and tense breathy, instead of modal, phonation as well as pauses in between the words add degrees of exasperation. At the other end of the elaboration-reduction scale, the maximally contracted form [ʔ̃ñ] with very weak segmental articulation in 'rather stylized intonation and rhythm', 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. [ññ]. The rhythmic timing also reflects the fuller form, with the central section being shortest, the final one longest. And finally, the intonation is most likely an *early peak* constellation, a high-low f_0 descent occurring in the opening phase of the final, accented section. In English, as in German, this peak synchronization signals finality in argumentation structure – the closure of a communicative turn [Kohler, 2005, 2006b]. A *medial peak* would not fit in the communicative situation, as described above, because it signals argumentative openness [Kohler, 2005, 2006b].

The German equivalent of English *I do not know*, in all its formal manifestations and semantic and pragmatic functions, is *keine Ahnung*. Figure 1 compares the neutral [k^hä:ñë 'ä:nöŋ] (fig. 1a), reduced [hä:ë 'ãö] (fig. 1b), elaborated ['k^hai:në 'ʔä:nöŋ] (fig. 1c), reinforced tense breathy ['khaï:në 'ʔä:nöŋ] (fig. 1d) utterances in their wave forms and spectra, f_0 as well as energy time courses. The neutral version has four clearly demarcated opening-closing vocal-tract movements, with glottalization taking over 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* would also be possible in the neutral version to signal argumentative openness. 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. The timing of the gestures is also kept, and an early peak pattern is still associated with the prominent third one, this time exclusively so to fit the conclusiveness of the communicative situation, as described for English. In the elaborated version, articulation and prosody are intensified: the gestural movements are

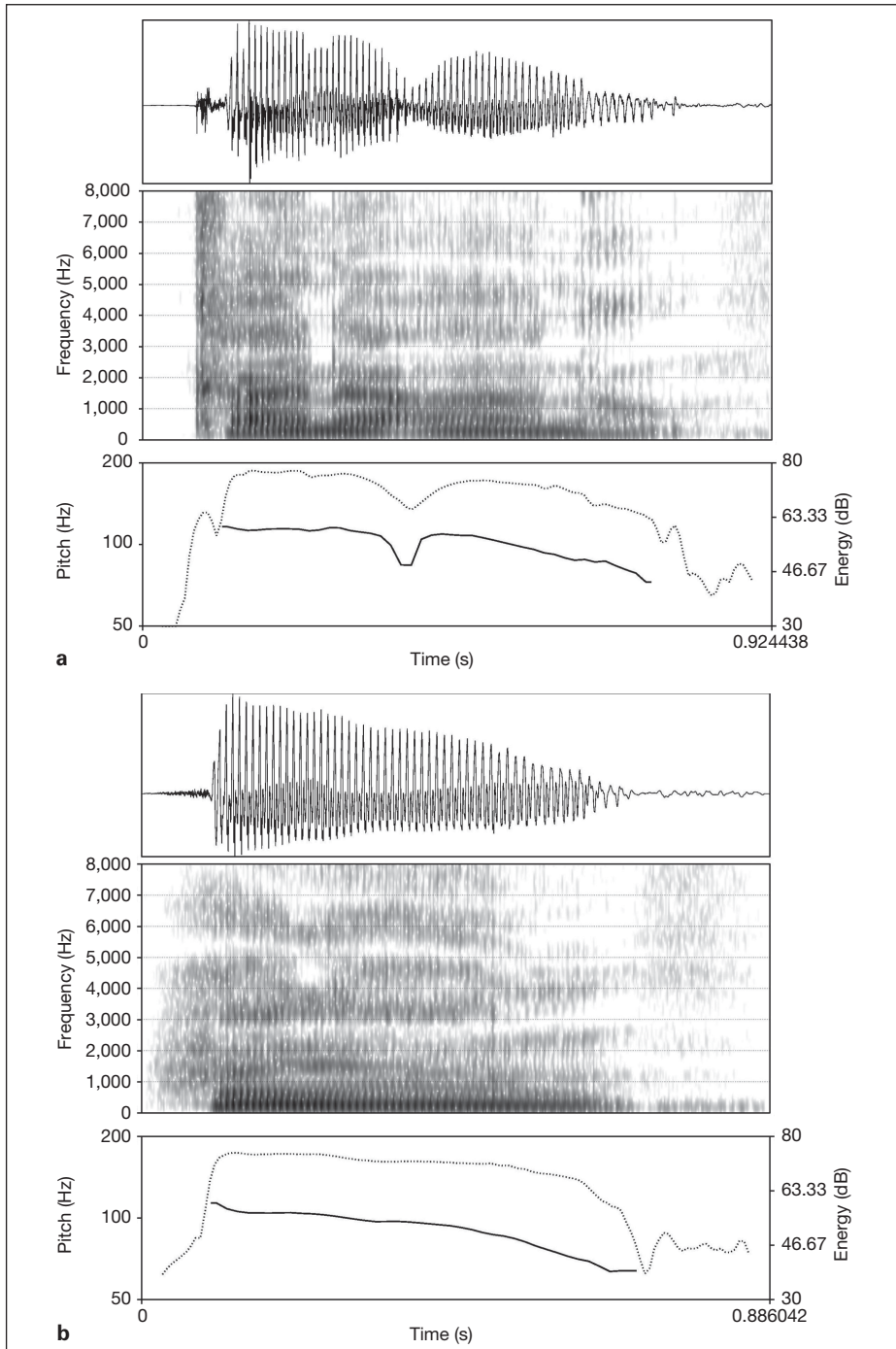
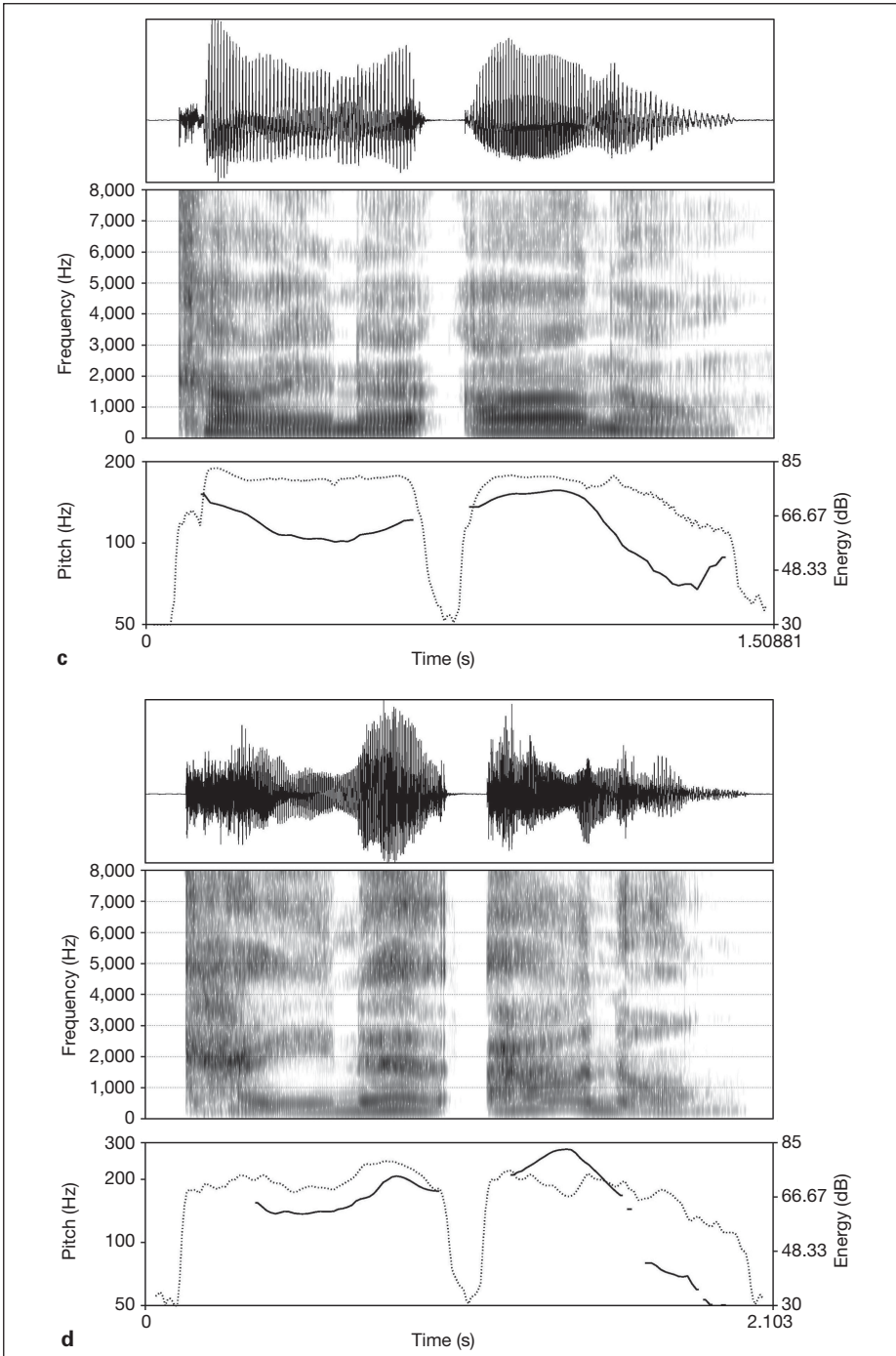


Fig. 1. Speech wave, spectrogram, fundamental frequency (plain line), and energy (dotted line) of *keine Ahnung*: neutral (a), reduced (b), elaborated (c), reinforced with negative intensification (d), spoken by the author. Online supplementary audios 1–4.



expanded in time and space [cf. also Lindblom's, 1990, H & H theory], 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, with an *early peak* pattern in the former and a *medial peak* with high-plateau intensification in the latter. Such an intonation would not go together with the reduced version, because it functions as high-key emphatic reinforcement and would therefore clash with the low-key function of the reduced form. When modal phonation is replaced by tense breathiness throughout the utterance, and the second, unstressed syllable *ne* receives an additional *medial peak* accent, an effect of negative intensification [Niebuhr, 2010], expressing exasperation, is introduced.

These examples demonstrate that communicative function determines the phonetic output as a whole, and that segmental as well prosodic properties are shaped jointly for subtle functional coding of meaning in speech interaction. The preceding discussion shows two things. On the one hand, the question of how to bridge the divide between segments and prosodies has been raised repeatedly, but on the other hand, segmental analysis has set the research pattern in mainstream phonetics and phonology and dominated it for more than a century, which is also evidenced by the terminological reference of prosodies to 'suprasegmentals'. What we need in the future is to study phenomena in either field with constant awareness of the interrelation between these levels of observation within a communicative frame. If we analyse segments we need to examine their variation in different long-term prosodic settings, because segments vary with them. If we analyse prosodies, we need to look at how their time courses are synchronized with the temporal sequencing of vocal tract opening-closing movements, and how segmental strictures within these movements support, and can even take over, prosodic functions, so prosodies also depend on segments. This interrelated approach equally argues against a concept of 'infraprosodies', as advocated in Port [2008, 'All is prosody: phones and phonemes are the ghosts of letters'], because it can offer no explanation why segmental abstractions from opening-closing gestures have occurred prior to, and independently of as well as outside, alphabetic writing systems.

The focus on segments and the segment-prosody divide has certainly provided an 'as-if' procedural heuristics [Vaihinger, 1920] in phonetic analysis, a useful service as a scaffolding for building phonetic descriptions, starting at the word level. But it becomes a hindrance to progress in the investigation of speech communication if it is invested with reality and becomes part of the building. The scaffolding needs to be pulled down when a sufficient body of knowledge about potential interrelations between the segregated parts has accumulated. It can be argued that we have now reached that stage, especially for well-described languages, of which German is one.

