Original Paper

Phonetica 2010;67:1–29
DOI: 10.1159/000■■■

Received: July 16, 2010
Accepted: August 18, 2010

On the Phonetics of Intensifying Emphasis in German

Oliver Niebuhr

Seminar für Allgemeine und Vergleichende Sprachwissenschaft,
Christian-Albrechts-Universität zu Kiel, Kiel, Germany

Abstract

This paper presents an exploratory study in the field of emphasis in German. It provides a comprehensive acoustic analysis for a type of emphasis that intensifies lexical meanings either positively or negatively. A speech corpus was recorded using an elicitation method adapted to yield natural-sounding, conversational, expressive speech under controlled conditions. Supporting the distinction between positive and negative intensification, two clearly different phonetic profiles emerged. The phonetic profiles of positive and negative intensification involve voice quality as well as the dynamics of the speech signal across its segmental and prosodic layers. By means of these profiles, the intensifying emphases were correctly classified by a discriminant analysis as positive or negative in around 90% of the cases. Moreover, indications were found for a third type of intensifying emphasis, which was called 'reinforcement'. Its multidimensional phonetic profile falls in between the ones of positive and negative intensification.

1. Introduction

1.1. The Initial Notion of a Non-Pitch Accent

The impetus for the presented research on German emphasis came from a problem that occurred during the prosodic labelling of a subsection of the 'Kiel Corpus of Spontaneous Speech', the 'Lindenstraße' dialogues [Kohler et al., 2006]. Figure 1 displays the word *runterkickt* ('kicks down') in the utterance *wie Boris Valerie die Treppe runterkickt* ('when Boris kicks Valerie down the stairs') in dialogue l061a018 from this corpus. German compound verbs like *runterkickt* have lexical stress on the adverbial particle (*runter-*), and by default, pitch accents at the utterance level are hooked onto these lexically stressed syllables. However, phonetically experienced labellers agreed that in this utterance both the adverbial particle *and* the stem were equal in terms of perceptual prominence and should hence be labelled as accented. As can be seen in figure 1, there is an extensive fall in fundamental frequency (F0) on the initial adverbial

KARGER

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

© 2010 S. Karger AG, Basel
0031–8388/■■/0■■–0■■■
$26.00/0
Accessible online at:
www.karger.com/pho

Oliver Niebuhr
Freyastrasse 10
DE–24939 Flensburg (Germany)
Tel. +49 461 940 3993
E-Mail niebuhr@linguistik.uni-kiel.de

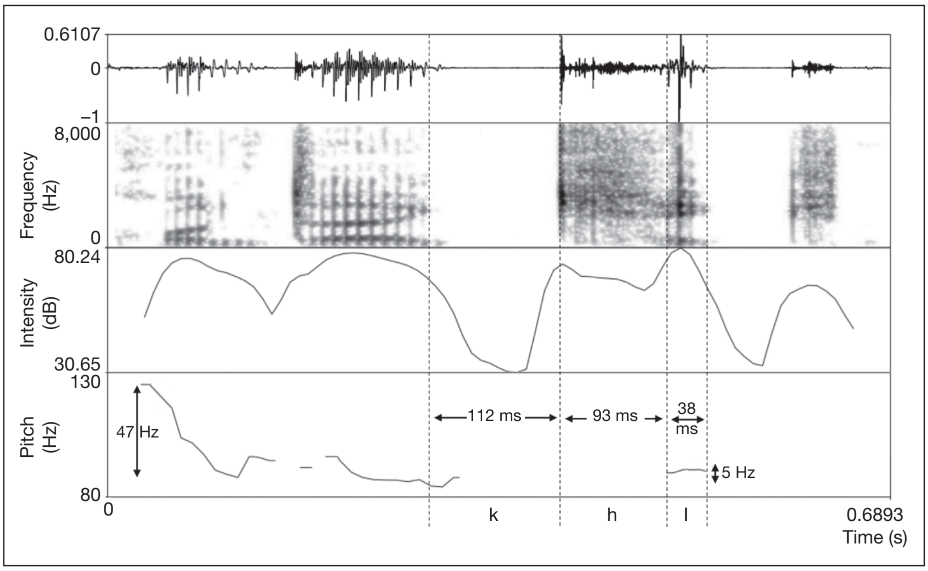

Fig. 1. Oscillogram, spectrogram, acoustic energy course, and F0 course of the word *runterkickt* ('kicks down'), produced with two equally prominent syllables (*run-* and *-kickt*) by speaker 'l06' in the German spontaneous dialogue corpus 'Lindenstraße'. F0 ranges and the consonant and vowel durations in the emphasized syllable are given.

particle *runter-* that allowed a clear identification of a type of pitch accent. The F0 range covered around 50 Hz (130–80 Hz). By contrast, the subsequent stem syllable *-kickt* does not show a similarly rising or falling F0 movement. Rather, F0 remains flat within a narrow range of about 5 Hz. So, in view of the crucial role of F0 in triggering perceptual prominence in German [Andreeva et al., 2007; Kohler, 2008], where does the high prominence of *-kickt* come from?

The discussion of this example ascribed the high perceptual prominence to an increased 'force' of speech production in the consonant-vowel sequence /kɪ/ of *-kickt* [following the terminology of, e.g., Sievers, 1901]. In the case of the syllable-initial voiceless stop /k/ this force manifested itself acoustically in the closure and aspiration phases that showed exceptionally long durations of 112 and 93 ms and 'swallowed up' a good part of the vowel /ɪ/, which only measured 38 ms. Release burst and aspiration of /k/ showed high acoustic energy levels. They almost reached the level of /ɪ/, which has a higher acoustic energy peak than the preceding vowels, despite its short duration. In consequence, the acoustic energy changes to both sides of /ɪ/ happen abruptly. Moreover, the increased 'force' gives the vowel a pressed voice quality that is indicated in figure 1 in the relatively high spectral energy of the upper frequencies (i.e. in the lower spectral tilt). Overall, the long asonorous aspiration phase that leads over to a barked-sounding and low, evenly pitched vowel is in strong syntagmatic contrast to the preceding sonorous and melodic accent, which attracts the attention of the listener and may be seen as the main factor for the raised prominence of *-kickt*. On this basis, the type of accent that was observed on *-kickt* was referred to as a *(non-pitch) force accent*, a different category from pitch accent [Kohler, 2003].

1.2. Emphasis for Contrast vs. Emphasis for Intensity

From the point of view of meaning, the additional prominence on the verb-final stem *-kickt* was interpreted as a kind of emphasis. Following the early phonetic analyses of English and French by Armstrong and Ward [1926] or Coustenoble and Armstrong [1934], two basic types of emphasis can be differentiated, i.e. 'emphasis for contrast' (or 'special prominence') and 'emphasis for intensity'. This differentiation was first made by Coleman [1914].

Emphasis for contrast corresponds to contrastive or corrective focus. It is well investigated across languages: Braun and Ladd [2003], Baumann et al. [2006, 2007] as well as Hermes et al. [2008] for German; Krahmer and Swerts [2001] as well as Hanssen et al. [2008] for Dutch; Cooper et al. [1985] for American English; Xu and Xu [2005] for British English and Mandarin Chinese; Dahan and Bernard [1996], Dohen and Loevenbruck [2004], as well as Dohen et al. [2004] for French; Gili Fivela [2002] for a variety of Italian, or Strangert [2003] for Swedish. Emphasis for contrast is based on adding extra prominence to the accented syllable of the word, while simultaneously suppressing the prominences of the following syllables. This strategy is pursued primarily by means of F0. That is, the syllable with the contrastive accent in the word shows a steep and/or extensive F0 rise that reaches beyond the surrounding F0 peaks. The rise is frequently followed by an immediate sharp fall, hence creating a pointed peak contour. The F0 of the subsequent syllables is lowered and/or flattened. Some studies also report a lowering of the F0 preceding the pointed contrastive accent peak. Over and above F0, contrastive emphasis may go along with a more careful (hyper-) articulation of the emphasized vowel as well as with minor lengthening of both the vowel and preceding consonants.

It is evident that the exponents of contrastive emphasis are almost diametrically opposed to the F0 and duration patterns that characterized the emphasized syllable *-kickt*. The syllable *-kickt* does not even show a rudimentary F0 rise, the vowel /ɪ/ is shortened and suppressed rather than hyperarticulated, the lengthening of the onset consonant cannot count as minor, and there is a high F0 peak preceding *-kickt*. In addition, the semantic-pragmatic context in which the emphasis on *'-kickt'* occurred is not contrastive, but suitable to trigger emphasis for intensity. Emphasis for intensity means that the semantics of the word that bears the emphasis is intensified. Armstrong and Ward [1926] illustrate the intensification concept by means of the utterance *There was an enormous queue waiting at the theatre.* By producing intensifying emphasis on *-nor-'*of *enormous*, speakers can stress the great length of the queue and in this way express either their admiration for the grandiose theatre performance or their irritation about the long waiting time. As to phonetic manifestation, Armstrong and Ward [1926] as well as Coustenoble and Armstrong [1934] describe and display briefly two very different ways to signal intensifying emphasis, however, without specifying which way of signalling is used under which conditions. The two signalling strategies can be paraphrased as follows:

(1) Emphasis for intensity can be characterized by a pronounced rising-falling F0 peak contour that spans the emphasized syllable and that has shallow slopes at both ends as well as by a high plateau in the centre. Underneath the dull-edged F0 peak, the emphasized syllable – and the vowel in particular – are lengthened.
(2) Emphasis for intensity can be produced with a narrowed and lowered pitch course with a short falling movement across the emphasized syllable.

The second way of signalling intensifying emphasis resembles the *-kickt* example (fig. 1).

1.3. Towards Positive and Negative Intensification

The entire ‘Lindenstraße’ database was searched systematically for further intensifying emphases. As the search was guided by the question ‘Is the semantics of the word intensified by phonetic means?’, it was primarily oriented towards meaning, but also integrated phonetic observations like high perceptual salience. The search yielded 40 additional instances. As outlined in Kohler [2005], 35 of the 40 tokens resembled *-kickt* and signalling strategy (2) for intensifying emphasis (cf. 1.2). That is, the emphatically intensified syllables had considerably lengthened onset consonants that were followed by shortened vowels with barked sound qualities. However, almost all of these 35 syllables also differed in one major characteristic from *-kickt*. They showed clear F0 movements that could be assigned to different types of pitch accents. But consistent with signalling strategy (2) the pitch-accent movements fell steeply and shortly across the emphatically accented syllables. Another peculiarity of the 35 emphasis tokens was that the segmental lengthening extended beyond the onset consonant of the emphatically accented syllable to the preceding sound. Together with a slowdown of the acoustic energy change towards the onset consonant (fig. 1) this creates the impression of a ‘ritardando’ before the emphatically accented syllable.

The remaining five instances of intensifying emphasis in the ‘Lindenstraße’ database were clearly different from the previous 35. The five emphasized syllables showed phonetic characteristics that resembled signalling strategy (1). Two examples, *doch* (‘yes [he does]’) and *Fest* (‘party’), are illustrated in figures 2a, b. As is immediately obvious, not the onset consonants, but the vowels showed major lengthening and were produced with sonorous breathy voice underneath high-rising F0 plateau contours that spanned a frequency range of 75–100 z or 8–13 semitones. The F0 and acoustic energy transitions at the vowel edges were not abrupt, but gradual, which contributed to the ‘smooth’ character of the vowels.