2. The Research Question

2.1. Acoustic Properties of Natural Speech Stimuli

Phonetic research has led to the persuasive proposition that segments shape prosodies and are themselves shaped in suprasegmental frames. For example, voiceless obstruents determine the f_0 course of peak contours [Gartenberg and Panzlaff-Reuter, 1991; Kohler, 1996]. On the one hand they raise f_0 onset, without the perceived contour class being affected. On the other hand they mask it, e.g. after short vowels in late peak contours of utterance-final monosyllables, where the f_0 descent is strongly truncated,

compared with the f_0 fall in early peaks due to time constraints [cp. Niebuhr, 2008, for German]. Grabe [1998] also found this truncation in her German data for a pattern she classified as H^*+L in ToBI notation. She compared it to compression in what she noted as the same pattern H^*+L in English. However, it follows from her data acquisition methodology that she used different contextualizations to obtain H^*+L samples in the two languages, favouring contrast and unexpectedness in German but new observation in English, which result in different peak contour categories, a late peak in German versus a medial peak in English, with different consequences for the manifestation of f_0 across syllables ending in voiceless obstruents [Kohler, 2006b]. So, what is interpreted as a difference of alignment strategies in the two languages in reality relates to two functionally distinctive peak patterns that belong to both languages. This also means that the non-occurrence of truncation in the rising pattern L^*+H , which Grabe [1998] found in her data of both languages, does not create a compression-truncation asymmetry for German. Rising contours are forced to reach their f_0 targets for the listener to be able to decode the final pitch level in the highly variable upper range of a speaker's voice, whereas the final pitch level of a curtailed fall can be reconstructed from its onset to the lower end of a speaker's voice range.

Contrariwise, segments have variable realization depending on prosodic patterns. The extreme case of this prosodic dependence of segment exponency is found in reduced versus full forms in function words [Kohler, 1999; Kohler and Niebuhr, 2011, Niebuhr and Kohler, in press]. Moreover, Baumann et al. [2007] provided spectrographic data for German of vowels in different prosodic patterns, showing more peripheral vowel articulation in narrow and contrastive as against broad focus: /u:/ is more raised and retracted, /o:/ more raised and/or retracted, /i:/ more raised and/or fronted, /a:/ more lowered and fronted. Contrastive focus, compared with the other two, is realized as a late peak pattern with a higher maximum f_0 and a greater pitch excursion in a longer nuclear syllable. The contrastive function is thus mapped onto an intensification of production parameters simultaneously in voice control and opening-closing articulatory movements.

This phenomenon is also manifested in aspiration differences of syllable-final [t] after short [ɪ] in late versus early pitch peak contours, where Niebuhr [2008] found significant distinctions for German in several acoustic properties. The f_0 descent is typically truncated in the late peak but reaches the low value of the speaker's voice range in the early peak. The curtailing of f_0 goes together with longer aspiration, a lower short-term energy maximum in the aspiration noise, and a higher-frequency spectral energy maximum at the aspiration onset with stronger downdrift across the noise. All these features point to an intensification of the closing phase in voice control and articulatory gesture: the occlusion is performed with considerable lengthening of the vocoid during mainly rising rather than continuously falling f_0 and is released more gradually with higher resistance to airflow, creating an initial higher-frequency spectral energy maximum and a longer and stronger downdrift. These acoustic differences in the aspiration noises of the late and early peaks are perceived as pitch differences, terminating the high-low pitch fall in the early peak and supplementing the small f_0 descent in the vowel of the late peak by a pitch fall in the aspiration. Thus, the articulatory dynamics heighten the prosodic peak patterns, which Niebuhr [2008] confirmed in a complementary perception experiment.

Furthermore, Niebuhr [2009b] showed how the realization of German word-final voiceless fricatives [ʃ x] and reduced vowels [ə ɐ] vary in valley as against peak patterns. Test utterances containing the target words were contextualized in short dialogue texts to trigger terminal-falling medial peak contours linked with asserting

statements, or high-rising valley contours linked with surprise questions. The mean centre of gravity in the voiceless fricatives is significantly higher and the CoG range significantly smaller in the high-rising valley than in the terminal-falling peak, resulting in higher pitch percepts. In [ɐ], F1 is significantly lower and F2 significantly higher in the high-rising valley at the centre and the offset of the vowel. In [ə], F1 is not significantly different, but F2 in valleys has a significant increase at the offset giving the auditory impression of a rising diphthong. These data again point to vowel articulation moving away from a central to a more peripheral position, and fricative strictures being formed more tightly with higher flow resistance, in high-rising f0 patterns. In both cases, this mirrors simultaneous intensification in voice control and opening-closing gesture.

The investigation presented here now looks more closely at this bidirectional prosody-segment interdependence of peak or valley contours and fricative strictures in opening-closing gestures by taking the following example from the appointment-making scenario of the Kiel Corpus of Spontaneous Speech for German [IPDS, 1995/1996] as its point of departure: *wie wäre es denn mit einem gemeinsamen Besuch in Stockholm bei der ICPHS* ('how about a joint visit to Stockholm at the ICPHS', g105a000 [Kohler, 2006b] (fig. 2).

Note that although the example has interrogative syntax, the question mark is omitted on the basis that, in terms of speech interaction, the speaker is not asking a question requiring an open polar reaction of acceptance or rejection but making a proposition, which she expects the dialogue partner will accept. It is this argumentation structure, reflecting the speaker's intention, that determines the prosodic implementation of the utterance. The interrogative syntactic structure is filled with a falling pitch pattern of a late peak contour [Kohler, 2006a] on the final accented word [ʃtək'hɔlm], and the same falling pattern is then repeated in the appended elliptic specification [itsepeha'ɛs]. The propositional argumentation structure accounts for the pitch being falling [Kohler, 2004]. At the same time, the speaker contrastively singles out a particular occasion for a meeting which is new in the appointment-making task, and by using the conditional verb form and the modal particle *denn* she indicates that the suggestion may be unexpected for the listener. This explains the late peak, which adds the speaker's affective evaluation to the contrast [Kohler, 2006a, b]. It is unimportant that the recipient might respond: *OK, great!* or *No, that won't work because I have to be in Australia at that time*. The listener is free not to take up the argument the speaker has signalled, and therefore the listener's response has no implications on the analysis of the syntax-prosody manifestation in the speaker's utterance in terms of speaker-related argumentation structure rather than external information structure.

The first late f0 peak is fully realized in the voiced segmental environment, whereas the second only has a very small f0 descent from the peak value, corresponding to <1 semitone, due to the following voiceless fricative context. Nevertheless, both contours are identified as the same late peak category. But the example also shows that in [ʃtək'hɔlm], the f0 peak contour takes place in an energy time course that rises fast in ['hɔlm] and then descends more slowly. So both prosodic parameters develop in parallel over time, and this combination of f0 and energy is clearly perceivable as reinforcement. In the case of [itsepeha'ɛs], this is less clear, but still recognizable: at the glottal-stricture juncture of [ha'ɛs], there is a dip in the speech signal amplitude followed by another increase and subsequent decrease, i.e. at the point where f0 starts

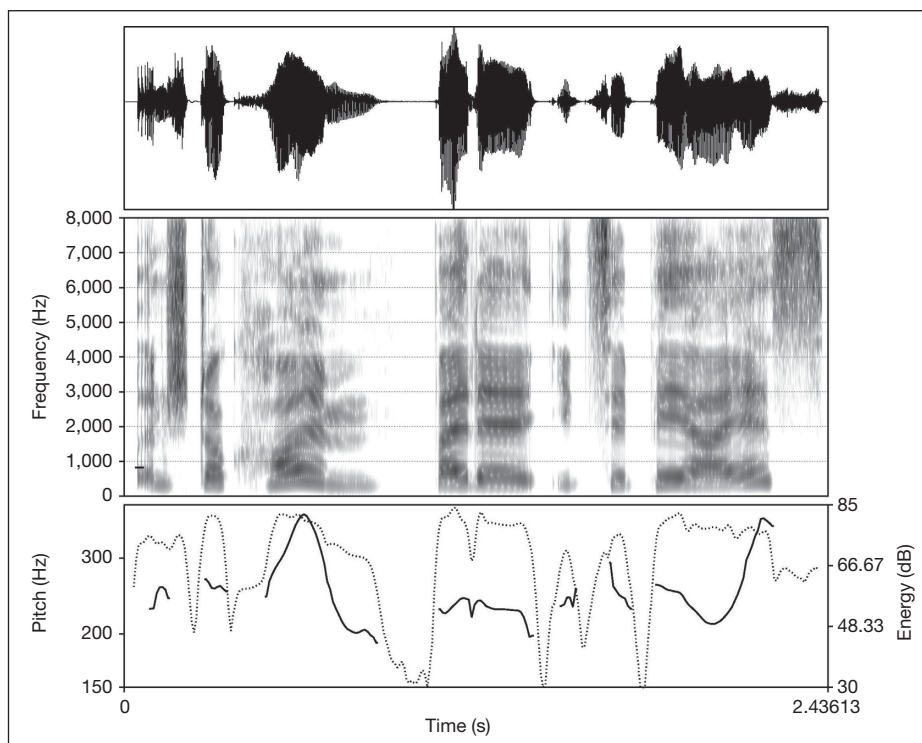


Fig. 2. Speech wave, spectrogram, fundamental frequency (plain line), and energy (dotted line) of *in Stockholm bei der ICPhS*, Kiel Corpus of Spontaneous Speech, Appointment Scenario, female speaker g105a000. Online supplementary audio 5.

rising to the peak value, which is followed by a very small fall, and the reinforcement is again perceivable, fully comparable to the reinforcement in [ˈhɔlm].

Since it is to be assumed that the characteristic production parameters of this utterance would be mapped onto perception and the cognitive processing of its meaning, a perception experiment was devised. Starting from the spontaneous speech sample, the author recorded *in Stockholm. auf der ICPhS. / in Stockholm? auf der ICPhS?* several times as affectively contrastive statements suggesting unexpectedness, with late peak contours, and as contrastive questions expressing incredulity, with late concave high-rising valley contours [Kohler, 2006a]. Figure 3a shows an instance of the sequence *in Stockholm. auf der ICPhS.* from this recording, corresponding to the Kiel Corpus example of figure 2; figure 3b gives an instance of the sequence *in Stockholm? auf der ICPhS?* In all cases, the late peak in the statement had a truncated fall in *ICPhS*, as against its full realization in *Stockholm*, in the same way as in the Kiel Corpus example (fig. 2, 3a). In all the questions, a late concave high-rising valley was associated with the accented syllables of *Stockholm* and of *ICPhS*; its maximum f_0 was substantially higher than the peak maximum in the statements, and the same final high value was reached in both segmental contexts (fig. 3b). In addition, a long high precursor contour preceded the valley in the

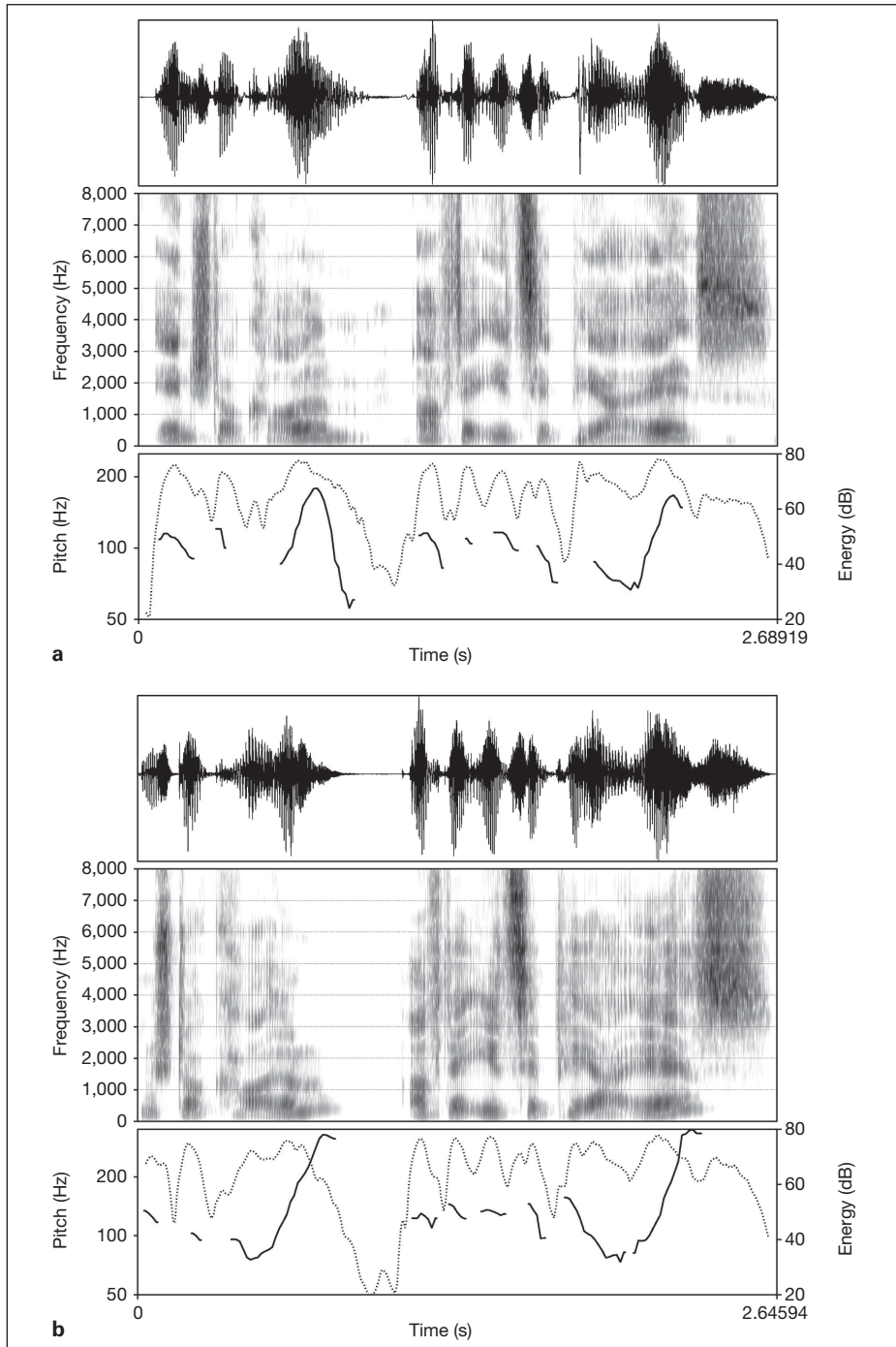


Fig. 3. Speech wave, spectrogram, fundamental frequency (plain line), and energy (dotted line) of the statements *in Stockholm. auf der ICPhS.* (a) and of the questions *in Stockholm? auf der ICPhS?* (b), spoken by the author. Online supplementary audios 6, 7.

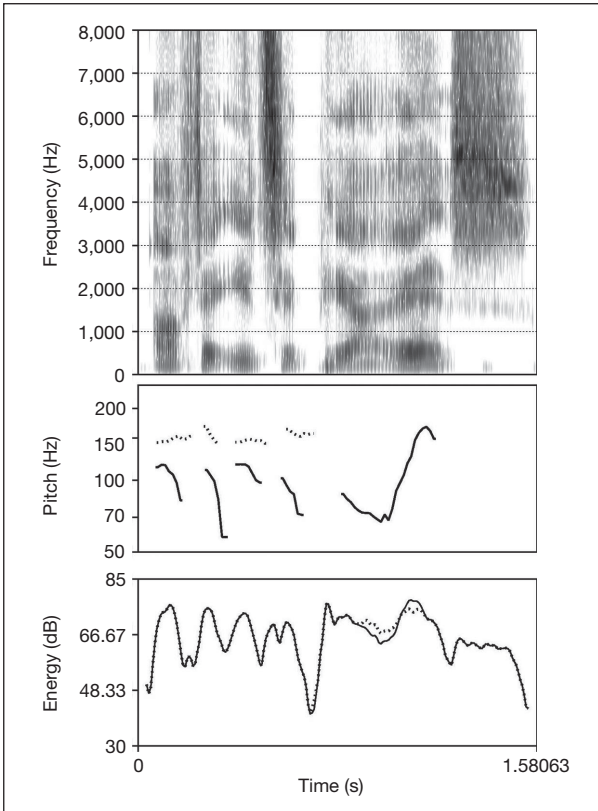


Fig. 4. Spectrogram, fundamental frequency, and energy of original statement *auf der ICP_hS*. produced by the author; *high precursor* (dotted in the f₀ window), *levelled energy* (dotted in the energy window) from the question realization are added to show the systematic differences. Online supplementary audios 8–11.

questions *auf der ICP_hS?*, around an average level of 140–150 Hz, as against a low level around 110 Hz before the peak in the statements. The frication of the final [s] of /itsepeha'ɛs/ was perceived as low-pitched in the statements, but high-pitched in the questions.

2.2. Selection of Templates for Stimulus Generation

The data described in 2.1 provided the frame for the test construction. Two utterances *auf der ICP_hS* were selected from the sets produced by the author in the statement and question forms with a contrastive late peak and a late concave high-rising valley, respectively. The statement utterance went into further processing, the question form was only used to obtain the question precursor, the final [s] and the maximum f₀ value in the accented vowel. Figures 4 and 5 represent the statement and question features that

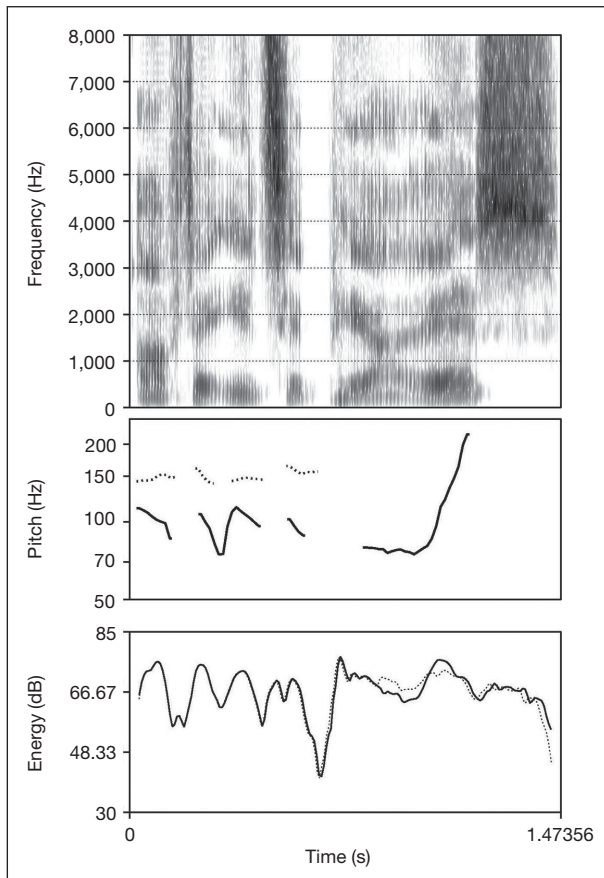


Fig. 5. Spectrogram, fundamental frequency, and energy of the modified signal of the original statement *auf der ICPHS*. (fig. 4): *rising f0* in the accented syllable and *high-pitch final fricative* were transferred from the original question. In the *f0* and the energy window the dotted and plain lines compare *precursor* and *energy*, respectively, from the question and statement realizations. The low precursor *f0* trace is a little different from the one in figure 4 after the resynthesis of the original statement utterance with the valley contour and the subsequent reanalysis for this graph. Online supplementary audios 12–15.

were systematically varied: 2 pitch accent contours \times 2 precursors \times 2 [s] \times 2 energy courses in the pitch accent. An abrupt energy rise was replicated in the statement, a smoother rise in the question. Figures 6a and b show the spectral sections (512 FFT points at 16 kHz sampling rate, bandwidth 50 Hz) 100 ms into the statement and the question [s] ([s] durations 291 and 281 ms) of figures 4 and 5, respectively. In low-pitched [s], there is a stepwise increase in spectral energy from 45 dB between 3 and 4 kHz to 52 dB between 4 and 5 kHz. In high-pitched [s], there is a sharp rise from 30 dB at 3 kHz to a spectral peak of 58 dB at about 4 kHz. This is followed by a slower energy drop in the higher spectrum to a higher level of 35 dB at 8 kHz, compared with the

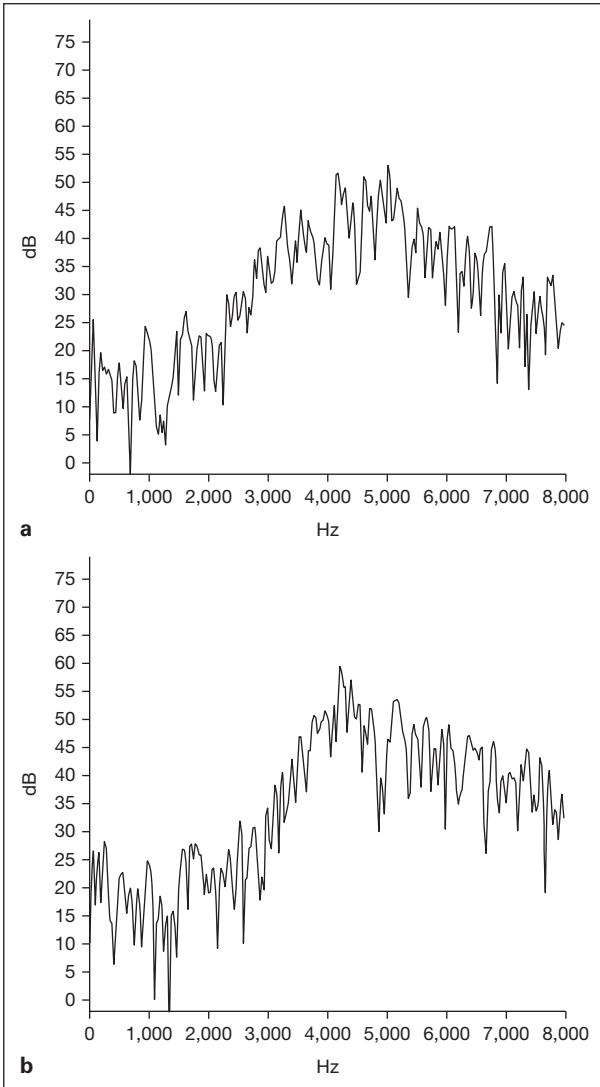


Fig. 6. Narrow-band spectral section 100 ms into the final [s] in the original statement *auf der ICPhS*. (a), and in the original question *auf der ICPhS?* (b), produced by the author and analysed in the Kiel IPDS software xassp with 512 FFT points at 16 kHz sampling rate, bandwidth 50 Hz.

low-pitched [s]. Thus, low-pitched [s] has higher energy in a lower part of the spectrum between about 2.8 and 3.5 kHz, high-pitched [s], on the other hand, has higher energy in the upper part of the spectrum above 4 kHz. Furthermore, the spectral peak moves to lower frequencies across the fricative in low-pitch [s], whereas its position is quite stable in the high-pitched [s]. This is clearly visible in the comparison of the two [s]

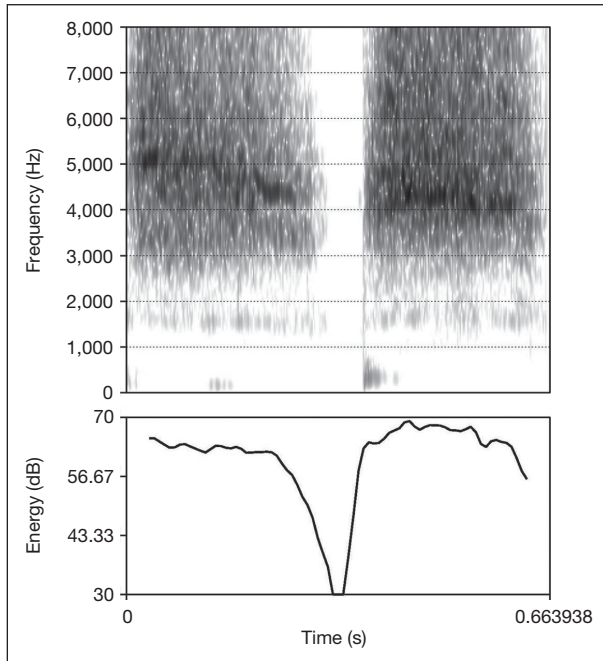


Fig. 7. Spectrograms of the final [s] in the original statement (left) and the original question (right) *auf der ICPHS* in the upper panel, energy trace across the two sounds in the lower panel. Online supplementary audio 16.

spectrograms in figure 7, and finally, the overall energy level of the fricative is higher in the high-pitched [s] (fig. 7).