Contrastive analyses of the intensifying emphases in words like *-kickt* on the one hand, and *doch* and *Fest* on the other suggested that the signalling strategies of intensifying emphasis that were already noted by Armstrong and Ward [1926] and Coustenoble and Armstrong [1934] are not simply variants of ‘the same thing’. Rather, the two signalling strategies represent *different types* of intensifying emphasis that differ in the *valence* of the intensification. The phonetic patterns that characterized *-kickt* and another 35 examples in the ‘Lindenstraße’ dialogues express a negative valence. The phonetic patterns that were found for the remaining 5 examples, including *doch* and *Fest*, express a positive valence. This idea is further supported by the semantic-pragmatic frames in which the intensifications were found. For example, among those words that behaved phonetically similar to *-kickt* (‘kicks down’) were also *bescheuerte* (’stupid’), *besoffene* (‘drunk’), *erstickt* (‘suffocated’), *‘rausgetorkelt’* (‘staggered out’), *angebrüllt* (‘screamed at’), *schwachsinnig* (‘idiotic’), and *beschissen* (‘lousy’). The other three words whose emphatic exponents accorded with *doch* (‘yes’) and *Fest* (‘party’) were *Spaß* (‘fun’), *ganz* (‘totally’) and *fertig* (‘ready’). On this basis, the two subtypes that were assumed to exist within intensifying emphasis were called *positive and negative intensification*, the latter term replacing the narrow (non-pitch) ‘force accent’ concept of

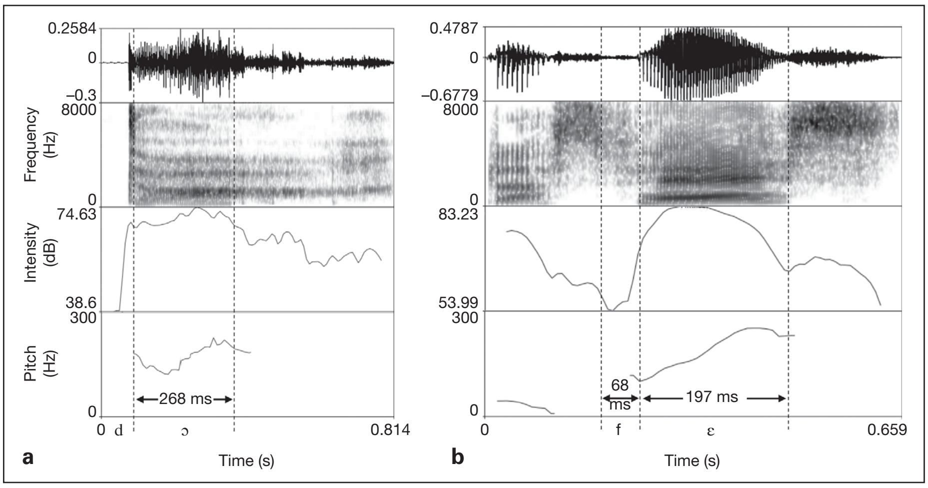

Fig. 2. Oscillogram, spectrogram, acoustic energy course, and F0 course of the emphatically intensified words *doch* ('yes [he does]') and *Fest* ('party'), produced by the female speaker 'l04' (**a**) and the male speaker 'l05′ (**b**) in the 'Lindenstraße' dialogues. The word *Fest* is shown preceded by the definite article *das* ('the'). F0 ranges and the CV durations of the emphasized syllables are given.

Kohler [2003]. Armstrong and Ward [1926] and Coustenoble and Armstrong [1934] did not take into consideration that their emphasis for intensification could be further subdivided, as they focused on phonetic form and disregarded possible meaning differences.

1.4. Aims and Assumptions

The idea of a distinction between positively and negatively intensifying emphasis was taken up by Kohler [2006a] and paved the way for an initial systematization of different types of emphasis in German. Following on, the present study pursues two linked aims. The first aim is to provide further meaning-guided evidence for the existence of both positive and negative intensification by testing whether semantic-pragmatic context frames can elicit separate phonetic profiles. The second main aim is to flesh out these profiles in a comprehensive and systematic acoustic analysis of segmentally and prosodically controlled speech material.

In view of 1.3, the parameters that are included in the acoustic analysis focus on the section of the speech signal closely around the onset consonant and vowel of the syllable that bears the intensifying emphasis. The following multiparametric phonetic differences are expected in this section.

(1) Duration: Both positive and negative intensification cause considerable lengthening in the vicinity of the CV sequence of the emphasized syllable.
(1a) Under positive intensification, this lengthening affects the vowel of the emphasized syllable.
(1b) Under negative intensification, the lengthening relates to the onset consonant of the emphasized syllable as well as to the immediately preceding speech sound.

(1c) The speech sounds that are not lengthened are shortened compared with the other type of intensification.

(2) F0: Both positive and negative intensifications are accompanied by F0 peak contours. However, unlike the peak contours of negative intensification the peak contours of positive intensification (2a) have a shallower fall, (2b) that starts later relative to the accented-vowel onset (hence creating a rise rather than a fall across the emphasized syllable), and (2c) that is preceded by a (longer) high F0 plateau (hence creating a blunt rather than a pointed peak).

(3) Acoustic energy: Both positive and negative intensification affect the acoustic energy dynamics in the vicinity of the CV sequence of the emphasized syllable.

(3a) Compared with the ‘smoothly sounding’ vowels of positive intensification, the ‘barked-sounding’ vowels of negative intensification have greater acoustic energy changes at both ends.

(3b) The speech segment that precedes the onset consonant in the case of negative intensification shows an acoustic energy transition into the onset consonant which is slower than in the case of positive intensification.

(4) Voice quality: The vowels of the emphasized syllables differ in voice quality. It is (4a) more breathy in the case of positive intensification, and (4b) more pressed in the case of negative intensification.

2. Method

2.1. Data Acquisition

A speech corpus was required that met a number of – partly contradictory – prerequisites. On the one hand, it had to show a high frequency of occurrence of emphatic intensifications, realized as naturally as possible by a number of Standard Northern German speakers. Moreover, in order to provide evidence for positive and negative intensification and their different phonetic profiles, it was necessary to have a sufficient number of cases produced in controlled contexts with regard to segmental and prosodic structure. Existing spontaneous speech corpora like the ‘Lindenstraße’ section in the ‘Kiel Corpus of Spontaneous Speech’ contain a substantial number of emphatic intensifications, but without the contextual control required for contrastive analyses. Contextual control is easier to get in read speech corpora, which, however, cannot be expected to show the relevant emphases. For this reason, a new speech corpus was generated that aimed at a compromise between read speech and a natural, expressive speaking style.

Instead of using prose passages or lists of individual unusual sentences, 7 short dialogues were constructed, concerned with issues of everyday life. The dialogues were complemented by 29 short everyday monologues, which, however, were easily interpretable as parts of dialogues and also addressed the dialogue partner explicitly via questions and/or personal pronouns (cf. ‘Appendix’). Furthermore, pairs of speakers were recorded while sitting face to face on opposite sides of a table. In the case of the dialogues, each subject had to take the role of one of the dialogue partners. During the monologues the situation was similar, except that one of the partners just listened and did not contribute to the conversation. However, s/he was directly addressable all the time. The designed texts and their face-to-face realizations together accounted for the normal use of speech as a means of communication, thus making it easier for the subjects to speak in a natural way. In addition, the pairs of speakers judged each others’ productions with regard to naturalness and repeated the corresponding dialogue/monologue, until both speakers agreed that a satisfactory version was achieved. For this purpose, the speakers were also allowed to slightly adjust the texts to their own vocabulary, for example, by introducing, substituting, and skipping words, or by changing the wording of passages.

The dialogue partners, 1 male and 1 female in each session, were married or good friends. They also knew the experimenter (the author O.N.) very well. At the beginning of each session the subjects were involved into small talk initiated by the experimenter, in the course of which they familiarized themselves with the recording situation, its facilities and the room itself (a sound-treated lecture room at the Institute of Phonetics and Digital Speech Processing, IPDS, of the University of Kiel). As for the small talk itself, the experimenter drew the subjects' attention towards topics that triggered jokes and laughter as well as different kinds of expressive negative statements. So, beyond the familiarization, the small talk also contributed to a relaxed and informal atmosphere in the laboratory. Moreover, it served to activate the subjects' knowledge of an expressive speaking style and of intensifying emphasis and to break down their inhibitions about using both in their utterances. Only subjects were selected for the recordings that were known to the experimenter as having an extrovert, expressive personality. During the small talk, the experimenter was careful not to influence the subjects' later productions of the texts. For example, while motivating the subjects to produce intensifying emphasis, he avoided its use in his own utterances, particularly in combination with the later key words (cf. 2.2).

A meaning-guided elicitation method was used in order to obtain the desired instances of intensifying emphasis. The basic idea of this method is to design utterances with selected lexical and grammatical characteristics and to arrange these utterances deliberately in a way that a specific semantic-pragmatic context is created. This context frame then triggers the elicitation of the respective meanings – in the present case positive and negative intensifications – on a particular key word. A major advantage of a meaning-guided elicitation is that it does not require actors or otherwise trained speakers who may be criticized for producing stereotypical, exaggerated speech [Iriondo et al., 2007]. Instead, naïve speakers can be used who only need to be controlled through minor, non-linguistic instructions.

The scope of the data acquisition was restricted to positive and negative intensification. A 'neutral' baseline condition was not included. In order to fan out intensifying emphasis and to determine phonetic profiles for the emerging subtypes, it is necessary to relate the subtypes to each other, but not to a 'neutral' condition. The only prerequisite is to have sufficient grounds for assuming that the analyzed tokens can count as emphatic through clearly defined expressive categories in a data acquisition paradigm.

2.2. Key Words

In the texts presented to the subjects, the key words were typographically highlighted. The highlighting supported the emphasis attraction of the key words and prevented them from being substituted or skipped by the subjects. In total, the dialogues and monologues contained 44 key words, 21 in contexts for positive and another 23 in contexts for negative intensification.

Segmental and prosodic structures of the 21 + 23 key words were controlled with regard to the following aspects. First, all key words occurred at the nuclear accent position in the utterances. Care was also taken that the target syllables were separated by at least one syllable from phrase boundaries or preceding accented syllables in order to avoid influences of tonal crowding and phrase-final or accent-clash-related lengthening on the measurements. Second, the target syllables that were to become emphatically accented started in all key words with a single consonant in the syllable onset. Third, the vowels in the target syllables of the key words were balanced with regard to their phonological quantity. Ten key words in the positive and 12 key words in the negative contexts had phonologically long vowels or diphthongs. In the other half of the key words, i.e. 11 in both positive and negative contexts, the vowels in the target syllables were phonologically short. Moreover, the two sets of phonologically long and short vowels were balanced with regard to front unrounded, open, and back rounded qualities. Fourth, the key words were selected to match semantically with positive or negative intensification. Thus, different key words were used in the positive and negative contexts, except for a small subset of key words whose semantics was considered neutral or ambiguous.

Other aspects in and around the key words were not directly controlled, but can be regarded as negligible for the measurements of the study insofar as they were randomly distributed across the positive and negative contexts. This concerned, for example, the phonological voicing feature of the single consonant in the onset of the target syllable, the presence and type of consonant in the coda of the target syllable, the type of speech sound preceding the target syllable (e.g., consonant/vowel, voiced/voiceless), and the distance (of 1–4 syllables) between the target syllable and another accented syllable or a prosodic boundary.

2.3. Data Recording

Prior to each recording session, the subjects were told that they were to read expressive dialogue and monologue texts of everyday situations while addressing the dialogue partner on the opposite side of the table. Moreover, the speakers were allowed to familiarize themselves with the texts and to modify them slightly according to their personal tastes. The instruction of the speakers, their familiarization with the texts, and the initial small talk with the experimenter took between 20 and 30 min. The speech signals were recorded by directional microphones placed on the table in front of each subject. A complete recording session took between 1 and 2 h. Four pairs of speakers from Northern Germany, i.e. 8 subjects altogether, were recorded, three pairs in their 30s, and another pair in their 60s.

The recording sessions comprised three parts. The first part was dedicated to the dialogues. The speakers performed two rounds with reversed roles. Then, as the second part, the subjects continued with the monologues. It was left to each subject, whether s/he started with the monologues designed for positive or negative intensification. At the end of the second part, each subject had produced the two sheets of monologues at least once. In the third part, the subjects returned to the dialogues and performed another two rounds with reversed roles. In all three parts, the 2 speakers repeated each dialogue or monologue, until both agreed that they had achieved a natural-sounding version. After each part, there was a short pause which was again filled with small talk between the 2 subjects and the experimenter.

2.4. Speech Corpus

The dialogues and monologues that were judged by both dialogue partners to be the final, most natural-sounding versions and that came from the second and third parts of the recording session were assembled to a speech corpus. Excluding the first two rounds of the dialogue productions was necessary as speakers became more relaxed and expressive in the course of this initial part of the recording session. They also became more critical and developed higher demands concerning the natural-sounding character of their utterances. The resulting speech corpus consists of 4 male and 4 female (i.e. 4 pairs of) speakers who produced overall around 45 min of speech or around 4,800 words. A total number of 352 of these words were key words, 168 (= 21 words × 8 speakers) from the context frame for positive and 184 (= 23 words × 8 speakers) from the context frame for negative intensification. They were analyzed with regard to duration, intonation (i.e. F0), acoustic energy, and voice quality.

2.5. Acoustic Analyses of the Key Words

2.5.1. Duration

Testing the assumptions (1a–c) required three segment durations for each key word. In the target syllables the durations of the consonant in the syllable onset as well as of the subsequent long or short vowel were measured. The third measurement concerned the duration of the sound segment that immediately preceded the target syllable.

2.5.2. Fundamental Frequency

Following assumptions (2a–c), the alignment and shape characteristics of the F0 peaks on the target syllables were measured. Figure 3 provides a summary of the measurements, which were all done in *Praat* [Boersma, 2001]. The F0-peak *alignment* was measured in terms of the fall onset at the end of the peak maximum relative to the onset of the accented vowel. A number of previous perceptual studies showed that the accented vowel – and its onset in particular – are the crucial references for describing intonational patterns of German [Niebuhr and Kohler, 2004; Niebuhr, 2006, 2007a, b]. As for the F0 reference, there were two reasons for preferring the fall onset over the rise offset or the peak maximum. First, it had to be taken into account that F0 peaks in the context of positive intensification could show extensive high plateaux which make the determination of a clear F0 peak maximum problematic [Silverman and Pierrehumbert, 1990]. Second, it is known from a number of studies across languages (including German) that the fall-onset alignment is a more reliable characteristic of

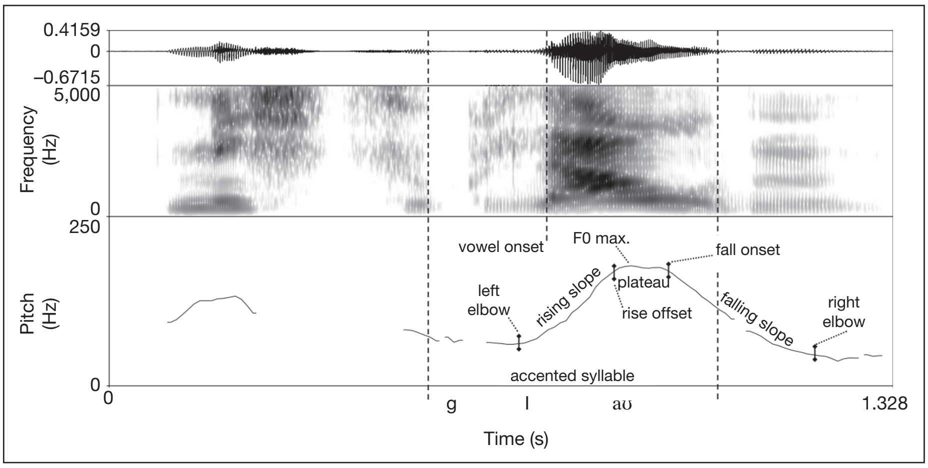

Fig. 3. Oscillogram, spectrogram, and F0 contour of the utterance *Nicht zu glauben* ('unbelievable'), produced by the female speaker ALI in the context frames for positive intensification. The F0 panel exemplifies the F0 measurements taken for each key word.

different intonation patterns than the rise-offset alignment, particularly in connection with high plateaux [Kohler, 1991a; Igarashi, 2005; Knight and Nolan, 2006].

The *shape* characteristics of the F0 peaks were represented in two measurements. First, the shape of the peak maximum was determined by measuring the duration of the high F0 level in between rise offset and fall onset. The high F0 level comprised all those F0 values that were at most one semitone lower than the absolute F0 peak maximum. The one-semitone threshold was shown in a number of studies to roughly correspond to the just noticeable difference of pitch change in speech or speech-like stimuli [Mack and Gold, 1986; Rosen and Fourcin, 1986]. Thus, the duration of the high F0 level indicates how peaky or plateau-like the shape of the F0 peak maximum is perceived. The second shape measure concerned the rising and falling F0 movements to both sides of the peak. A shape index was calculated by dividing the average slope of the F0 rise by the average slope of the F0 fall. The average slopes were measured in semitones per second and represented as moduli (i.e. as absolute values). A shape index larger than 1 showed that the average slope of the rise was steeper than the average slope of the fall. Inversely, F0 peaks in which the falls were on average steeper than the rises yielded shape indices smaller than 1.

Calculating the shape indices also required determining the left and right 'elbows' of the F0 peaks, i.e. their rise onsets and fall offsets. The elbow detection was done manually. Recent studies demonstrated that human labellers – and trained phoneticians in particular – are not worse than semi-automatic or automatic algorithms as regards the adequacy and consistency of the detected elbows [del Giudice et al., 2007; Petrone and D'Imperio, 2009]. The elbows had to be located at most two syllables away from the accented target syllable of the key word. Within this distance only those changes in the direction of the F0 course counted as elbows that were continued into the same frequency direction for at least four further F0 values (i.e. 40 ms). This consistent frequency change had to go along with a consistent increase in slope from one F0 value to the next. Finally, elbows that coincided with consonant boundaries were disregarded in order to exclude microprosodic perturbations.

2.5.3. Acoustic Energy

With regard to assumptions (3a) and (3b), characteristics of the acoustic energy course in the area of the target syllable were represented by four measures for each key word. First, similar to the peak-maximum alignment in the F0 dimension, the duration between the vowel onset of the target syllable and the acoustic energy maximum in the vowel was determined. Moreover, the dynamics in the

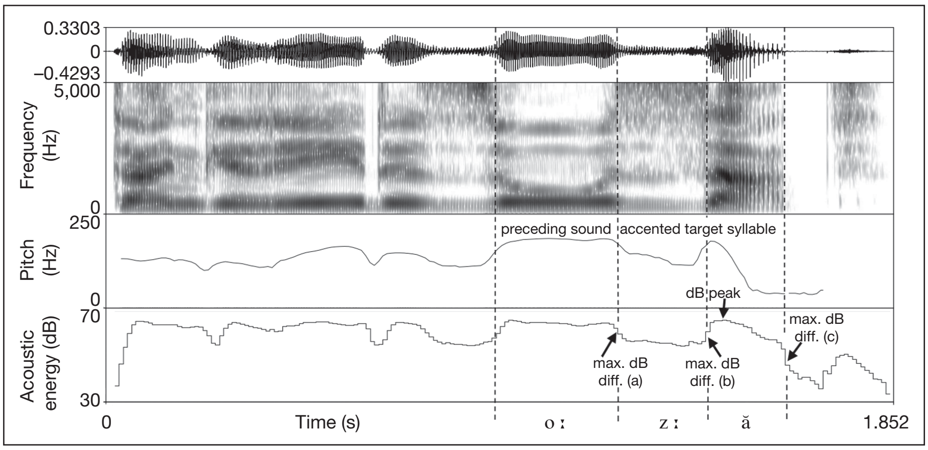

Fig. 4. Oscillogram, spectrogram, F0 contour, and acoustic energy contour of the utterance section *Dein Gerede so satt* ('so fed with up with your ramblings') produced by the male speaker SKR in the context frames for negative intensification. The key word *satt* ('fed up', broken lines) was meant to attract the emphatically intensified accent. The lower acoustic energy panel shows the four measurements that were taken for each key word.

changes of acoustic energy were estimated at the following three points: (a) towards the consonantal onset of the target syllable, (b) at the consonant-vowel transition in the target syllable, and (c) at the vowel offset of the target syllable.

The estimations were based on measurements of the maximum dB differences between adjacent dB values in the falling or rising acoustic energy courses at (a)–(c). In order to facilitate comparisons between the dynamics of rising and falling courses, the maximum dB differences were measured as moduli. It may further be assumed that the estimated dynamics of acoustic-energy changes reflect articulatory dynamics, particularly in the transition from consonants to vowels or vice versa.

The measurements of acoustic energy peaks and maximum differences were done in *Praat* by means of discretized acoustic energy courses. That is, the values of the acoustic energy course that the algorithm in *Praat* calculated every 10 ms were not linearly interpolated, but displayed in a staircase fashion. An example that also summarizes the four acoustic-energy measurements for each key word is given in figure 4. The values of the acoustic energy course were based on an analysis window of 25 ms. The relatively small analysis window provided a high temporal resolution to capture the rapid changes that were expected for intensifying emphasis. Finally, the measurements accounted for the mathematical correlation between F0 and acoustic energy. That is, variation in the acoustic energy course that stemmed from concomitant F0 variation was filtered out.

2.5.4. Voice Quality

Assumptions (4a) and (4b) were tested by calculating the dB differences between the amplitudes of the first two harmonics H1 and H2. Based on a number of empirical studies (that were primarily concerned with the degree of breathiness), the harmonic amplitude differences will be interpreted in the present study along a continuum from pressed to breathy voice. Lower (and predominantly negative) values indicate a more pressed and higher (and predominantly positive) values a more breathy voice [Fischer-Jørgensen, 1967; Klatt and Klatt, 1990; Keating and Eposito, 2007]. Modal voice shows intermediate H1–H2 differences that are roughly located ±5 dB around 0 dB. The H1–H2 difference was determined by means of narrow-band spectral section windows (50 Hz bandwidth) in *wave surfer* (www.speech.kth.se/wavesurfer/*)* at the centres of the vowels of the target syllables, or closely around the vowel centres, if no plausible values for H1 and H2 could be found at that points of time. Figure 5 shows an edited screenshot that illustrates the determination of the H1–H2 differences in *wave surfer*.

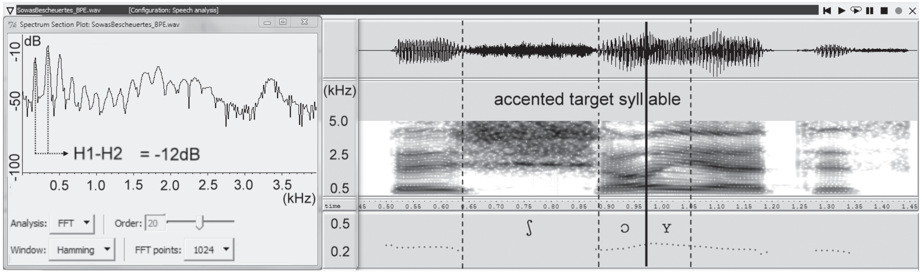

Fig. 5. Edited screenshot of the speech-signal analysis in *wave surfer*. The right panel displays the oscillogram, spectrogram, and F0 contour of the key word *Bescheuertes* (‘dumb’). It was produced by the male speaker BPE in a context frame for negative intensification of the target syllable *-scheu* (broken lines, right panel). The continuous black line refers to the point at the centre of the diphthong [ɔʏ] where the narrow-band spectral section window was taken. It is shown in the left panel.

3. Results

The overall 168 and 184 key words that were produced in the context frames for positive and negative intensification were split up into two subsamples along the phonological quantity of the vowel in the target syllable (long vs. short). The short-vowel subsample consisted of 80 key words from the positive and 96 key words from the negative context frames. In the long-vowel subsamples, there were 88 key words from each of the two context frames. The subdivision of the overall sample of key words was mainly done with regard to the duration measurements. It is evident that these measurements will be different for target syllables with long and short vowels. Thus, this factor was to be kept separate from the assumed effects of positive and negative intensification. However, beyond the obvious duration issue, there is also cross-linguistic empirical evidence that vowel quantity or the concomitant duration differences affect the temporal dynamics of F0 and acoustic energy contours [Ladd et al., 2000; Mády et al., 2005].