2.3. Semantic Differential Framework

The investigation of the segment-prosody interrelation proceeds by subordinating the analysis to speech function as the tertium comparationis, which means that it not only tries to establish segmental and prosodic convergence but relates it to functional variables. For this reason, the analysis is carried out in the framework of the Semantic Differential [Osgood et al., 1957], adapting the technique applied in Kohler [2005]. Seven-point scales are used to capture the semantic space covered by the four acoustic variables *f0 direction*, *f0 precursor*, *fricative spectrum*, *energy time course parallel to f0 pattern*. As in Kohler [2005], scales are selected in four semantic and pragmatic domains that can be assumed to be part of the functional frame of any speech communication. In order to keep the very demanding test short only one scale per domain was chosen this time. The choice of scales reflects the semantic and pragmatic description of the German utterances in 2.1: (1) speaker acting in dialogue: *fragend – feststellend* (questioning – asserting); (2) expressiveness of the speaker: *erregt – ruhig* (excited – calm); (3) weighting of argumentation: *mit Nachdruck – ohne Nachdruck* (forceful – not forceful); (4) relationship between speaker and hearer: *gegensätzlich – einvernehmlich* (contrary – agreeable).

2.4. Hypotheses for a Perception Experiment

As the goal of this paper is to investigate the relation between the four signal parameters of 2.2 and the four semantic scales of 2.3 the perception experiment was designed and run on the basis of the following hypotheses.

Hypothesis 1. It can be assumed from the German data description in 2.1 and in Kohler [2006a, b] that, for an elliptic phrase without a formal syntactic interrogative indicator to signal a question, a final high f₀ rise is necessary. On the other hand, a limited final f₀ drop from a less high level must be sufficient before a voiceless fricative to signal a statement. Hypothesis 1 postulates that high-rising valleys are decoded as *questioning*, late peaks as *asserting*, although the pitch drop before the voiceless fricative is minimal.

Hypothesis 2. The German data description in 2.1 shows that by default production, contrastive questions expressing incredulity combine high-rising late concave valleys with high precursors, and affectively contrastive statements expressing unexpectedness combine late peaks with low precursors. In either case, the distance between the initial and the final f₀ level in the utterance is reduced. When the precursors are swapped this distance is increased: the appellative high-rising pitch in the listener-oriented question [Ohala, 1983, 1984] starts at low-level f₀, and the expressive low-falling pitch, across vowel and fricative, in the speaker-oriented assertion starts at high-level f₀. This leads to the postulate of hypothesis 2 that exchanging the precursors strengthens the *contrary* function between speaker and hearer in both question and assertion.

Hypothesis 3. The German data description in 2.1 shows that the spectral composition of the final [s] in [itsepeha'es] contributes to the signalling of a contrastive late peak or a late valley: in the valley contour it has higher overall energy and energy concentration higher in the spectrum, sounding higher-pitched, thus continuing the rising pattern of the vowel, whereas in the peak contour its lower spectral energy and resulting lower pitch continue the falling pattern. Hypothesis 3 postulates that exchanging the final [s] in the original late-peak and late-valley utterances reduces the *asserting* function of the late peak and the *questioning* function of the late valley, and increases the *forceful* function in the late peak.

Hypothesis 4. The German data description in 2.1 shows that a late peak has a fast rise in energy corresponding to the fast rise of f₀. Thus before a voiceless obstruent, a late-peak f₀ contour differs from a late high-rising valley f₀ contour not only by the presence of a small descent from a lower maximum value, and a different shape (convex vs. concave), of the rise, but also by a more abrupt energy increase. Hypothesis 4 postulates that levelling the energy time course in a late f₀ peak reduces *contrariness* and *forcefulness* associated with it; on the other hand, a late high-rising valley with the energy course of a late peak gains in *forcefulness*.

3. Stimulus Generation and Data Collection

3.1. Stimuli

The stimuli for the perception experiment were generated on the basis of the two utterances *auf der ICPHs* described in 2.2. Their total durations and the durations of their precursors are practically identical: statement 1,480 and 892 ms, question 1,463 and 866 ms. Even the individual segments, including the final [s], differ little in their durations ([s] = 281 ms in the question, 291 ms in the statement). It can therefore be accepted that the two precursors, the final [s], and the total

utterances are perceptually equivalent as regards their duration structures, making it possible to exchange precursor and final [s] in the statement utterance by the counterparts from the question utterance.

The original statement utterance provides the late peak: **po (peak original)** (fig. 4). Its peak maximum is 166 Hz, and there is a drop to 149 Hz (1.9 semitones) over 50 ms. This pitch descent is about twice as large as in the Kiel Corpus example, but the end point is still about an octave above the value the same speaker reached in *Stockholm*. So, the f0 contour is again shaped by the following segment [s]. Moreover, in this case, vocalic breathiness sets in 50 ms before the local friction for [s], causing temporal and dynamic irregularities in the glottal pulses, which the pitch algorithm in Praat is incapable of coping with, so the f0 measurements stop at the end of modal voice.

From **po** a second stimulus **pm (peak modified)** was generated in CoolEdit by levelling the amplitude envelope over [ha'ɛs]: the low amplitudes were raised, the subsequent high ones lowered to create an energy course that rises less abruptly (fig. 4). The result sounds less forceful.

From the 2 peak stimuli **po** and **pm**, 2 valley stimuli **vo** and **vm** were generated in Praat by modifying the f0 course across [ha'ɛs]: the f0 value at the end of modal voicing, before the onset of breathiness leading up to the fricative, was raised to 246 Hz, corresponding to the original question utterance, but the f0 value and time position of the original peak maximum (166 Hz, about 40 ms before the end of modal voicing) were kept as an additional pitch point to generate a valley contour that reaches the same value at the same time in the accented syllable as in the peak contour and replaces the fall of the latter by the additional rise in the same structural time frame. Furthermore, the convex shape of the f0 course was changed to concave by inserting another f0 point of 84 Hz 60 ms after the low point (74 Hz) and 90 ms before the original peak point. f0 points were then connected linearly: **vo**, **vm** (valleys from original and modified peaks, fig. 5).

In each of the resulting 4 stimuli, two [s] versions were used: (a) the final [s] was kept in its lower-pitched statement spectral form (**l**), including the 50 ms of vocalic breathiness before the local friction, and (b) the **l**-[s] friction and 30 ms of the preceding vocalic breathiness were replaced by splicing the higher-pitched spectral form (**h**) from the question utterance, where such vocalic breathiness had not occurred: **plo**, **pho**, **plm**, **phm**, **vlo**, **vho**, **vlm**, **vhm** (fig. 4, 5). This procedure ensured that all instances of the original peak-spectral fricative had 50 ms of vocalic breathiness as a further acoustic attribute, irrespective of falling or rising f0; contrariwise, for all instances of the valley-spectral fricative, the strongest breathiness in the original peak utterance was removed, thus providing further acoustic differentiation, in the peak as well as the valley context.

From these 8 stimuli a further set of 8 was generated by splicing the high precursor (**H**) from the original question utterance in place of the low precursor (**L**): **pLlo**, **pLho**, **pLlm**, **pLhm**, **pHlo**, **pHho**, **pHlm**, **pHhm**, **vLlo**, **vLho**, **vLlm**, **vLhm**, **vHlo**, **vHho**, **vHlm**, **vHhm** (fig. 4, 5).

3.2. Test Design

The experiment tests the following 4 factors, the independent stimulus variables: (1) f0 course in the accented vowel (falling/rising), (2) precursor (low/high f0 level), (3) final [s] (high/low pitch), and (4) energy course in the accented vowel (abruptly raised/levelled). The dependent variables are the judgements on the four 7-point semantic scales with an indecisive centre 0 between two poles and 3 values 1–3 on either side.

3.3. Stimulus File and Questionnaire

The 16 stimuli were copied 4 times and randomized in a test file, formatting it with beeps and pauses. A questionnaire was compiled, consisting of answer sheets in which each of the 16 stimuli was associated with each of the 4 scales, 8 stimuli per page, in the order they appeared in the sound file. The answer sheets were preceded by detailed written instructions of the test procedure and followed by a page for personal information and comments on the difficulty of the task and the naturalness of the stimuli. The instructions explained the contextualization of the test utterances *auf der ICPHS*, which subjects were asked to use as reference for their answers ('Appendix'). Part of the instructions was

also a trial run with 4 stimuli, 1 for each semantic scale. Answers consisted in marking the degree on the semantic scales by ticking the appropriate number between the two poles, which were named at the left and the right of each scale in the ordering given in 2.4.

3.4. Subjects and Test

Twenty-four native speakers of German, students of linguistics at the beginning of a course on prosody, participated in the experiment in groups of up to 8. The sessions took place in a sound-treated seminar room of the Institute of Phonetics and Digital Speech Processing at Kiel University. The questionnaires were distributed, and the instructions were also presented orally from a laptop via loud-speaker, as was the test file. The total experiment took about 30 min: 7 min for the recorded instructions, 13 min for the scaling test, 10 min for questions before, and comments after, the test.

4. Results and Statistics

4.1. Descriptive Data Grouping

The numerical values ticked in the questionnaire scales were transferred to lists with negative or positive signs for the left or right half of each scale, respectively. These discrete numbers are regarded as varying along a continuous scale (because subjects might have returned intermediate values if the test paradigm had provided for them), and the distances between successive judgement levels are treated as equal. The resulting interval scale allows the calculation of means and standard deviations, making it possible to run analyses of variance. But since listeners provided discrete judgements, differences between the means of corresponding opposites of the independent variables must bear a meaningful relationship to this procedural aspect of the test. Therefore a further condition was imposed on the interval scale that the means should differ by at least 1 point ± 0.2 (standard error of the mean) along a semantic scale, thus adding credibility to the use of an interval scale.

Figures 8–11 present the means of judgements by the 24 subjects for each of the 16 stimuli, which differ in the independent variables *pitch direction*, *precursor*, *final fricative*, *energy*, separately for each of the 4 dependent variables of the semantic scales *questioning/asserting*, *excited/calm*, *forceful/not forceful*, *contrary/agreeable*. The discussion will follow the given orderings of dependent and independent variables.

4.1.1. The Dimension 'fragend/feststellend' (Questioning/Asserting) (fig. 8, Scale 1)

- (1) The parameter *pitch direction* (**p** vs. **v**) has a strong influence in this dimension, **v** being *questioning*, **p** *asserting*. All **p** stimuli have means in the *asserting* half of the scale, all **v** stimuli in the *questioning* half. Moreover, corresponding opposites are separated by >3 scale points.
- (2) The parameter *precursor* (**L** vs. **H**) does not cause response differences in this dimension.
- (3) The **p** and **v** sets differ with regard to the influence of the parameter *final fricative* (**l** vs. **h**). Across the **v** set, each **l** is less questioning than the corresponding **h**. Corresponding means differ by ± 0.7 scale points. The opposite does not happen in the **p** set: the assertive character is not influenced by the final fricative. There

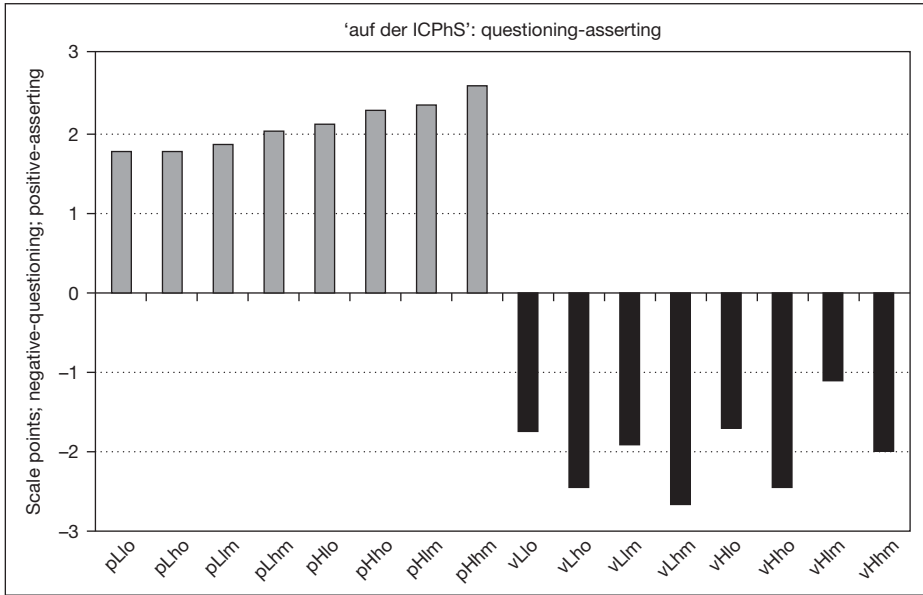


Fig. 8. Means of judgements by the 24 subjects for each of the 16 stimuli on Scale 1 *questioning – asserting*. Stimulus series in the above order in online supplementary audio 17.

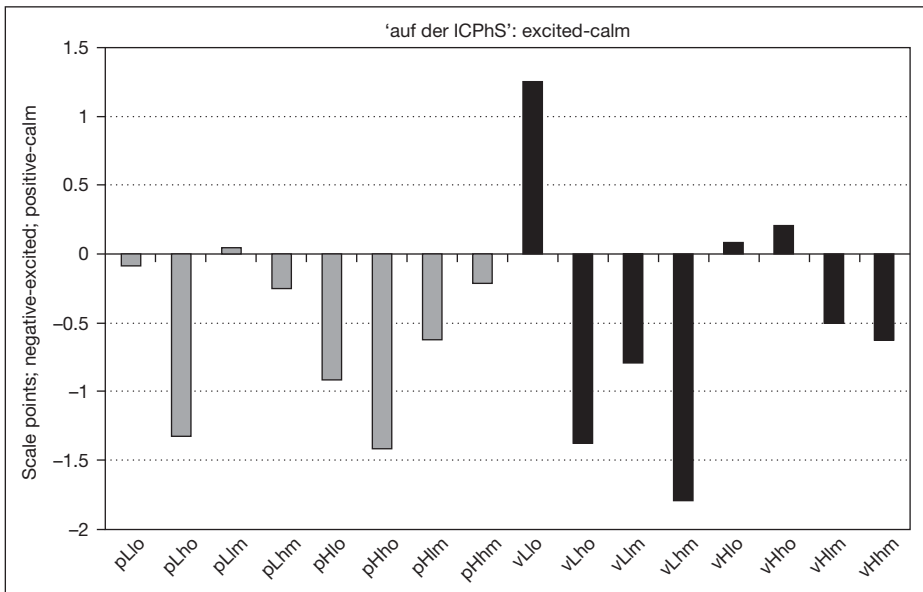


Fig. 9. Means of judgements by the 24 subjects for each of the 16 stimuli on Scale 2 *excited – calm*. Stimulus series in the above order in online supplementary audio 17.

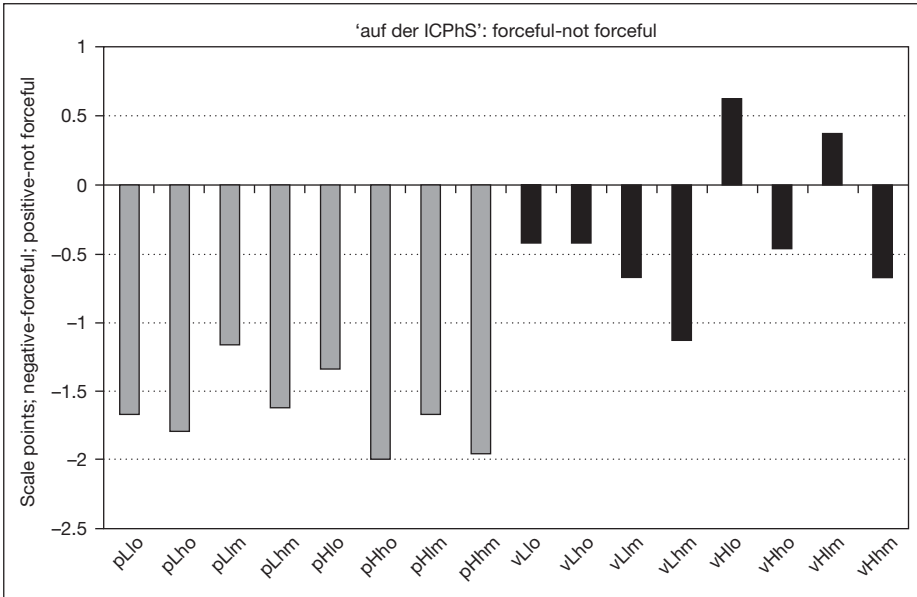


Fig. 10. Means of judgements by the 24 subjects for each of the 16 stimuli on Scale 3 *forceful – not forceful*. Stimulus series in the above order in online supplementary audio 17.

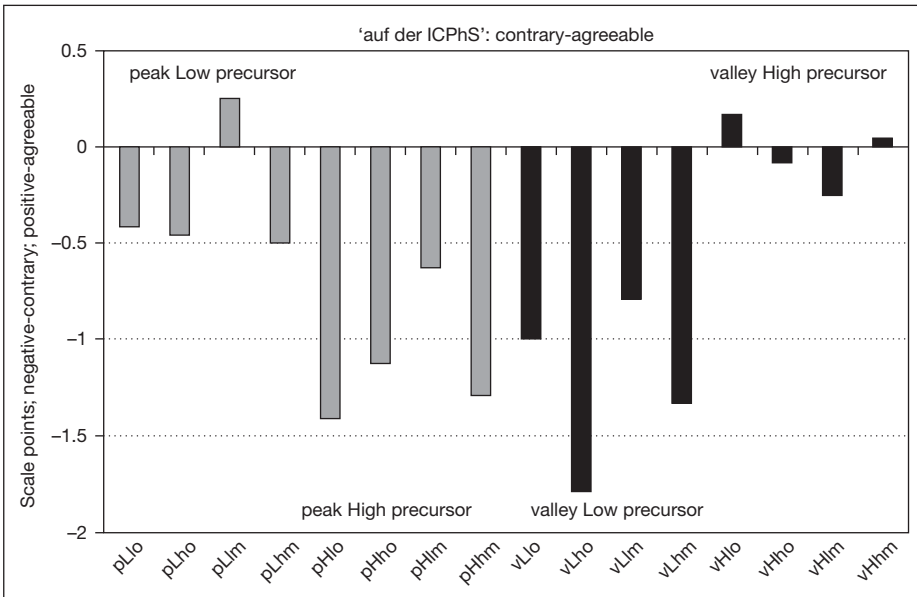


Fig. 11. Means of judgements by the 24 subjects for each of the 16 stimuli on Scale 4 *contrary – agreeable*. Stimulus series in the above order in online supplementary audio 17.

is thus an interdependence of the parameters *pitch direction* and *final fricative*.

- (4) The parameter *energy* (**o** vs. **m**) does not influence responses in this dimension across the data set as a whole.

4.1.2. The Dimension 'erregt/ruhig' (Excited/Calm) (fig. 9, Scale 2)

- (1) In this dimension, there is no clear pattern for the parameter *pitch direction* (**p** vs. **v**) across the complete data set.
- (2) The parameter *precursor* (**L** vs. **H**) does not show a consistent effect across the complete data set either. However, the **v** set differs from the **p** set in two respects: (a) The **H** subset has means around the 0 scale point, whereas the **L** subset shows much larger departures from it. (b) The **L** subset is further structured according to the factor *final fricative*, which is absent from the **H** subset (see (3)).
- (3) The parameter *final fricative* (**l** vs. **h**) differs strongly in this dimension across the **v** set, being more in the direction of *excited* for **h**. This is, however, tied to the combination with initial **L**, an indication of an interdependence between *final fricative*, *pitch direction* and *precursor*, with >1 scale point. There is no corresponding consistent pattern in the **p** set, where only **pLho** differs by >1 scale point from **pLlo** in the direction *excited*.
- (4) The only patterned differences that emerge for the parameter *energy* (**o** vs. **m**) in this dimension are **pho** feature bundles being judged more *excited* than **phm** by >1 scale point, which is again a case of interdependence of parameter features, i.e. *energy* and *pitch direction*, in addition to the influence of *final fricative*.

4.1.3. The Dimension 'mit Nachdruck/ohne Nachdruck' (Forceful/Not Forceful) (fig. 10, Scale 3)

- (1) The parameter *pitch direction* differs in this dimension across the complete data set. As against the **v** set, the **p** means are all located in the *forceful* half of the scale and are generally *more strongly forceful*, with differences between corresponding opposites >1 scale point, except for **pLlm/vLlm** and **pLhm/vLhm** ≈ 0.5 .
- (2) No clear patterns emerge in this dimension for the parameter *precursor* (**L** vs. **H**).
- (3) The parameter *final fricative* (**l** vs. **h**) differs in this dimension across the **p** set, being more strongly *forceful* for **h**, but staying well below 1 scale point in all corresponding opposites. In the **v** set, the distances exceed 1 scale point in combination with *high precursor*, but the absolute scale values are low on both sides of the 0 point. This shows yet another interdependence of parameter features: *pitch direction*, *final fricative*, *precursor*.
- (4) There is no patterned difference in this dimension for the parameter *energy* (**o** vs. **m**).

4.1.4. The Dimension 'gegensätzlich/einvernehmlich' (Contrary/Agreeable) (fig. 11, Scale 4)

- (1) The parameter *pitch direction* differs in this dimension across the complete data set, **p** being more *contrary* in combination with initial **H** than **L**, and vice versa for **v**. **pLlo** and **pHlo** differ by about 1 scale point, the other pairings in the **p** set by 0.7–0.8. In the **v** set, the difference is >1 for all but **vLlm** and **vHlm**.
- (2) The parameter *precursor* (**L** vs. **H**) does not show any patterned differences in this dimension by itself, but there are the interdependences with *pitch direction* in (1).