The statistical analysis of the data in the long- and short-vowel subsamples was done in two steps whose results are given in 3.1 and 3.2. First, a multivariate analysis of variance (MANOVA) was performed based on 11 dependent variables that corresponded to the acoustic measurements in the dimensions of duration, F0, acoustic energy, and voice quality:

(1) duration of the single consonant at the onset of the target syllable, ‘DURc’,
(2) duration of the long/short vowel of the target syllable, ‘DURv’,
(3) duration of the sound segment that immediately precedes the onset consonant of the target syllable, ‘DURpresound’,
(4) alignment of the onset of the F0 fall relative to the vowel onset of the target syllable, ‘F0fallalign’,
(5) duration of the high F0 level of the rising-falling peak on the target syllable, ‘F0plateau’,
(6) shape index of the F0 peak on the target syllable, ‘F0shape’,
(7) alignment of the acoustic energy maximum in the vowel of the target syllable relative to the vowel onset, ‘Ealign’,

(8) maximum acoustic energy difference at the consonant onset of the target syllable, ‘Eonsetc’,
(9) maximum acoustic energy difference after the vowel onset of the target syllable, ‘Eonsetv’,
(10) maximum acoustic energy differences at the vowel offset of the target syllable, ‘Eoffsetv’,
(11) difference between the amplitude of the first and second harmonic (H1–H2) around the centre of the vowel of the target syllable, ‘VQh1–h2’.

In the MANOVA the 11 dependent variables were set in relation to two fixed factors: ‘context frame’ and ‘vowel quantity’. Each factor had two levels. In the case of ‘context frame’, the two levels corresponded to the expected meaning-guided emphases produced on the target syllables of the key words. The levels were thus called ‘positive intensification’ and ‘negative intensification’. In the case of the fixed factor ‘vowel quantity’, the two levels represented the subdivision of the overall sample of key words into the subsamples of key words with long and short vowels.

The second step of the statistical treatment of the data consisted of two discriminant analyses that were performed separately for the long- and short-vowel subsamples. They determined the extent to which differences in the MANOVA formed multiparametric phonetic profiles of emphatic intensification that allow a reliable assignment of the key words to the original positive and negative context frames. Ultimately, the discriminant analyses were to provide a basis of deciding whether positive and negative intensification may be regarded as two distinctly coded and hence separable functional categories in speech communication.

3.1. Results of the MANOVA

The main results of the MANOVA are summarized in table 1. As can be seen, all 11 phonetic parameters differed highly significantly between the key words for positive and negative intensification. For example, there was a diametrically opposed shift of the consonant and vowel durations in the target syllable. Compared with key words for positive intensification, the key words for negative intensification had shorter vowels (independent of phonological quantity). Simultaneously, the preceding syllable-initial consonants were longer than in the key words for positive intensification. The durational differences between positive and negative intensification also extended to the speech sound that was located immediately before the target syllable. Like the syllable-initial consonant, this speech sound was longer in the context of negative than in the context of positive intensification. Moreover, the MANOVA yielded a pronounced difference in voice quality. The mean H1–H2 values were clearly positive in connection with positive and clearly negative in connection with negative intensification (the negative mean is close to the *Bescheuertes* example in fig. 5).

As for the F0 parameters, the key words for positive intensification showed F0 peaks with wider high plateaux that started to fall considerably later after the vowel onset than in the key words for negative intensification. The latter key words had on average shape indices slightly below 1. Together with the smaller high plateaux on key words for negative intensification this indicates pointed F0 peaks with slightly faster falling than rising slopes. By contrast, the F0 peaks in the key words for positive

Table 1. Descriptive and inferential statistics of the MANOVA for the fixed factor 'context frame' that refers to the key words/target syllables for positive or negative intensification

Dependent variables	Context frame/key words		Measures of MANOVA		
	positive intensification	negative intensification			
	mean	mean	d.f.	F	p
DURc	137.26	173.69	1	77.74	<0.001
DURv	162.54	122.80	1	38.65	<0.001
DURpresound	84.84	95.57	1	62.58	<0.001
F0fallalign	137.6	57.37	1	60.82	<0.001
F0plateau	65.26	47.47	1	45.24	<0.001
F0shape	1.25	0.94	1	23.58	<0.001
Ealign	114.87	68.39	1	52.39	<0.001
Eonsetc	5.04	4.59	1	18.11	<0.001
Eonsetv	4.83	5.38	1	12.93	<0.001
Eoffsetv	4.99	5.48	1	15.14	<0.001
VQh1-h2	4.79	–12.21	1	47.17	<0.001

The mean values of the duration variables, including 'F0fallalign', 'F0plateau', and 'Ealign' are given in milliseconds (ms). The means of the acoustic energy and voice quality variables are in dB. The statistics are based on 168 positive and 184 negative tokens.

intensification were not just dull-edged. According to the mean shape index of 1.25, the peaks also had clearly asymmetrical shapes with fast rising and slow falling movements.

For acoustic energy the results show that the maximum dB differences at the onsets and offsets of the vowels in the target syllables were greater in key words for negative than in key words for positive intensification, which suggests that the acoustic energy changes into and out of the syllable nucleus were more dynamic in key words with negative intensification. However, with regard to the differences in 'Eonsetc' it was the key words for *positive* intensification that showed a more dynamic change (i.e. decrease) in acoustic energy *towards* the target syllable. Finally, parallel to the alignment differences in the onset of the F0 fall, the acoustic energy maxima were aligned later after the vowel onset in key words for positive than for negative intensification.

Beyond the main effects of the factor 'context frame', the other fixed factor 'vowel quantity' also yielded significant results. That is, there were a number of differences between the key words with long and short vowels. Regarding the F statistics, the strongest effects of vowel quantity concerned the durations of the vowels in the target syllables as well as the temporal dynamics of the F0 and acoustic energy courses. On average, the durations of the long vowels were more than twice the durations of the short vowels (192.02 vs. 97.81 ms; d.f. = 1; F = 317.33; $p < 0.001$). Following the shift of the vowel offset to a later point in time, F0 and acoustic energy maxima were reached and/or left later in long than in short vowels. The distances between the vowel onsets and the acoustic energy maxima amounted to 133.08 ms in long and 53.94 ms in short vowels (d.f. = 1; F = 182.74; $p < 0.001$). Likewise, the falling movement after the high F0 plateau set in around 140 ms from the vowel onset in long and around 60 ms from the vowel onset in short vowels (d.f. = 1; F = 58.56; p < 0.001). In terms of the dynamics of the F0 and acoustic energy changes, the high F0 plateau was around 15 ms smaller in short than in long vowels (49.73 vs. 63.66 ms; d.f. = 1; F = 23.90; p < 0.001). In addition, the maximum dB step into the vowel was larger for short than for long vowels (5.49 vs. 4.69 dB; d.f. = 1; F = 31.73; p < 0.001), which suggests a faster rise in acoustic energy from the preceding consonant to the vowel.

Furthermore, the vowel quantity also had small effects on the voice quality of the vowel and on the duration of the onset consonant of the target syllable. Inverse to the change in vowel duration, the preceding consonant was longer before short than before long vowels (161.26 vs. 149.11 ms; d.f. = 1; F = 5.11; p = 0.024). However, with 10 ms on average the difference in consonant duration was relatively small compared with the difference in vowel duration. With regard to voice quality, the H1–H2 differences were slightly greater for long than for short vowels (–0.10 vs. –0.80 dB; d.f. = 1; F = 9.75; p = 0.002).

Finally, it must be noted that for most of the phonetic parameters there were significant interactions between the two fixed factors 'context frame' and 'vowel quantity'. In the case of F0 and acoustic energy, the interactions were due to the fact that the phonetic differences between the key words for positive and negative intensification were greater in the long vowel than in the short vowel condition. This concerned 'F0fallalign', 'F0plateau', 'Ealign', and 'Eoffsetv'. However, for duration variables 'DURc' and 'DURv' the situation is more complex. As was noted above, the positive and negative context frames had inverse effects on the consonant and vowel durations. The durations in the CV dyad were shifted in favour of the vowel in key words for positive and in favour of the consonant in key words for negative intensification. In this context the interactions of 'DURc' and 'DURv' with the two fixed factors 'context frame' and 'vowel quantity' mean that the duration shifts rely primarily on the vowel in syllables with long vowels and on the consonant in syllables in short vowels.

3.2. Results of the Discriminant Analyses

The measurements of the 11 dependent variables formed significantly different phonetic profiles for positive and negative intensification along the canonical discriminant functions of both the long-vowel and the short-vowel analyses. Based on these separated profiles, the great majority of the key words were assigned correctly to their original type of positive or negative context frame. The discriminant analysis of the subsample of key words with short vowels yielded 74.4% correct assignments. In the discriminant analysis of the long-vowel key words, the correct assignment rate was only slightly worse, i.e. 71.0%. However, not all 11 phonetic parameters were equally relevant in discriminating key words for positive and negative intensification. In addition, it was striking that the importance of the parameters also varied depending on whether the target syllable in the key word had a long or a short vowel. In order to illustrate these points, table 2 provides the standardized canonical discriminant coefficients – i.e. the 'weights' – of the individual phonetic parameters in the two discriminant analyses together with other quality criteria of the canonical discriminant functions.

For the subsample of short-vowel key words table 2 shows that the (diametrically opposed) changes of the consonant and vowel durations in the target syllable were crucial for discriminating between key words with positive and negative intensification.

Table 2. Standardized canonical discriminant coefficients of the 11 dependent variables in the discriminant analyses based on the subsample key words with short vowels or long vowels in their target syllables (n=176 for each subsample)

Dependent variables	Short-vowel key words	Long-vowel key words
	standardized discriminant coefficient	standardized discriminant coefficient
DURc	*–0.383*	–0.085
DURv	*0.468*	0.104
DURpresound	–0.224	*–0.409*
F0fallalign	*0.257*	–0.011
F0plateau	–0.053	0.145
F0shape	*–0.332*	*–0.334*
Ealign	0.193	*0.353*
Eonsetc	–0.039	0.041
Eonsetv	*0.340*	–0.147
Eoffsetv	–0.174	*0.267*
VQh1–h2	–0.007	*0.409*
Wilks' lambda	0.666	0.756
Eigenvalue	0.501	0.323

The five dependent variables that have the highest coefficients and that hence make the greatest contribution to the discrimination of positive and negative intensification are given in italics. Additionally, Wilks' lambda and eigenvalue, are shown for each discriminant analysis.

In the F0 dimension, it was the shape index and the onset of the F0 fall after the high plateau that played the most important role. From the acoustic energy measures, only the dynamics at the vowel onset of the target syllable – in terms of the maximum dB difference – contributed substantially to distinguishing key words with positive and negative intensification. The voice quality measure can be regarded as irrelevant.

The corresponding set of top-five discriminating parameters in the subsample of long-vowel key words looked quite different. For example, the voice quality measure was one of the two most important discriminating characteristics in the subsample of long-vowel key words. The other crucial characteristic was the duration of the speech sound preceding the target syllable. The consonant and vowel durations in the target syllable itself were only of minor importance, contrary to the subsample of short-vowel key words. The only relevant similarity between the two subsamples with short- and long-vowel key words lies in the F0 dimension. That is, the shape index contributed clearly to distinguishing positive and negative intensification in both short- and long-vowel key words. However, different from the subsample of short-vowel key words, the alignment of the fall at the end of the high F0 plateau was irrelevant in the subsample of long-vowel key words. Instead, the alignment of the acoustic energy maximum in the vowel was important. Finally, it was the acoustic energy dynamics at the vowel *offset* that significantly supported the discrimination of positive and negative intensification in the subsample of long-vowel key words, as opposed to the dynamics of the vowel *onset* in the subsample of short-vowel key words.