- (3) No pattern emerges for the parameter *final fricative* in the **p** set; in the **v** set, there is a tendency for final **h** to be judged more contrary in combination with initial **L**, but the distance is only 0.54 scale points for **vLlm/vLhm** (0.8 for **vLlo/vLho**).
- (4) There is no patterned difference in this dimension for the parameter *energy* (**o** vs. **m**).

Both on this Scale and on Scale 2, there are response averages close to 0. They reflect listeners' judgements closely around the indecision point 0 of the scale because the variability in terms of standard deviations is of the same order of magnitude as for the response averages that deviate clearly from 0.

4.2. Inferential Statistics

4.2.1. Overview MANOVA-ANOVA

A repeated-measures MANOVA was run on the 4 independent variables, each with 2 levels, *direction* (1 peak, 2 valley), *precursor* (1 low, 2 high), *fricative* (1 low, 2 high), *energy* (1 high, 2 low) and 4 dependent semantic scales *S1*, *S2*, *S3*, *S4* for the complete data set, followed by univariate repeated-measures ANOVAs on the 4 independent variables for each scale. The results are summarized in table 1.

In the MANOVA, the factors *dire* and *fric* are highly significant, explaining 96% and 60% of the variance in the respective parameter. The factor *prec* turns out moderately significant, explaining only 37% of the variance. Since the interactions *dire*prec* and *prec*fric* are highly significant, explaining 63% and 72% of the variances, *prec* has to be ignored as a main effect across the complete data set; *ener* is not significant as a main effect, but the interaction *dire × ener* is highly significant, explaining 72% of the variance in this parameter. There are no other significant interactions. These results of the MANOVA justify running 4-factorial univariate repeated-measures ANOVAs on each scale for more finely grained pattern analysis.

4.2.2. Scale 1: 'Questioning/Asserting'

On Scale 1, *dire* is highly significant, explaining 93% of the variance with a 4-scale-point difference between the means of the 2 levels; *prec* is also highly significant, explaining 32% of the variance, but the difference between means is only 0.43, so it should be discarded as a main effect with regard to the imposed boundary condition of at least 1 ± 0.2 points between scale means; *fric* is significant, but again differences of means stay low with 0.44; however, the interaction *dire × fric* is significant, too, at a similar test level and partial eta-squared, with a means difference of 0.91 in the *fric* factor for *valley*. Thus *fric* can also be excluded as a main effect, but needs to be recognized as a significant *dire × fric* interaction. This was already clear from the descriptive grouping in 4.1.1, and it is further strengthened by a significant two-tailed t test of correlated samples on the **vl** and **vh** sets: $t = 5.031$, d.f. = 95, $p = 0.000$; mean (**vl**) = -1.63, mean (**vh**) = -2.45, diff = 0.82. So, the inferential statistics in conjunction with boundary conditions on factor separation provide analytic justification for the data description given in 4.1.1.

4.2.3. Scale 2: 'Excited/Calm'

On Scale 2, only *fric* is significant but the difference of means is again well below 1 scale point. However, there is a highly significant interaction *prec × fric*,

Table 1. F values, significance levels and partial eta-squared in the repeated-measures MANOVA and ANOVAs for the 4 independent variables *dire*, *prec*, *ener*, *fric* and their interactions on the four semantic scales S1–S4

a MANOVA: within-subject effects, hypothesis d.f. = 4, error d.f. = 20

Variable	Wilks' lambda	F	Significance	Partial eta-squared
<i>dire</i>	0.049	97.08	0.000	0.951
<i>prec</i>	0.630	2.935	0.046	0.370
<i>ener</i>	0.789	1.335	0.292	0.211
<i>fric</i>	0.402	7.452	0.001	0.598
<i>dire</i> × <i>prec</i>	0.367	8.620	0.000	0.633
<i>dire</i> × <i>ener</i>	0.280	12.838	0.000	0.720
<i>prec</i> × <i>fric</i>	0.283	12.642	0.000	0.717

No other interactions are significant.

b ANOVAs: univariate tests, Greenhouse-Geisser correction for violated homogeneity of variances, hypothesis d.f. = 1, error d.f. = 23

Variable	Scale	F	Significance	Partial eta-squared	Scale means differences
<i>dire</i>	S1	299.109	0.000	0.929	4
	S2	0.357	0.556	0.015	
	S3	36.637	0.000	0.614	1.3
	S4	0.050	0.824	0.002	
<i>prec</i>	S1	10.601	0.003	0.316	0.43
	S2	0.056	0.815	0.002	
	S3	1.231	0.279	0.051	
	S4	1.017	0.324	0.042	
<i>ener</i>	not significant in S1–S4				
<i>fric</i>	S1	6.914	0.015	0.231	0.44
	S2	15.056	0.001	0.396	0.66
	S3	11.390	0.003	0.331	0.52
	S4	5.346	0.030	0.189	0.3
<i>dire</i> × <i>prec</i>	S1	0.116	0.736	0.005	
	S2	5.518	0.028	0.193	1_1-2 0.39 negative increase 2_1-2 0.47 negative decrease
	S3	6.058	0.022	0.208	1_1-2 0.18 negative increase 2_1-2 0.63 negative decrease
	S4	28.279	0.000	0.551	1_1-2 0.83 negative increase 2_1-2 1.2 negative decrease
<i>dire</i> × <i>ener</i>	S1	0.133	0.718	0.006	
	S2	37.864	0.000	0.622	1_1-2 0.68 negative decrease 2_1-2 0.97 negative increase
	S3	2.489	0.128	0.098	
	S4	0.412	0.527	0.018	
<i>dire</i> × <i>fric</i>	S1	6.616	0.017	0.223	1_1-2 0.41 2_1-2 0.91
	S2	2.349	0.139	0.093	
	S3	0.828	0.372	0.035	
	S4	0.009	0.923	0.000	
<i>prec</i> × <i>fric</i>	S1	1.009	0.326	0.042	
	S2	16.361	0.001	0.416	1_1-2 1.29 2_1-2 0.02

Table 1. Continued

Variable	Scale	F	Significance	Partial eta-squared	Scale means differences
<i>dire</i> × <i>prec</i> × <i>fric</i>	S3	2.404	0.135	0.095	
	S4	2.877	0.103	0.111	
	S1	0.610	0.443	0.026	
	S2	5.317	0.030	0.188	1_1_1-2 0.77 2_1_1-2 1.81 1_2_+ 2_2_ pairs no diff
<i>ener</i> × <i>fric</i>	S3	1.838	0.188	0.074	
	S4	0.830	0.372	0.035	
	S1	0.452	0.508	0.019	
	S2	10.747	0.003	0.318	1_1-2 1.06 2_1-2 0.25
	S3	0.248	0.624	0.011	
	S4	0.735	0.400	0.031	

No other interactions are significant on any scale.

which explains 42% of the variance (>39% of *fric*) and shows a means difference of 1.29 in the *fric* factor for *low precursor*. The relevance of this significant interaction for data interpretation is further supported by a significant two-tailed t test of correlated samples on the **vLl** and **vLh** sets: $t = 7.045$, $d.f. = 47$, $p = 0.000$; mean (**vLl**) = 0.23, mean (**vLh**) = -1.58, $diff = 1.81$. This points to an interdependence of *prec* × *fric* with *dire*, which shows up in a significant threefold interaction *dire* × *prec* × *fric*.

Although *ener* is not significant as a main effect, the interaction *ener* × *fric* is, explaining 32% of the variance, with a means difference of 1.06 in the factor *fric* for *high energy*. This result is further strengthened by a significant two-tailed t test of correlated samples on the **pho** and **phm** sets, i.e. data that differ by a combination of *energy* and *fricative* levels in a *peak* context: $t = 5.555$, $d.f. = 47$, $p = 0.000$; mean (**phm**) = -0.23, mean (**pho**) = -1.38, $diff = 1.15$. There is thus also an interdependence of *energy* with *direction*, showing up in another highly significant interaction of *dire* × *ener*, which explains 62% of the variance. The means differences are 0.97 negative increase for *valley* and 0.68 negative decrease for *peak* in the factor *ener*, showing opposed bindings of *ener* with *peak* and *valley*.

The results on Scale 2 are less straightforward than those on Scale 1. But what shows up very clearly is a significant interdependence between *pitch direction* – **v(alley)**, *final fricative* – **h(igh)**, and *precursor* – **L(ow)** to signal the semantic dimension *excited*. There is also a significant interdependence between *energy* – **o (high)** and *final fricative* – **h(igh)** to signal the semantic dimension *excited*, tied to *pitch direction* – **p(eak)**. The inferential statistics support the descriptive data grouping in 4.1.2.

4.2.4. Scale 3: 'Forceful/Not Forceful'

On Scale 3, *dire* is highly significant, explaining 61% of the variance, with a 1.3 scale point difference between factor means. The other significant factor, *fric*, stays

well below 1 scale point between means, but still explains 33% of the variance. Yet the descriptive statistics of 4.1.3 yield clear differences between the **vl** and **vh** sets, and more specifically between the subsets **vHl** and **vHh**. In both cases, two-tailed t tests of correlated samples return significance: $t = 2.78$, d.f. = 95, $p = 0.007$, mean (vl) = -0.02 , mean (vh) = -0.67 , diff = 0.65; $t = 3.396$, d.f. = 47, $p = 0.001$, mean (vHl) = 0.5, mean (vHh) = -0.56 , diff = 1.06. In the **vL** set, however, the *fric* feature is not significant. But there is a consistent significant trend across the **p** set: $t = 2.455$, d.f. = 95, $p = 0.016$, mean (pl) = -1.46 , mean (ph) = -1.84 , diff = 0.38. There is thus a strong *fric* effect interdependent with *valley* and *high precursor*, tying in with the significant interaction *dire* × *prec*, which explains 21% of the variance, and there is a weaker effect interdependent with *peak*, with much smaller differences of means. But since the pattern is consistent it is safe to recognize it as a factor in the data. The inferential statistics thus strengthen the descriptive data grouping of 4.1.3.

4.2.5. Scale 4: ‘Contrary/Agreeable’

On Scale 4, only *fric* is significant. Yet its effect strength is low, as it only explains 19% of the variance, and the difference between means again stays well below 1 scale point. It should therefore be discarded as a main effect. But the descriptive statistics of 4.1.4 show a clear difference between the **vLl** and **vLh** sets, i.e. for *fric* in combination with *valley* and *low precursor*, and a two-tailed t test of correlated samples returns this as significant: $t = 2.33$, d.f. = 47, $p = 0.024$; mean (vLl) = -0.9 , mean (vLh) = -1.56 , diff = 0.66, notwithstanding a scale means difference that stays below the set threshold. However, since Cohen’s [1988] d

$$d = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{(s_1^2 + s_2^2)/2}} \text{ with means diff} = 0.66, s_1^2 = 3.41, s_2^2 = 2.08$$

yields a medium effect strength of 0.4 for this difference, there is justification in recognizing a *fric* effect in interdependence with *valley* and *low precursor*.

The interaction *dire* × *prec* is highly significant, explaining 55% of the variance; the means differences are 0.83 negative increase in the factor *prec* for *peak* and 1.2 negative decrease for *valley*, a case of opposed bindings of *prec* with *peak* and *valley*, which is further supported by two-tailed t tests of correlated samples of the **pL** versus **pH** and **vL** versus **vH** sets: $t = 3.68$, d.f. = 95, $p = 0.000$; mean (pL) = -0.28 , mean (pH) = -1.11 ; $t = 5.14$, d.f. = 95, $p = 0.000$; mean (vL) = -1.23 , mean (vH) = -0.03 . The inferential statistics results thus buttress the descriptive grouping in 4.1.4.

5. Interpretation of Results in Relation to Hypotheses

On Scale 1, *pitch direction* has the greatest influence on the dimension *questioning* for **v** and *asserting* for **p**. The *high-rising valley* to a value half an octave above the maximum peak value in the falling pattern is crucial in an elliptic syntactic construction that functions as a question: it may be regarded as its prototypical feature. *Final fricative* reduces *questioning* in the **v** set, when [s] is low-pitch, compared with high-pitch in the original question production. This is the effect of overall lower versus higher energy and of energy concentration lower versus higher in the spectrum of the final [s]; duration differences are negligible. The greater acoustic weight of the higher

fricative pitch functions as the prototypical manifestation of the fricative in a high-rising question, where tonal pitch is thus continued in the pitch of the final noise. On the other hand, low pitch would be the prototypical manifestation of the fricative in a falling statement, where tonal pitch is again continued in the pitch of the final noise, and where, furthermore, the terminal utterance leads to vocal breathiness in the transition from the vowel to the fricative.

On Scale 2, none of the parameters have an influence on the dimension *excited/calm* just by themselves, but *rising pitch*, *high-pitch final fricative* and *low-pitch precursor* combine in raising responses in the direction of *excited*; the first two variables strengthen high pitch of the accented syllable and its association with *questioning* (4.1.1, 4.2.2), the third parameter feature establishes a contrast between a low precursor and a high-pitched accented syllable, which, in turn, heightens the perception of *excitement* in *questioning*. Contrariwise, *falling pitch*, *high energy* and *high-pitch final fricative* also combine in raising responses in the direction of *excited*: *falling pitch* triggers *asserting* (4.1.1, 4.2.2), and its combination with the other two parameter features heightens the perception of *excitement* in *asserting*.

On Scale 3, *falling pitches* are judged more *forceful* than *rising* ones. There is a tendency of *forcefulness* increasing in *falling pitch* in combination with *high-pitch final fricative*; this reflects their combined effect on higher prominence, which is in turn judged more *forceful*. In *rising pitch*, *final fricative* must be combined with *high precursor* to produce an effect, but it stays at a small absolute magnitude on the scale around the indecision point, since the main effect of *rising pitch* is away from *forcefulness*. In the prototypical high-rising question with *high precursor*, *low-pitch final fricative* reduces *forcefulness* still further. On the other hand, a *low precursor* increases the distance between utterance-initial and -final pitch levels, which works against a reduction in *forcefulness*.

On Scale 4, none of the four parameters have an effect on the dimension *contrary/agreeable* on their own, but there is an interdependence between *pitch direction* and *precursor*: *falling pitch* is judged more *contrary* when it is introduced by a *high precursor* than by a *low precursor*, whereas for *rising pitch* it is the other way round. In both cases the pitch contrast between the beginning and the end of the utterance is widened. This result converges with the findings of Grabe et al. [1997], who carried out a Semantic Differential investigation into the combinations of Dutch high %H or low %L preheads with immediately following H*L and L*H pitch accents, which correspond to medial peak and high-rising valley categorizations used in this paper. Judgements on the 'Friendliness', 'Politeness', and 'Irritation' scales were affected by agreement or disagreement of the prehead tone with the initial pitch accent tone. Polar sequences of the two tones yielded more positive (friendlier, more polite, less irritated) judgements on the scales, non-polar sequences more negative ones. The 'Friendliness', 'Politeness', and 'Irritation' scales refer to the relationship between speaker and hearer in communicative interaction, as does the *contrary – agreeable* scale in the German data. However, in the latter case, the collocation of a high prehead with a late peak increases *contrary* judgements in spite of a polar-tone succession. This indicates that it is not the equivalence or divergence of the initial pitch accent tone and the preceding prehead tone that is relevant in perception and cognitive processing but the relation between the utterance-initial and the utterance-final pitch level, creating either a narrow or a wide range. The wide range intensifies the contrastive assertion of the late peak and the incredulous question of the late

concave high-rising valley, in both cases conveying an attitude of contrariness in the speaker-listener relationship. This semantic interpretation of a global utterance contour presupposes the recognition of a much wider processing window in speech perception than the postulate of intonational morphemes for prehead and nuclear tones allows [Grabe et al., 1997].

The results can now be related to the initial hypotheses.

Hypothesis 1 has been confirmed: high-rising valleys are decoded as *questioning*, late peaks as *asserting*, although the pitch drop before the voiceless fricative is minimal.

Hypothesis 2 has been confirmed: increasing the distance between the utterance-initial and -final tonal levels by either lowering the high precursor to the late valley or by raising the low precursor to the late peak strengthens the *contrary* function.

Hypothesis 3 has been partly confirmed: *low fricative* reduces the *questioning* function in the late valley, but *high fricative* does not reduce the *asserting* function of the late peak; it does, however, increase the *forceful* function in the late peak.

Hypothesis 4 has not been confirmed: levelling the energy time course in a late f0 peak does not reduce *contrariness* and *forcefulness*, nor does a late high-rising valley with the energy course of a late peak gain in *forcefulness*. But there seems to be a weak effect of *energy* in the context of *peak + high fricative*: *high energy* strengthens the impression of *excitedness*.

Over and above confirming or partially confirming the initial hypotheses 1, 2, 3, the results show that the four independent variables function on several semantic scales simultaneously, adding shades of meaning to a prototypical core in interaction with each other. Although high-rising and falling pitch explain most of the variance by prototypically differentiating questions and statements in elliptical phrases, pitch direction also functions on the *forceful* scale, late peaks being intrinsically more *forceful* than valleys. Moreover, peaks go together with *low precursor* to convey statements, valleys with *high precursor* to convey questions. If this prototypical binding is reversed, *contrariness* is superimposed on the basic statement or question function.

High-pitched fricative is part of the prototypical question function in addition to *high-rising valley*, since the question feature diminishes with a *low-pitched fricative*. The fricative variable further functions on the *forceful* scale, where *high-pitched* fricatives contribute to *forceful* in varying gradation depending on the combination with *pitch direction*. Finally, it also functions on the *contrary* and the *excited* scales. The *contrariness* of the non-prototypical link of valley with *low precursor* in questions is attenuated by *low-pitched fricative* and heightened by *high-pitched fricative*. Similarly, *low-pitched fricative* signals more towards *calm*, *high-pitched fricative* towards *excited*, in the non-prototypical link of valley with *low precursor* in questions.

The signal variables *pitch direction*, *final fricative* and *precursor* thus form an interdependent net for coding the semantics of statements and questions by adding components of meaning in the *excitedness*, *forcefulness* and *contrariness* dimensions to the core differentiation between *contrastive asserting* and *questioning* in late peaks and high-rising valleys of elliptic constructions.

The fourth variable, *energy*, shows a very small influence in these experimental data, restricted to *high* versus *low energy* in combination with *late peak + high final fricative* for signalling strongly *excited* as against indecisive. This result is surprising since the raised energy in the original statement is clearly audible and interpretable

as a more *forceful contrastive* statement. The reason for the weak effect is probably a combination of influences.

On the one hand, the difference between raised and reduced energy was perhaps not generated satisfactorily, on the other hand, the combinations of *late peak* with *high-pitch final fricative* or *high precursor* may have been stronger signals of *forceful* or *contrary*, respectively, for the listeners in this experimental task.

Generally, the listeners may have had problems with the semantic scales, more particularly with their naming. They were asked after the test which scale they found the easiest, which the most difficult. For 23 subjects, the easiest scale was questioning/asserting, for 1 subject excited/calm. For 15 subjects, the most difficult scale was contrary/agreeable, for 4 subjects forceful/not forceful, and for 5 subjects excited/calm.

The ease of the *quest/assert* scale is reflected in the clear-cut results of the *dire* factor. This contrasts with the much weaker effects of any single factor on the *contr/agree* scale, although the strong interaction of *dire* × *prec* indicates that, contrary to the listeners' subjective impressions, they produced very systematic profiles.

The subjects were also asked after the test to judge how natural the test utterances sounded, on a 7-point scale, ranging from -3 (most unnatural) through 0 to +3 (most natural). Their answers were: -2 (2 subjects), -1 (5 subjects), 1 (4 subjects), 2 (12 subjects), 3 (1 subject). So, naturalness was quite well achieved; therefore, the sound quality most likely did not have an adverse effect in the case of a weak acoustic property.

6. Conclusion and Outlook

This paper provides results of an investigation into the sound-sense relationship in the semantic field of questions and statements. Systematic structures relating four phonetic parameters to four semantic scales have emerged. The strongest effects are related to the prosodic feature of pitch direction and the segmental feature of the final fricative spectrum following falling or rising f_0 . These two features from either side of the prosody-segment divide are mutually dependent in their combined signalling power in all four semantic functions. The interdependence works in two directions:

(a) The fricative masks a substantial part of the f_0 fall, which can be reconstructed by the listener from the speaker's base-level f_0 , whereas the f_0 rise needs to be fully executed to signal the more freely variable, and thus not reconstructible, maximum f_0 level to the listener.

(b) The fricative continues the tonal pitch direction in its spectral energy distribution, low after falling, high after rising f_0 . Thereby it assists the listener in the reconstruction of an f_0 fall, and heightens the question function in an f_0 rise. Exchanging the two spectral fricative versions in the falling or rising f_0 pattern reduces the question function of an f_0 rise, and adds semantic components on the other three scales in both f_0 falls and rises.

Thus, segments shape prosodies and are shaped by them in varying ways in the coding of semantic functions. This implies that the analysis of sentence prosodies needs to integrate the manifestation of segments. Moreover, the data of this paper point to an even wider prosodic window than the f_0 contour in the accented syllable. The *high* or

low precursor modifies the communicative function of the whole utterance, depending on whether it increases the distance between the initial and final utterance pitches in *high precursor + fall* or *low precursor + rise*, as against the smaller intervals in the reversed combinations. In these constellations listeners obviously react to the utterance as a whole, which tells us that the short-time prosodic windows of falling or rising pitch accents and high or low precursors need to be incorporated in a more global time window [Grabe et al., 1997].

That is a further indication that the analysis of pitch contours should not be reduced to sequences of local tone levels as the ultimate organization of pitch prosody. In communicative interaction, speakers set broader prosodic time windows of varying sizes, and listeners respond to them. This calls to mind the European tradition of intonation analysis, first and foremost the London School of Phonetics and its function-oriented modifications of basic global tunes in English [O'Connor and Arnold, 1961], as well as Grønnum's [1992] 'sentence intonation contour slopes varying with sentence function' in Danish, Gårding's [1979, 1982] functional generative model of Swedish accents and sentence intonation, and Bruce's [1977] word accents in phrase accent frames of focus and sentence mode. So, future phonetic research can greatly benefit from bringing this tradition back into the limelight and from bridging the prosody-segment divide by analysing speech in time windows whose extensions are determined by the demands of communicative functions instead of by analysis paradigms.

Acknowledgement

The author gratefully acknowledges the advice on statistical methodology and analysis which he received from Ernst Dombrowski of the Department of Psychology at Kiel University, and the helpful comments on a previous draft of this paper from Bill Barry, as well as insightful suggestions for revision from two anonymous reviewers.

Appendix

Excerpt from the Instructions explaining the contextualization for the test utterances *auf der ICPHS*, which subjects were asked to use as reference for their answers.

Um die Aufgabe so natürlich wie möglich zu gestalten, wird sie in folgenden Situationskontext eingebettet:

1995 fand in Stockholm der Internationale Phonetikerkongress statt, *The International Congress of Phonetic Sciences* – die ICPHS(-Tagung). Zwei deutsche Phonetiker, aus Kiel und Saarbrücken, unterhalten sich einige Zeit vorher telefonisch darüber, wann sie sich mal wieder treffen.