However, it can also be seen in table 2 that the canonical discriminant functions of the two subsamples of long- and short-vowel key words yielded rather high values for Wilks' lambda of 0.666 (d.f. = 11; $p < 0.001$) or 0.756 (d.f. = 11; $p < 0.001$). The corresponding eigenvalues were only 0.501 or 0.323. That is, even though the two discriminant functions were successful in forming phonetic profiles and hence in distinguishing between positive and negative intensification, a considerable amount of the overall variance in the data was left unexplained. This issue is taken up again in 4.2 of the following discussion section, where it will be argued that the speakers produced a third type of intensifying emphasis in the context frames of both positive and negative intensification.

The two discriminant analyses also included correlation matrices for the 11 phonetic parameters. The corresponding results were quite similar between the subsamples of short-vowel and long-vowel key words. The most important correlations are listed in table 3. They are all highly significant ($p < 0.001$). With regard to segment durations, table 3 shows that the consonants and vowels in the target syllables are strongly negatively correlated. That is, the vowel duration increased at the cost of the preceding consonant duration and vice versa. This inverse duration relationship was already noticed in the context of the MANOVA (table 1). Table 3 shows further that from the perspective of the vowel of the target syllable the duration of the onset consonant shrunk or expanded in accordance with the duration of the speech sound that immediately preceded the target syllable, i.e. onset consonant and preceding speech sound were highly positively correlated ($r = 0.645/0.585$). Across acoustic dimensions, there were a number of positive correlations between F0 parameters and the duration of the vowel in the target syllable. With increased vowel duration, the high F0 plateau also became longer and the F0 fall set in at a greater distance from the vowel onset. Yet, the delay of the F0 fall did not result in a steeper falling slope relative to the rising slope. On the contrary, the positive correlation between the vowel duration and the shape index shows that the falling slope became successively shallower than the rising slope with increasing vowel duration. Finally, the correlations in table 3 revealed that relative to the vowel onset of the target syllable, changes in alignment of the F0 fall were paralleled by changes in the alignment of the acoustic energy maximum.

4. Discussion

4.1. Basic Interpretation of the Results

4.1.1. Phonetic Profiles of Positive and Negative Intensification

The results of the acoustic analyses showed clearly that in and around the target syllables of the key words there were extensive local changes in segment duration, F0, acoustic energy, and voice quality. Particularly with regard to the increase in segment durations and the creation and left-shift of high F0 plateaux, the revealed local changes go beyond the parametric patterns that are known to characterize syllables traditionally categorized as pitch-accented in German and other languages. For example, as for duration, the mean values of phonologically long and short vowels that were found by Simpson [1998] on the basis of German read and spontaneous speech corpora vary mainly between 70 and 120 ms (for long vowels durations were around 20–30 ms longer than for short vowels). Similar durations were measured by Mády et al. [2005], Möbius and van Santen [1996], and Kohler [1991a]. Kohler [1991a] further noted that the presence of a nuclear pitch accent increases durations by about 10–20% at the

Table 3. Coefficients (r) of important and statistically highly significant (Pearson's product moment) correlations between the 11 dependent variables in the two discriminant analyses

Dependent variable	DURc	DURpresound	F0fallalign	F0plateau	F0shape
DURv	–0.703/–0.566	–0.557/–0.559	0.652/0.674	0.590/0.673	0.579/0.625
Ealign			0.753/0.738		

In each column, the left value refers to the key words with short and the right value to the key words with long vowels in their target syllables (n = 176 for each of the two subsamples).

syllable level. Consonant durations are more variable, but overall shorter than vowel durations (e.g., due to syllable position, occurrence in clusters, lexical stress, accent, or rhythmic structure. Measurements by Möbius and van Santen [1996] suggest that German consonant durations were virtually all below 100 ms [cf. also Klatt, 1979, and van Santen, 1993, for other languages].

The mean durations of the consonants and vowels that characterized the target syllables in the context frames of positive and negative intensification were around 50% higher than the values above. Quite a few consonants and vowels even exceeded 350 ms and were hence longer than entire (pitch-accented) syllables in German read and spontaneous speech. Such an additional and considerable increase of consonant and vowel durations in relation to normal pitch accentuation was found across languages, including German, to be a characteristic feature of various form-meaning relationships that can be subsumed under emphatic accentuation [Klatt, 1979; Cooper at al., 1985; Dahan and Bernard, 1996; Braun and Ladd, 2003; Strangert, 2003; Dohen and Loevenbruck, 2004; Baumann et al., 2006, 2007]. Likewise, the intensity levels of the sounds in the target syllables clearly exceed values known from normal accentuation and are more similar to emphatic values [Lehiste, 1970; Geumann, 2001]. In summary, it is reasonable to assume on the basis of the exemplified parametric comparisons that – guided by the created context frames, and supported by the special features of the recording procedure that facilitated expressive speech – the key words of the present study were consistently realized by the 8 speakers with emphatic accents on the target syllables.

Furthermore, the results show that the phonetic profiles differ clearly between the emphatic accents that were produced in the positive and negative context frames. Neither of the two profiles is consistent with contrastive emphasis (cf. 1.2). With reference to the mean values presented in table 1, figure 6a and b provide a graphical summary of the multiparametric differences, separated into the subsamples of key words with long or short vowels in the target syllables. For the graphical summary, the measured maximum dB differences were interpreted in terms of dynamics of acoustic energy changes. As the degree of openness of the vocal tract is a major determinator for the acoustic energy level, the dynamics of acoustic energy changes also reflect the dynamics of the opening or closing articulatory movements in consonant-vowel or vowel-consonant transitions. Moreover, in line with the empirically determined correlation between H1 and H2 values and voice quality, lower or higher values were taken as indicating a more pressed or more breathy voice, respectively (cf. 2.5.4). The mean

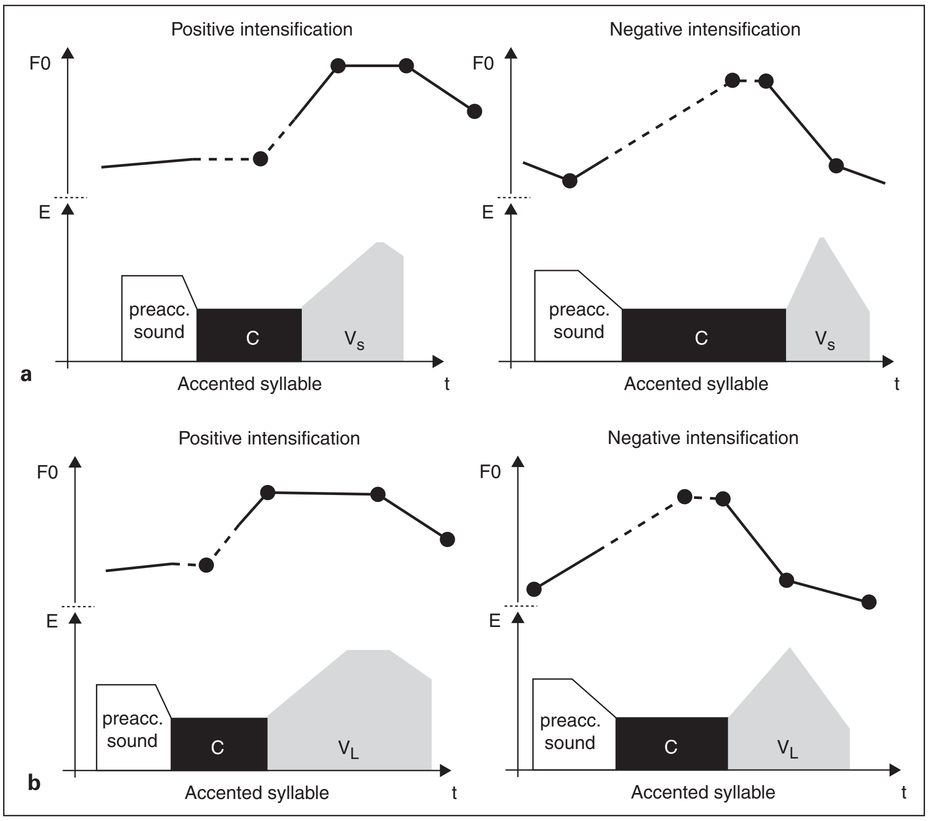

Fig. 6. Schematic illustrations of the phonetic profiles of positive and negative intensification that emerged from the key words with (**a**) short vowels (V_S) and (**b**) long vowels (V_L) in the accented target syllables. The shapes of the polygons in the lower panels represent the acoustic energy (E) courses. The upper panels sketch the characteristic F0 peak contours. Broken lines point to the possibility of voiceless-onset consonants. The different shades of the segment polygons refer to the differences in voice quality (i.e. lighter = breathier). All illustrations are based on the means of table 1. F0 ranges are oriented towards actually found values.

F0 curves partly consist of broken lines to account for the fact that voiceless onset consonants in some of the target syllables interrupted F0. However, the 'missing' F0 sections represented by the broken lines are only rough approximations, as in the actual productions the F0 curves changed their (fall-onset) alignments so that sections like the high plateau eluded the voiceless onset consonants [Kohler, 1991a; van Santen and Hirschberg, 1994].

By means of the phonetic differences illustrated in figures 6a and b, it was possible for the discriminant analyses to predict with significant reliability whether the corresponding key words were produced in the positively or negatively intensifying context frames. Taking the phonetic profiles and the predictability of the original context frames into account, it can be concluded that there are (at least) two types of differently coded intensifying emphasis in German. The semantic-pragmatic context frames in which these two types of intensifying emphasis occurred support their denotations as *positive and negative intensification*.

In view of figures 6a and b, the basic differences between the phonetic profiles of positive and negative intensification can be conceptualized as inversely timed syntagmatic contrasts in signal dynamics that highlight or suppress sonorous features at the cost or in favour of asonorous features. In the case of positive intensification, the articulation- and phonation-induced acoustic changes increase or hold their high dynamics until the accented-vowel onset of the target syllable. The onset consonant of the accented target syllable as well as the preceding speech sound are both quickly articulated, short segments with a sharp acoustic energy drop in between. At the right edge of the onset consonant the high signal dynamics finds its phonatory expression in a steep F0 rise into the accented vowel. Then, after the accented-vowel onset, the acoustic changes slow down drastically or even seem to pause for a moment in some key words (as far as this is possible for a biological system). The vowel duration is substantially increased, and the relatively high (positive) H1–H2 difference of around +5 dB is characteristic for a sonorous, slightly breathy voice (see Keating and Eposito [2007] and 2.5.4). The sonorous character of the vowel is further supported by the fact that the F0 rise freezes at a high level and creates an extensive plateau that is left around the end of the vowel in a shallow fall (see also the F0 courses in fig. 2, 3). The acoustic energy increases or decreases slowly at the onsets and offsets of the vowel, which accords with decelerated articulatory movements from or to the surrounding consonants (see also fig. 2). Thus, overall, positive intensification highlights sonorous features in and around the target syllable.

The opposite holds for negative intensification. The vowel is a quickly articulated short segment with great and abrupt acoustic energy changes at both edges and a pointed F0 peak that falls steeply to a low level right after the accented-vowel onset (see the F0 and acoustic energy courses in fig. 4, 5). The low-falling F0 adds to the *asonorous* characteristic of the vowel, which is mainly created by a pressed and irregular voice quality (compare the highly negative H1–H2 values with those for creaky voice in Keating and Eposito [2007]). In fact, there were quite a few negatively intensified key words in which sections of the falling F0 slope in the target syllable were disregarded by the F0 algorithm in *Praat* due to lack of periodicity. Furthermore, contrary to positive intensification, the segment durations prior to the vowel increase successively from the pre-accented speech sound to the accented consonant at the onset of the target syllable (see also fig. 1, 4, 5). The gradual increase in segment duration is reflected in the following fact. While the durations of onset consonant and preceding sound are positively correlated, the latter is still clearly shorter than the former. The deceleration also goes along with a slowly decreasing acoustic energy towards and a shallowly rising F0 movement across the onset consonant. In summary, for negatively intensified target syllables the signal dynamics slows down before the accented vowel and then speeds up abruptly after the accented-vowel onset.