A. Wir haben uns lange nicht gesehen. Wir müssen uns unbedingt mal wieder treffen.

B. Bist du in Stockholm? **Auf der ICPHS?**

B. Ich schlage vor, wir treffen uns in Stockholm. **Auf der ICPHS.**

Wenn Sie sich in diese Situation versetzen, dann wird es für Sie bestimmt recht einfach sein, die Melodie der Äußerungen **Auf der ICPHS?** bzw. **Auf der ICPHS.** zwischen den Polen der Bedeutungsskala «*fragend* – *feststellend*» einzuordnen. Die erste beurteilen Sie sicher als stark «*fragend*» und würden deshalb das linke Kästchen mit der 3 ankreuzen, während Sie die zweite am Pol «*feststellend*» ansiedeln. Feststellungen und Fragen können nun aber in sich selbst wieder mit vielen Nuancen melodisch differenziert werden. Zum Beispiel in folgenden Reaktionen von B:

B: Bist du nicht in Stockholm? **Auf der ICPHS?**

B: Wir treffen uns doch in Stockholm. **Auf der ICPHS.**

Das heißt, die Sprechmelodie muss noch auf anderen Skalen beurteilt werden, die sich auf Nachdruck, Gegensatz, Lebhaftigkeit beziehen. Im Hörexperiment werden daher vier verschiedene Bedeutungsskalen verwendet, nämlich außer der bereits genannten noch «*lebhaft – ruhig*», «*mit Nachdruck – ohne Nachdruck*» und «*gegensätzlich – verbindlich*».

Sie werden in diesem Experiment immer die gleiche Wortfolge **Auf der ICPHS** hören, aber mit wechselnden Melodien. Ihre Aufgabe als Hörer ist es, die Melodie zwischen den Polen der Bedeutungsskala einzuordnen. Dabei sollen Sie sich stets den skizzierten Kommunikationskontext vorstellen.

References

- Abramson, A.; L-Thongkum, Th.: Voice register in Suai (Kuai): an analysis of perceptual and acoustic data. *Phonetica* 61: 147–171 (2004).
- Arvaniti, A.: Rhythm, timing and the timing of rhythm. *Phonetica* 66: 46–63 (2009).
- Arvaniti, A.; Ladd, D.R.; Mennen, I.: Stability of tonal alignment: the case of Greek prenuclear accents. *J. Phonet.* 26: 3–25 (1998).
- Barry, W.; Andreeva, B.; Koreman, J.: Do rhythm measures reflect perceived rhythm? *Phonetica* 66: 78–94 (2009).
- Baumann, S.; Becker, J.; Grice, M.; Mücke, D.: Tonal and articulatory marking of focus in German. Proc. 16th ICPHS, Saarbrücken 2007, pp. 1029–1032.
- Bergem, D. van: Acoustic vowel reduction as a function of sentence accent, word stress, and word class on the quality of vowels. *Speech Commun.* 12: 1–23 (1993).
- Bruce, G.: Swedish word accents in sentence perspective (CWK Gleerup, Lund 1977).
- Cohen, J.: *Statistical power analysis for the behavioral sciences*; 2nd ed. (Erlbaum, Hillsdale 1988).
- Cumming, R.: The interdependence of tonal and durational cues in the perception of rhythmic groups. *Phonetica* 67: 219–242 (2010).
- Cumming, R.: The language-specific interdependence of tonal and durational cues in perceived rhythmicity. *Phonetica* 68: 1–25 (2011).
- Firth, J.R.: Sounds and prosodies. *Trans. Philological Soc.* 1948: 127–152 (1948).
- Fry, D.B.: Duration and intensity as physical correlates of linguistic stress. *J. acoust. Soc. Am.* 27: 765–768 (1955).
- Fry, D.B.: Experiments in the perception of stress. *Lang. Speech* 1: 126–152 (1958).
- Fry, D.B.: The dependence of stress judgments on vowel formant structure; in Zwirner, Bethge, Proc. 6th Int. Congr. Phonetic Sci., pp. 306–311 (Karger, Basel 1965).
- Gårding, E.: Sentence intonation in Swedish. *Phonetica* 36: 207–215 (1979).
- Gårding, E.: Swedish prosody: summary of a project. *Phonetica* 39: 288–301 (1982).
- Gartenberg, R.; Panzlaff-Reuter, C.: Production and perception of F0 peak patterns in German. *Arbeitsber. Inst. Phonetik Univ. Kiel (AIPUK)* 25: 29–113 (1991).
- Grabe, E.: *Comparative intonational phonology: English and German*; PhD thesis University of Nijmegen (1998).
- Grabe, E.; Gussenhoven, C.; Haan, J.; Marsi, E.; Post, B.: Preaccidental pitch and speaker attitude in Dutch. *Lang. Speech* 41: 63–85 (1997).
- Grønnum, N.: *The groundworks of Danish intonation – an introduction*. Museum Tusulanum Press, Copenhagen 1992.
- Hawkins, S.: Roles and representations of systematic fine phonetic detail in speech understanding. *J. Phonet.* 31: 373–405 (2003).
- Hawkins, S.; Smith, R.: Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Ital. J. Ling.* 13: 99–188 (2001).
- Honikman, B.: *Articulatory settings*; in Abercrombie, Fry, MacCarthy, Scott, Trim, In honour of Daniel Jones, pp. 73–84 (Longman, London 1964).
- Huggins, A.W.F.: Just noticeable differences for segment duration in natural speech. *J. acoust. Soc.* 51: 1270–1278 (1972).
- IPDS: *The Kiel corpus of spontaneous speech*, vol. 1, 2, CD-ROM No. 2, 3 (IPDS, Kiel 1995/1996).
- Isačenko, A.V.; Schädlich, H.-J.: *Untersuchungen über die deutsche Satzintonation*. *Studia Grammatica* VII, pp. 7–67 (Akademie-Verlag, Berlin 1966).
- Kohler, K.J.: *Modelling prosody in spontaneous speech*; in Sagisaka, Campbell, Higuchi, *Computing prosody. Computational models for processing spontaneous speech*, pp. 187–210 (Springer, New York 1996).
- Kohler, K.J.: *Articulatory prosodies in German reduced speech*. Proc. 14th ICPHS, vol. 1, pp. 89–92, San Francisco 1999.
- Kohler, K.J.: *Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions*; in Fant, Fujisaki, Cao, Xu, *From traditional phonology to modern speech processing. Festschr. for Professor Wu Zongji's 95th Birthday*, pp. 205–214 (Foreign Language Teaching and Research Press, Beijing 2004).
- Kohler, K.J.: *Timing and communicative functions of pitch contours*. *Phonetica* 62: 88–105 (2005).
- Kohler, K.J.: *What is emphasis and how is it coded?* Proc. 3rd Int. Conf. Speech Prosody, Dresden 2006a, pp. 748–751.

- Kohler, K.J.: Paradigms in experimental prosodic analysis: from measurement to function; in Sudhoff, Lenertová, Meyer, Pappert, Augurzyk, Mleinek, Richter, Schließer, *Methods in Empirical Prosody Research*, pp. 123–152 (de Gruyter, Berlin 2006b).
- Kohler, K.J.: Rhythm in speech and language: a new research paradigm. *Phonetica* 66: 29–45 (2009a).
- Kohler, K.J.: Patterns of prosody in the expression of the speaker and the appeal to the listener; in Fant, Fujisaki, Shen, *Frontiers in phonetics and speech science*, pp. 287–302 (The Commercial Press, Beijing 2009b).
- Kohler, K.J.; Niebuhr, O.: The phonetics of emphasis. *Proc. 16th ICPHS, Saarbrücken 2007*, pp. 2145–2148.
- Kohler, K.J.; Niebuhr, O.: On the role of articulatory prosodies in German message decoding. *Phonetica* 68: 57–87 (2011).
- Ladd, D.R.: *Intonational phonology* (CUP, Cambridge 1996).
- Ladd, D.R.; Faulkner, D.; Faulkner, H.; Schepman, A.: Constant ‘segmental anchoring’ of F0 movements under changes in speech rate. *J. acoust. Soc. Am.* 106: 1543–1554 (1999).
- Ladd, D.R.; Mennen, I.; Schepman, A.: Phonological conditioning of peak alignment in rising pitch accents in Dutch. *J. acoust. Soc. Am.* 107: 2685–2696 (2000).
- Ladefoged, P.: *Three areas of experimental phonetics* (Oxford University Press, London 1967).
- Laver, J.: *The phonetic description of voice quality* (Cambridge University Press, Cambridge 1980).
- Lindblom, B.: Explaining phonetic variation: a sketch of the H&H theory; in Hardcastle, Marchal, *Speech production and speech modelling*, pp. 403–439 (Kluwer Academic Publishers, Dordrecht 1990).
- Lindblom, B.; Maddieson, I.: Phonetic universals in consonant system; in Hyman, Li, *Language, speech and mind*, pp. 62–78 (Routledge, London 1988).
- MacNeilage, P.: *The origin of speech* (Oxford University Press, Oxford 2008).
- Mücke, D.; Grice, M.; Becker, J.; Hermes, A.: Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *J. Phonet.* 37: 321–338 (2009).
- Niebuhr, O.: The signalling of German rising-falling intonation categories – the interplay of synchronization, shape, and height. *Phonetica* 64: 174–191 (2007).
- Niebuhr, O.: Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. *J. acoust. Soc. Am.* 124: 1252–1263 (2008).
- Niebuhr, O.: Fundamental frequency-based rhythm effects on the perception of local syllable prominence. *Phonetica* 66: 95–112 (2009a).
- Niebuhr, O.: Intonation segments and segmental intonations. *Proc. 10th Interspeech Conf., Brighton 2009b*, pp. 2435–2438.
- Niebuhr, O.: On the phonetics of intensifying emphasis in German. *Phonetica* 67: 170–198 (2010).
- Niebuhr, O.; Kohler, K.J.: Perception of phonetic detail in the identification of highly reduced words. *J. Phonet.* (in press).
- Nolan, F.; Asu, E.L.: The Pairwise Variability Index and coexisting rhythms in language. *Phonetica* 66: 64–77 (2009).
- O’Connor, J.D.; Arnold, G.F.: *Intonation of colloquial English* (Longmans, London 1961).
- Ohala, J.: Cross-language use of pitch: an ethological view. *Phonetica* 40: 1–18 (1983).
- Ohala, J.: An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica* 41: 1–16 (1984).
- Osgood, C.E.; Suci, G.J.; Tannenbaum, P.H.: *The measurement of meaning* (University of Illinois Press, Urbana 1957).
- Port, R.F.: All is prosody: phones and phonemes are the ghosts of letters. *Proc. 4th Int. Conf. Speech Prosody, Campinas 2008*.
- Sluijter, A.M.C.; van Heuven, V.J.: Effects of focus distribution, pitch accent and lexical stress on the temporal organisation of syllables in Dutch. *Phonetica* 52: 71–89 (1995).
- Sluijter, A.M.C.; van Heuven, V.J.: Spectral balance as an acoustic correlate of linguistic stress. *J. acoust. Soc. Am.* 100: 2471–2485 (1996).
- Sluijter, A.M.C.; van Heuven, V.J.; Pacilly, J.J.A.: Spectral balance as a cue in the perception of linguistic stress. *J. acoust. Soc. Am.* 101: 503–513 (1997).
- Vaihinger, H.: *Die Philosophie des Als Ob* (Meiner, Leipzig 1920). Translated by C.K. Ogden, *The philosophy of ‘as if’* (Routledge & Kegan, London 1924).