4.1.2. Assumptions

German speakers use two different form-meaning relationships within intensifying emphasis that differ in their valence and may be called positive and negative intensification. The duration differences between positive and negative intensification that were represented by the antagonistic (though not proportional) changes in the CV dyad of the target syllable as well as by the relative lengthening or shortening of the preceding speech sound are in line with assumptions (1a–c). The findings are also consistent with the assumptions concerning F0 alignment and shape characteristics (2a–c), segment-dependent acoustic energy dynamics (3a, b), and the more breathy or pressed voice qualities that occur in the vowels of positively and negatively intensified target syllables (4a, b).

4.2. A Further Type of Intensifying Emphasis? – The Notion of Reinforcement

In deriving phonetic coding schemes for positive and negative intensification from the analyzed key words, it is important to keep in mind that the applied discriminant analyses achieved a significant, but not perfect separation and classification of positively and negatively intensified tokens (table 2). On the whole, 96 of the overall 352 key words were misclassified in terms of original and predicted context frames. There are three possible reasons for the misclassifications: (1) The classifications of the discriminant analyses are actually adequate and the 8 speakers produced positive intensification in negative and negative intensification in positive context frames. (2) The speakers did not produce key words with emphatically intensified, but with normally stressed and/or pitch-accented target syllables. (3) The speakers produced emphatically intensified key words, but used a further, third type of intensification that is compatible with both positive and negative context frames.

As regards possibility (1), it must always be reckoned that eliciting communicative meanings without giving the speakers explicit and direct instructions will generate a certain amount of variability. Yet, it is highly unlikely that the recording procedure and the created semantic-pragmatic as well as situational context frames left enough room for interpretation to elicit positive intensification in negative and negative intensification in positive context frames. However, it is known that speakers can deliberately clash lexical with prosodic meanings in order to create irony [Hirst, 2004]. But as the experimenter was aware of this fact, care was taken that the speakers of the present study were not given any reason to doubt the truth content of their produced utterances or the seriousness of their task, which would have been prerequisites for irony. So, although it cannot be excluded that a small number of key words was still realized in an ironic fashion, possibility (1) cannot account for the total amount of misclassifications. The same holds for possibility (2). As was stated in 4.1.1, the key words that were clearly classified as positively and negatively intensified showed parametric patterns, particularly segment durations and F0 movements in and around the target syllable, that go far beyond the levels of normally pitch-accented syllables. This is also true for the misclassified key words. Therefore, the misclassified key words must also have been realized with some kind of emphasis.

Possibility (3) was tested in a first approach by means of a small perception experiment. The 96 utterances that contained the misclassified key words were used as stimuli. They were judged independently by 3 subjects, i.e. 1 male and 2 female graduate students from the Linguistics Department of the University of Kiel. The 3 subjects made two binary judgements after each utterance. First, they judged whether the respective key word was or was not produced with emphasis. Second, they judged on the basis of meaning-directed native-speaker interpretation whether (a) the key word was negatively or positively intensified, or whether (b) neither of the two terms was applicable. In order to make the second judgement, the experimenter explained the communicative background of negative and positive intensification as well as of emphasis in general to the 3 graduate students. However, the explanations did not include any concrete sound examples or specifications of phonetic characteristics. The 3 subjects listened to the stimuli in separate silent rooms over headphones as often as necessary before making their judgements. Each subject received the 96 stimuli in a differently randomized order. Judgements were made by ticking boxes on prepared answer sheets that provided for each stimulus the stimulus number, the respective key word, and two boxes captioned with 'emphatic/non-emphatic' and 'positive, negative/neither nor'. The experiment took 1.5–2 h.

As expected, virtually all key words – i.e. 91 of 96 – were judged by the 3 subjects as being produced with emphasis. The remaining five key words were still judged by 2 of 2 subjects as emphatic. More crucially, however, there were 78 key words for which all 3 subjects agreed that they can be interpreted as neither positively nor negatively intensified. Forty were produced in the positive context frames. The other 38 come from negative context frames. Closer inspection by the experimenter revealed that the 40 + 38 key words can be regarded as a phonetically and functionally homogeneous subset. This subset matched with emphasis tokens that have already been noticed and described by Niebuhr [2005] with reference to speech recordings from TV shows enacting court trials. Based on contrastive functional analyses, Niebuhr [2005] argued that the emphasis tokens represented a type of intensification. However, it does not concern the positive-negative dimension, but is directed to another fundamental communicative dimension. The speaker uses this emphasis to express that the corresponding information is absolutely indisputable and hence has to be accepted unconditionally by the dialogue partner. Hence, s/he intensifies the *truthfulness* of the emphasized information without making a positive or negative judgement. Therefore, this type of intensifying emphasis was called *reinforcement* (following a suggestion of Ernst Dombrowski, IPDS, University of Kiel).

In summary, the small perception experiment suggests, in line with possibility (3), that the data of the present study was 'contaminated' by a subset of 78 key words that were similarly distributed across the positive and negative context frames, but produced with a third type of intensifying emphasis: reinforcement. As to the remaining 18 misclassifications, it must be taken into account that the presence of an unexpected third type of intensifying emphasis that is forced into the binary classifications of the discriminant analyses makes the discriminant functions less sharp and hence inevitably raises the misclassification rate, irrespective of whether the third emphasis type does or does not overlap phonetically with positive and/or negative intensification.

Therefore, the two discriminant analyses of the short- and long-vowel key words were redone on the basis of a *tripartite* classification consisting of positive intensification, negative intensification, and reinforcement. Table 4 summarizes the main results of the redone discriminant analyses. As can be seen, the performance of the statistical model improved considerably. Wilks' lambda values of less than 0.2 in combination with high eigenvalues of more than 2.0 show that the discriminant functions were much better in separating positively and negatively intensified key words, when they were detached from reinforced key words beforehand. The overall correct classification rate of the emphasis types (including reinforcement) went up from formerly 71.0% in the long-vowel or 74.4% in the short-vowel subsamples to 93.2 or 89.8%, respectively. As for positive and negative intensification, the absolute number of misclassified key words was reduced to 8 in the long-vowel and 12 in the short-vowel condition.

Table 4. Results of the discriminant analyses aiming at the tripartite differentiation between positive intensification, negative intensification, and reinforcement

Dependent variables	Discriminant analyses short-vowel key words			Discriminant analyses long-vowel key words		
	pos. int.	neg. int.	reinforcem.	pos. int.	neg. int.	reinforcem.
DURc	110.83	199.01	164.70	125.81	167.53	156.03
DURv	113.25	86.62	96.10	259.78	137.88	172.71
DURpres.	78.32	97.87	95.20	76.72	100.49	94.12
F0fallalign	109.80	8.87	78.76	264.65	34.10	115.72
F0plateau	64.87	43.54	39.50	96.33	38.91	49.67
F0shape	1.56	0.81	1.02	1.73	0.65	0.88
Ealign	72.72	38.58	54.97	212.06	69.08	107.23
Eonsetc	5.22	4.53	4.74	5.49	4.39	4.30
Eonsetv	4.60	5.84	6.12	3.73	5.30	5.50
Eoffsetv	4.66	5.74	5.53	4.51	5.48	5.57
VQh1–h2	6.71	–14.51	–2.56	9.74	–13.98	1.91
Wilks' lambda	0.17; d.f. = 22; p < 0.001***			0.07; d.f. = 22; p < 0.001***		
Eigenvalues	function1 = 2.07; function2 = 0.98			function1 = 10.85; function2 = 0.24		
Class. rates	89.8%			93.2%		

The analyses were done separately for the subsamples of key words with short (left) or long (right) vowels in their target syllables (n= 176 for each subsample). The upper part of the table presents mean values for the 11 dependent variables, the lower part shows quality measures of the discriminant functions.

Furthermore, the given mean values show that by separating out the reinforced key words the phonetic profiles of positive and negative intensification became more distinct, but remained qualitatively constant. This constancy also includes the standardized canonical discriminant coefficients that specify the weights of the parameters in the coding of positive and negative intensification. The top-five coefficients remained the same as in table 2, however, their ranking changed slightly. That is, after extracting reinforced key words, the parameters 'F0shape' and/or 'F0fallalign' that represented the slopes of the F0 peak and the fall-onset alignment became more important, at the cost of the segment duration parameters.

As regards the extracted reinforced key words themselves, it can be seen from the mean values in table 4 that reinforcement is – not only in terms of its meaning, but also in terms of its phonetic characteristics – a kind of hybrid or intermediate form of positive and negative intensification. In the (extended) durations of the onset consonant of the target syllable and the preceding speech sound, reinforcement was quite similar to negative intensification. Moreover, like negative intensification, reinforcement was also characterized by a F0 peak that shows a pointed maximum (or a relatively short high plateau) and slowly rising, quickly falling slopes (the shape indices <1). However, despite the pointed F0 peak the falling movement set in much later than for negative intensification. In fact, the mean value of the onset of the F0 fall was roughly midway between the values of positive and negative intensification. The same is true for the mean values that concern the duration of the vowel of the target syllable and the alignment of the acoustic energy maximum after the vowel onset. In terms of voice quality, the mean value of reinforcement was closer to the one of positive than of negative

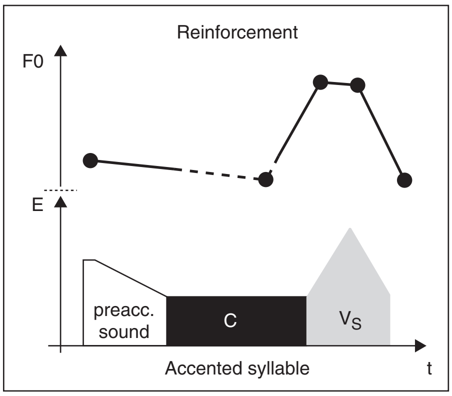

Fig. 7. Schematic illustration of the phonetic profile that emerged for reinforcement, exemplified by means of the key words with short vowels (V_S) in the accented target syllables. The shapes of the polygons in the lower panel represent the acoustic energy (E) courses. The upper panel sketches the characteristic F0 peak. The broken-line section points to the fact that some onset consonants were voiceless. All illustrations are based on the means of table 4. The displayed F0 range is oriented towards actual values, which were predominantly larger than for positive and negative intensification.

intensification, which indicates that the vowels were not produced with pressed voice. In terms of the (great) acoustic energy dynamics at the vowel boundaries of the target syllable reinforced key words resembled key words with negative intensification.

In summary, the multiparametric phonetic profile of reinforcement shows a – slightly attenuated – variant of the syntagmatic contrast in signal dynamics that was also found for negative intensification. There is a deceleration in the articulation- and phonation-induced acoustic changes towards the accented vowel, followed by a sudden acceleration after the accented-vowel onset. However, unlike in negative intensification, the deceleration-acceleration sequence of reinforcement is not linked with a suppression of sonorous features in and around the target syllable. Figure 7 provides a graphical summary of the phonetic profile that emerged for reinforcement. As all relevant differences of reinforcement to positive intensification on the one hand and negative intensification on the other are similar in the long- and short-vowel conditions, figure 7 shows only the short-vowel condition.

4.3. General Aspects

4.3.1. General Remarks on the Analysed Parameters

The subdivision of the data into two sets according to phonological vowel quantity was done since empirical studies showed that the effects of long vs. short vowels in accented target syllables are not only restricted to the obvious parameter of vowel duration. Differences in vowel quantity can also affect the duration of the preceding consonant as well as the dynamics of the F0 and acoustic energy courses in the vicinity of the accented target syllable. In fact, the MANOVA revealed that effects of the phonological vowel quantity in the target syllable were found for duration, F0, acoustic energy, and voice quality. The effects of vowel quantity do not constitute or explain the different phonetic profiles of positive and negative intensification. Rather, the effects were *superimposed* on the different profiles.

With regard to duration, MANOVA showed that the onset consonants of the target syllables had significantly greater durations before short than before long vowels. The effect of vowel quantity on the duration of the onset consonant in the present study may be explained with reference to the sequences of acceleration and deceleration in the phonetic profiles of positive and negative intensification as well as reinforcement that primarily involved vowel or onset consonant in short- or long-vowel syllables, respectively (fig. 6). As regards the acoustic energy course, the present study revealed close relationships to the temporally parallel F0 course. Most importantly, a highly significant positive correlation was found between the alignments of the acoustic energy maximum and the onset of the F0 fall relative to the vowel onset [Kohler, 1991b, p. 187; Niebuhr and Pfitzinger, 2010]. In addition, the positive correlation must have been supported by the fact that the alignment of the acoustic energy maximum was affected in the same way by vowel quantity as the alignment of the F0 fall. That is, the acoustic energy maximum occurred earlier after the vowel onset in short than in long vowels, as was found in previous studies [Mády et al., 2005].

With respect to the F0 measurements, the high plateaux of the rising-falling contours and hence also the onsets of the F0 falls were aligned later after the accented-vowel onset in long than in short vowels. Table 3 showed that vowel duration and fall-onset alignment highly positively correlated. Similar alignment shifts due to vowel quantity or duration were reported by van Santen and Hirschberg [1994] and Ladd et al. [1999, 2000]. Moreover, the positive correlations between vowel duration and shape index or F0 plateau duration match well with the data reported by Knight and Nolan [2006]. They showed that in the case of F0 plateau contours there is a more consistent alignment relationship between the fall onset and the right edge of the accented syllable than between the rise offset and the left edge of the accented syllable. Finally, the H1–H2 differences were affected by vowel quantity in a way that points to a slightly more pressed voice in the case of short vowels. It is difficult to find an explanation for this effect at the current state of research. However, similar observations have been made before [Sherman and Linke, 1952].

4.3.2. On the Phonetic Scope of Positive and Negative Intensification

The Kiel Intonation Model [Kohler, 1991a, 2006b] is built around the accented vowel. That is, different types of pitch accents are defined with reference to their timing relative to the onsets and offsets of the vowel. Experimental studies by Niebuhr [2006, 2007a] support this vowel-centred pitch accent concept, but indicate further that it is the increasing or decreasing acoustic energy rather than the spectral discontinuity at the vowel boundaries that represents the acoustic reference for pitch-accent timing. The findings of the present study also suggest that the accented vowel – and its onset in particular – are crucial timing landmarks in the expression of positive and negative intensification. However, the timing relative to the accented-vowel onset does not just concern F0 turning points (like the peak maximum), but signal dynamics in general. The accented-vowel onset separates deceleration and acceleration processes that showed up in the form of segment durations as well as F0 and acoustic energy transitions. The multiparametric coding of positive and negative intensification and of reinforcement crosses the established borderline between the segmental and the prosodic layer of the speech chain. Positive and negative intensifications are not just prosodic or even intonational phenomena. They are likewise segmental phenomena.

The current study also presented initial evidence that the expression of positive and negative intensification not only goes beyond the prosodic layer, but also beyond the accented target syllable of the key word. For example, the speech sound preceding the target syllable co-varied in duration with the onset consonant of the target syllable.

Moreover, supplementary observations show that positive and negative intensification also affected the pronunciation of speech sounds that *followed* the target syllable. For example, there were a number of sequences of adjacent velar and/or alveolar stops in the data. These sequences either formed coda clusters in negatively and positively intensified syllables like /kt/ in *schmeckt* and *stinkt*, or they went across the coda of the intensified and the onset of the subsequent syllable, as in the case of *fett geworden* (/t.g/). In the context of positive intensification the sequence-initial velar or alveolar stop was separately released and aspirated in 83.3% of the cases (n = 18). However, in combination with negative intensification there was not a single sequence with separately released and aspirated initial stops (n = 24). A further observation concerns the frequent German syllables that end in /_ən/ or /_əl/. In the present data, they occurred frequently after positively and negatively intensified syllables as, for example, in *Kotzen*, *fassen*, or *riesen, Spiegel*. It is well known that /_ən/ and /_əl/ syllables are susceptible to schwa elision, resulting in a syllabic nasal or lateral. In fact, when preceded by negatively intensified syllables, the schwas in the /_ən/ or /_əl/ syllables of the present study were elided in 89.3% of the cases (n = 56). This elision rate is slightly higher than the base rate of around 83% that was determined by Kohler [1995] by means of 5,215 items in the Kiel Corpus database. By contrast, if /_ən/ or /_əl/ occurred in the present study after positively intensified syllables, then the schwa *remained* there in 70% of the cases (i.e. elision rate = 30%; n = 40). Both the preference of a separately aspirated stop over silence and the maintenance of schwa vowels in /_ən/ or /_əl/ match well with the deceleration and enhancement of sonorous features that were claimed to characterize positive as opposed to negative intensification.

The global temporal aspects in the production and perception of intensifying emphasis must be fleshed out in future studies. This includes addressing the question of how positive and negative intensifications interact with manifestations of positive and negative emotions in speech. Emotions cause longer-term changes in physiological settings which affect the muscles and hence the biomechanical properties of the larynx and of the active articulators in the speech production apparatus. Thus, emotions influence the phonetic signal at the levels of F0, voice quality, acoustic energy, and the general dynamics of articulatory and phonatory movements [Banse and Scherer, 1996; Mozziconacci, 1998; Ní Chasaide and Gobl, 2004; Magno Caldognetto et al., 2004]. Although emotions concern the same parameters as positive and negative intensification, it is evident that the dynamically variable and complexly coordinated phonetic structures of positive and negative intensification operate in a shorter time frame than the longer-term parameter settings of positive and negative emotions. Yet, it is quite possible that such emotions cause adaptations in the signalling of positive and negative intensifications. Emphases and emotions are also clearly separable on functional grounds, i.e. emphases characterize the message, whereas emotions characterize the (state of the) speaker.

5. Conclusion

Based on the phonetic findings, the present study lends support for the basic differentiation between ‘emphasis for contrast’ and ‘emphasis for intensity’, as already suggested by Coleman [1914], Armstrong and Ward [1926] and Coustenoble and Armstrong [1934]. The meaning-based approach resulted in a refinement of the emphasis-for-intensity concept by yielding separate phonetic profiles for ‘positive’ and ‘negative intensification’.

Additionally, there are indications for a third type of intensifying emphasis, 'reinforcement'. In view of its equal distribution across the positive and negative context frames, and the outcome of a small perception experiment, reinforcement does not intensify the positive or negative valence of the corresponding word. Instead, reinforcement intensifies the truthfulness of the information that is represented by the emphasized word. Phonetically, reinforcement shows some characteristics of negative intensification, but most parameters are in between positive and negative intensification.

The term 'emphasis' has been used rather loosely in the past, covering a variety of phenomena across the fields of propositional, expressive and attitudinal meaning. For some, emphasis is restricted to contrastive emphasis of narrow focus, i.e. to propositional meaning. But such a limitation of speech functions to linguistic functions in the narrow sense does not do justice to the much broader field of speech communication in cultural, social, and situational settings where the transmission of information is constantly intertwined with expressive and attitudinal aspects of meaning. A system of emphasis must be built around such communicative meaning. From the present study, a diversity of structured form-meaning relationships has emerged constituting part of a network of emphasis in communication. Future research will have to elaborate this network. The next steps will be the specific context-guided elicitation of reinforcement in a further production study and the cross-validation of positive and negative intensification in a perception study with meaning-oriented judgements. The phonetic characteristics of contrastive emphasis should also be revisited taking the notion of reinforcement into account. Additionally, more general theoretical reflections are necessary on what types of emphasis can be expected to occur with regard to the elements (e.g., speaker and hearer) and the argumentative structure of communication.

6. Acknowledgements

Special thanks are due to Klaus J. Kohler. His ideas and his support contributed considerably to the investigation of intensifying emphasis in German. Furthermore, I am greatly indebted to the helpful and thought-provoking comments and suggestions by my two anonymous reviewers as well as by Hélène Loevenbruck, Richard Ogden, Mariapaola D'Imperio, Brechtje Post, and Martine Grice. Finally, I would like to thank Herbert Fuchs and Michael Weinhold for their support in recording the speakers, and Ernst Dombrowski and all other members of the Kiel Prosody Group for numerous fruitful discussions.

Appendix

The texts given to the subjects (accented syllables of key words are given in bold face).

1. Dialogues
A: Professor Müller hält seine Vorlesungen schon seit geraumer Zeit auf Englisch.
B: Ja, und hast Du mal auf sein Englisch geachtet?
A: Was ist damit?
B: Es hat sich kaum verbessert.
A: Findest Du? Ich finde, es hat sich enorm verbessert.

A: Hast Du Tobi schon gesehen?
B: Nein. Aber das wundert mich auch nicht. Er ist doch immer zu spät.
A: Ja, nur diesmal ist er nicht nur ein wenig zu spät. Er ist viel zu spät.

A: Karl hat gestern im juristischen Staatsexamen «voll befriedigend» bekommen.
B: Das ist aber nicht besonders gut.
A: Nein, ganz im Gegenteil, das ist sehr gut. Bei den Juristen werden höhere Noten nur sehr sehr selten vergeben.

A: Hast Du gestern die Nominierung der Nationalmannschaft für die WM gesehen?
B: Ja hab ich. David Odonkor ist ja auch dabei.
A: Genau. So eine Fehlentscheidung. Der hat doch wirklich gar kein Potential.
Das sehe ich anders. Der hat gewaltiges Potential. Er muss nur die Möglichkeit bekommen, es auch zu zeigen.

Ich hab mir gestern einen neuen Laptop gekauft, zwei Jahre gebraucht für tausend Euro. Gut, oder?
Was hast Du bezahlt? Tausend Euro? Das ist zu viel! Das ist **viel** zu viel! Wer hat Dich denn dabei bloß beraten?!

A: Gestern hat die Nationalelf 7:0 gegen Luckenwalde gewonnen. War doch eigentlich klar, oder?
B: Sicher. Aber es hat sich auch einiges in der Mannschaft getan. Die Abwehr hat sich verbessert. Die Laufbereitschaft hat sich verbessert! Und erst das Angriffsverhalten! Wow, das hat sich **enorm** verbessert! Sensationell! So können wir doch den WM-Titel holen.

A: Wirklich toll, diese filigrane Zeichnung. All diese Details! Wie machst Du das nur?
B: Nun ja, man braucht einen sehr feinen Pinsel und dazu noch eine sehr ruhige Hand. Aber das aller aller wichtigste ist **sehr** viel Geduld.

A: Heute war doch die Eröffnung des neuen IKEA-Marktes. War bestimmt ein Spektakel, oder?
B: Absolut! Mann, ich sag dir, das war ein **gewaltiges** Spektakel! Wohin das Auge reicht nur Menschen!

2. Monologues with Positive Context Frames
Das ist ja **Wahn**sinn!
Oh, da kommt unser Essen. Mmm, wie das **riecht**!
Das ist ja fan**tas**tisch!
Nicht zu glauben!
Was für eine Wohnlage. Diese **Ru**he hier!
Hm, **lec**ker! Das schmeckt!! Du bist ein wahrer **Mei**sterkoch!
Ach, wie **an**genehm, so still! Hier will ich gleich einziehen.
Ich möchte gern ein Steak. Und weißt Du was? Ein **rie**sen Stück!
Hey wow, das hast Du **toll** gemacht! Das Publikum war be**gei**stert!
Ich **lie**be diesen alten **Lim**burger. Wie das stinkt! **Herr**lich!
Mensch!! Spitze!! Du hast echt Talent als Heimwerker!
Er hat es **geschafft**. Das war eine **Mei**sterleistung!
Der Berentzen war schon lecker! Mmm…und erst der **Feig**ling!
Ah, Er ist da! **End**lich!

3. Monologues with Negative Context Frames
Das ist doch zum **Kot**zen!
Ver**dammt** noch mal!! **Pass** doch auf, wo Du hingehst!!
Oh Mann, guck dir **das** an! Sowas Be**schiss**enes hab' ich schon lang nicht mehr gesehen!
Das ist doch **ab**soluter Schwachsinn!
Oh, **Shit**! Jetzt hab' ich den Schlüssel drin gelassen, und die Tür ist zugefallen. Sowas Be**scheu**ertes kann nur mir passieren!
Uah, igitt! Schmeiß den Kram bloß weg! Schmeckt nicht! Echt **wi**derlich das Zeug!
Sieh dich doch mal im Spiegel an. Das ist ja **ek**lig. Du bist richtig **fett** geworden.

Ich hab' dein Gerede so **satt**! Sei bloß still!! Sonst ver**gess**' ich mich!!
Weißt Du was Du bist? Ein **mie**ses Stück!
Jetzt ist die **scheiß** Karre endgültig kaputt. Das hast Du **toll** gemacht! Ich bin echt be**gei**stert!
Sag mal, hast Du in 'ner Klärgrube gebadet?? Boa, das stinkt! Zum **Kot**zen!!
Jetzt sitzt der Karren echt im Dreck. Na spitze! Und was jetzt?
Er hat das **Au**to ruiniert. Nicht zu **fass**en! Das war eine echte **Mei**sterleistung!
Näh, du bist ein **wi**derlicher **Feig**ling!
Oh Mann, na **end**lich!! Das wurde aber auch Zeit! Hast Du eine **Ah**nung, wie lange ich hier schon stehe!? Wo hast Du gesteckt?!

References

Andreeva, B.; Barry, W.; Steiner, I.: Producing phrasal prominence in German. Proc. 16th ICPhS, Saarbrücken 2007, pp. 1209–1212.

Armstrong, L.E.; Ward, I.C.: A handbook of English intonation (Heffer, Cambridge 1926).

Banse, R.; Scherer, K.R.: Acoustic profiles in vocal emotion expression. J. Personality social Psychol. *70*: 614–636 (1996).

Baumann, S.; Becker, J.; Grice, M.; Mücke, D.: Tonal and articulatory marking of focus in German. Proc. 16th ICPhS, Saarbrücken 2007, pp. 1029–1032.

Baumann, S.; Grice, M.; Steindamm, S.: Prosodic marking of focus domains – categorical or gradient? Proc. 3rd Int Conf. Speech Prosody, Dresden 2006, pp. 301–304.

Boersma, P.: Praat, a system for doing phonetics by computer. Glot int. *5*: 341–345 (2001).

Braun, B.; Ladd, D.R.: Prosodic correlates of contrastive and non-contrastive themes in German. Proc. 8th Eur. Conf. on Speech Commun. Technol., Geneva 2003, pp. 789–792.

Coleman, H.O.: Intonation and emphasis. Miscellanea phonet. *1*: 6–26 (1914).

Cooper, W.E.; Eady, S.J.; Mueller, P.R.: Acoustical aspects of contrastive stress in question-answer contexts. J. acoust. Soc. Am. *77*: 2142–2156 (1985).

Coustenoble, H.N.; Armstrong, L.E.: Studies in French intonation (Heffer, Cambridge 1934).

Dahan, D.; Bernard, J.-M.: Interspeaker variability in emphatic accent production in French. Lang. Speech *39*: 341–374 (1996).

del Giudice, A.; Shosted, R.; Davidson, K.; Salihie, M.; Arvaniti, A.: Comparing methods for comparing pitch 'elbows'. Proc. 16th ICPhS, Saarbrücken 2007, pp. 1117–1120.

Dohen, M.; Loevenbruck, H.: Pre-focal rephrasing, focal enhancement and post-focal deaccentuation in French. Proc. 8th Int. Conf. on Spoken Lang. Processing, Jeju Island 2004, pp. 313–316.

Dohen, M.; Loevenbruck, H.; Cathiard, M.-A.; Schwartz, J.-L.: Identification of the possible visible correlates of contrastive focus in French. Proc. 2nd Int. Conf. Speech Prosody, Nara 2004, pp. 77–80.

Fischer-Jørgensen, E.: Phonetic analysis of breathy (murmured) vowels in Gujarati. Indian Linguistics *28*: 71–139 (1967).

Geumann, A.: Vocal intensity: acoustic and articulatory correlates. Proc. 4th Conf. on Motor Control, Nijmegen 2001, pp. 70–73.

Gili Fivela, B.: Tonal alignment in two Pisa Italian peak accents. Proc. 1st Int. Conf. Speech Prosody, Aix-en-Provence 2002, pp. 339–342.

Hanssen, J.; Peters, J.; Gussenhoven, C.: Prosodic effects of focus in Dutch declaratives. Proc. 4th Int. Conf. Speech Prosody, Campinas 2008, pp. 609–612.

Hermes, A.; Becker, J.; Mücke, D.; Baumann, S.; Grice, M.: Articulatory gestures and focus marking in German. Proc. 4th Int. Conf. Speech Prosody, Campinas 2008, pp. 457–460.

Hirst, D.: The phonology and phonetics of speech prosody: between acoustics and interpretation. Proc. 2nd Int. Conf. Speech Prosody, Nara 2004, pp. 163–169.

Igarashi, Y.: How many falling intonation patterns in Russian? Categories of F0 alignment. Proc. IIAS Conf. Between Stress and Tone, Leiden 2005, pp. 32–34.

Iriondo, I.; Planet, S.; Claudi Socoró, J.; Alías, F.; Monzo, C.; Martínez, E.: Expressive speech corpus validation by mapping subjective perception to automatic classification based on prosody and voice quality. Proc. 16th ICPhS, Saarbrücken 2007, pp. 2125–2128.

Keating, P.A.; Eposito, Ch.: Linguistic voice quality. UCLA Working Papers Phonet. *105*: 85–91 (2007).

Klatt, D.H.: Synthesis by rule of segmental duration in English sentences; in Lindblom, Öhman, Frontiers of speech communication research, pp. 287–299 (Academic Press, London 1979).

Klatt, D.H.; Klatt, L.C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. acoust. Soc. Am. *87*: 820–857 (1990).

Knight, R.-A.; Nolan, F.: The effect of pitch span on intonational plateaux. J. Int. Phonet. Ass. *36*: 21–38 (2006).

Kohler, K.J.: Prosody in speech synthesis: the interplay between basic research and TTS application. J. Phonet. *19*: 121–138 (1991a).

Kohler, 1991b

Kohler, K.J.: Einführung in die Phonetik des Deutschen (Schmidt, Berlin 1995).
Kohler, K.J.: Neglected categories in the modelling of prosody: pitch timing and non-pitch accents. Proc. 15th ICPhS, Barcelona 2003, pp. 2925–2928.
Kohler, K.J.: Form and function of non-pitch accents. Arbeitsber. Inst. Phonetik dig. Sprachverarb. (AIPUK) *35a*: 97–123 (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 of experimental prosodic analysis: from measurement to function; in Sudhoff, Lenertová, Meyer, Pappert, Augurzky, Mleinek, Richter, Methods in empirical prosody research, pp. 123–152. Language, Context, and Cognition 3 (De Gruyter, Berlin 2006b).
Kohler, K.J.: The perception of prominence patterns. Phonetica *65*: 257–269 (2008).
Kohler, K.J.; Peters, B.; Scheffers, M.: The Kiel corpus of spontaneous speech, vol. IV. German: Video Task Scenario (Kiel DVD No. 1). http://www.ipds.uni-kiel.de/kjk/ pub_exx/kk2006_6/InfoDVD1.pdf (2006).
Krahmer, E.; Swerts, M.: On the alleged existence of contrastive focus. Speech Commun. *34*: 391–405 (2001).
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).
Lehiste, I.: Suprasegmentals (MIT Press, Cambridge 1970).
Mack, M.; Gold, B.: The effect of linguistic content upon the discrimination of pitch in monotone stimuli. J. Phonet. *14*: 333–337 (1986).
Mády, K.; Tronka, K.Z.; Reichel, U.D.: Syllable cut and energy contour in vowels: a comparative study on German and Hungarian. ZAS Papers Ling. *42*: 197–217 (2005).
Magno Caldognetto, E.; Cosi, P.; Drioli, C.; Tisato, G.; Cavicchio, F.: Modifications of phonetic labial targets in emotive speech: effects of the co-production of speech and emotions. Speech Commun. *44*: 173–185 (2004).
Möbius, B.; van Santen, J.P.H.: Modeling segmental duration in German text-to-speech synthesis. Proc. Int. Conf. on Spoken Lang. Processing, Philadelphia 1996, pp. 2395–2398.
Mozziconacci, S.: Speech variability and emotion: production and perception; PhD thesis TU Eindhoven, Eindhoven (Printers of the University, 1998).
Ní Chasaide, A.N.; Gobl, C.: Voice quality and f0 in prosody: towards a holistic account. Proc. Int. Conf. Speech Prosody, Nara 2004, pp. 189–196.
Niebuhr, O.: Die Diskussion zur Emphaseforschung mit besonderer Berücksichtigung des Druckakzentphänomens. http://www.ipds.uni-kiel.de/on/downloads/EMPH_HTML/ Emphase.html (2005).
Niebuhr, O.: The role of the accented-vowel onset in the perception of German early and medial peaks. Proc. 3rd Int. Conf. Speech Prosody, Dresden 2006, pp. 109–112.
Niebuhr, O.: Categorical perception in intonation: a matter of signal dynamics? Proc. 8th Interspeech Conf., Antwerp 2007a, pp. 642–645.
Niebuhr, O.: Perzeption und kognitive Verarbeitung der Sprechmelodie: theoretische Grundlagen und empirische Untersuchungen; in Steube, Language, context, and cognition, vol. 7, pp. 1–430 (de Gruyter, Berlin 2007b).
Niebuhr, O.; Kohler, K.J.: Perception and cognitive processing of tonal alignment in German. Proc. Int. Symp. on Tonal Aspects of Lang. (TAL): Emphasis on Tone Languages, Beijing 2004, pp. 155–158.
Niebuhr, O.; Pfitzinger, H.P.: On pitch-accent identification: the role of syllable duration and intensity. Proc. 5th Int. Conf. Speech Prosody, Chicago 2010.
Petrone, C.; D'Imperio, M.: Is tonal alignment interpretation independent of methodology? Proc. 10th Interspeech 2009, Brighton 2009, pp. 2459–2462.
Rosen, S.; Fourcin, A.: Frequency selectivity and the perception of speech; in Moore, Frequency Selectivity in Hearing, pp. 373–487 (Academic Press, London 1986).
Sherman, D.; Linke, E.: The influence of certain vowel types on degree of harsh voice quality. J. Speech Hear. Disorders *17*: 401–408 (1952).
Sievers, E.: Grundzüge der Phonetik. Bibliothek indogermanischer Grammatiken 1 (Breitkopf & Härtel, Leipzig 1901).
Silverman, K.E.; Pierrehumbert, J.B.: The timing of prenuclear high accents in English; in Kingston, Beckman, Papers in Laboratory Phonology I, pp. 72–106 (Cambridge, Cambridge University Press, Cambridge 1990).
Simpson, A.P.: Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung. Arbeitsber. Inst. Phonetik dig. Sprachverarb. (AIPUK) *33*: 1–233 (1998).
Strangert, E.: Emphasis by pausing. Proc. 15th ICPhS, Barcelona 2003, pp. 2477–2480.
van Santen, J.P.H.; Hirschberg, J.: Segmental effects on timing and height of pitch contours. Proc. Int. Conf. on Spoken Lang. Processing, Yokohama 1994, pp. 719–722.
van Santen, J.P.H.: Quantitative modelling of segmental duration. Proc. Workshop on Human Lang. Technol., Princeton 1993, pp. 323–328.
Xu, Y.; Xu, C.X.: Phonetic realization of focus in English declarative intonation. J. Phonet. *33*: 159–197 (2005).