# What is Emphasis and How is it Coded?

Klaus J. Kohler

Institute of Phonetics and Digital Speech Processing (IPDS) Christian-Albrechts-University at Kiel, Germany kjk@ipds.uni-kiel.de

# Abstract

The meaning category emphasis is examined with regard to its semantic, pragmatic, and affective components and their prosodic coding in German, English, and Dutch. In particular, a distinction is made between emphasis for focus, which singles out elements of discourse by making them more salient than others, and emphasis for intensity, which intensifies the meaning contained in the elements. To evaluate intensity negatively a force accent comes into play, which is signalled by non-pitch features. The question of universals is also addressed. Sound illustrations may be found in [19].

#### 1. Introduction

The term *emphasis* is in current use for multivaried phenomena that do not constitute a single unitary category. Ladd [15] repeatedly refers to "focus and emphasis", equating emphasis with narrow focus (e.g. p. 9), and saying that "unambiguous narrow focus pronunciations involve what may be called 'emphatic stress", and that there is the "paralinguistic possibility of gradiently modifying the realisation of ...patterns so as to single out individual words" (pp. 200f). This is the use of the term for highlighting elements of utterance by giving them special prominence; scaling H and L tones signals degrees of emphasis without affecting the linguistic identity of the contour (p. 39). This view refers back to the analyses of the Anna/Manny type by Pierrehumbert [18], and Liberman and Pierrehumbert [17]. Ladd sees emphatic peak raising in the paralinguistic vs linguistic and gradient vs categorical dichotomies and stipulates that "in languages like English, linguistic accent location and paralinguistic degree of emphasis even though they are both suprasegmental – should be distinguished" (p. 202). Ladd and Morton [16] investigate this issue of paralinguistic gradiency further in connection with peak height contrast.

But beside this reference to singling out and contrasting, Ladd also uses emphasis in a different sense. In referring to an additional emphatic falling declarative accent on an unstressed final syllable in Italian he concludes that "this adds emphasis or conveys some sort of special emotional involvement on the part of the speaker" (p. 129). This addresses the affective component of meaning, which Gussenhoven and Rietveld [7] investigated along the semantic scale of surprise for Dutch.

The problem with Ladd's use of emphasis is that it is not clearly defined under semantic, pragmatic, and affective aspects, and that he does not examine the acoustic cues to emphatic stress because he does not regard them as relevant in his evaluation of linguistic H and L tone constellations (p. 200). A renewed, clarifying discussion of the issue of emphasis needs to analyse the function - signal relationship, taking all communicative functions into account, linguistic and paralinguistic, and to stop analysing the phenomena primarily under the categorical vs gradient dichotomy.

Gussenhoven [6], Chen, Gussenhoven, and Rietveld [4], and Chen [3] remedied these shortcomings by introducing a terminological differentiation between the informational and the affective components of the Effort Code, which they named emphatic and surprised, respectively. They systematically varied the acoustic parameters of F0 peak height and F0 peak alignment in sets of test stimuli for Dutch and English listeners to mark perceived degrees of informational or affective interpretation along the semantic scales EMPHATIC\_NOT EMPHATIC and SURPRISED\_NOT SURPRISED. There is a positive correlation between peak height and the degree of emphaticness, but also between peak height and the degree of surprise; on the other hand, the factor of peak alignment is significant on the *surprise* scale.

However, five problems persist:

(1) the unproven assumption that listeners associate a uniform semantics with the scale name emphasis, which in popular understanding can refer either to informational highlighting or to expressive evaluation (cf. [11]);

(2) the premature association of the categories emphasis and surprise, and their acoustic manifestations, to a biological Effort Code at the outset of the investigation, instead of deducing patterns from the results that may be linked to an externally defined code;

(3) the linguistic : paralinguistic and discrete : gradient dichotomies;

(4) the lack of a comprehensive account of the semantic, pragmatic, and affective domains of accentuation, of which the authors' emphasis is one component;

(5) a more detailed account of the phonetic parameters that code these functions of accentuation, including articulatory properties, their energy and timing, beside F0.

This paper is concerned with problems (3) to (5).

# 2. Accent coding

#### 2.1. Lexical stress and sentence accent

Before the coding of meaning categories, such as *emphasis*, by accentual features can be solved successfully, we need to define accent as an element in the prosodic systems of individual languages, and we have to relate it to the acoustic properties that can signal it for a listener, and to the physiological and articulatory antecedents in the speaker's production. Among the languages of the world, there is a group including the Germanic and Slavonic languages, Italian and Spanish, but not French, among the Romance languages, which have lexical stress, i.e. the abstract phonological specification of a syllable as a position in the word that acts as a placeholder for acoustic properties when the word is *accented* in the utterance.

It does not make sense to ask what the phonetic exponents of word stress are, because they are only accessible through accent at the utterance level, e.g. in the common metalinguistic frame of isolated word citation forms, and therefore vary with

the selected utterance frame. *Lexical stress* and *sentence accent* are thus to be differentiated theoretically and terminologically in prosodic analysis. The term *sentence accent* refers to the perceptual salience of some words over others in utterances, and is not to be confused with *phrase accent* as a phrase boundary marker in AM phonology. This relational perceptual salience of *sentence accents* is linked to semantic, pragmatic, and affective structuring of speech communication, at the levels of the individual language, of typological language groups, and of language universals.

The Germanic languages use *sentence accents* differently from, e.g., French. In Dutch, English, or German, they can highlight individual words by melodic excursions for contrast with others, which is not the general pattern in French, where syntactic means take over. Furthermore, French does not only lack the category of *lexical stress*, but also the concept of a phonetic word: it is therefore the phonetic phrase that receives a melodic accent to indicate syntagmatic structuring. The highlighting of words within phrases is not primarily achieved by means of pitch inflection but by the *accent d'insistance*, which relies on initial consonant lengthening and acoustic energy in the first instance.

#### 2.2. Accent coding by pitch features

In the West Germanic languages, *sentence accent* is not a binary decision present vs absent, but comprises at least four distinctive levels, *unaccented*, *default accented*, *partially deaccented*, *reinforced*, which are labelled as **/0**, **2**, **1**, **3**/ in the *Kiel Intonation Model (KIM)* [9].These four levels may be illustrated by the following contextualizations of the test sentence "Mary/Anna came with Manny".

- Speaker A: "Peter invited a few of his friends to a party in his flat. Mary came with Manny." : both *Mary* and *Manny* get default sentence accent.
- Speaker B: "No, Anna came with Manny": only *Anna* is fully accented, *Manny* is either completely deaccented if the contrast between the two women is all Speaker B wants to highlight, or it is partially deaccented if Speaker B picks up Speaker A's background information as relevant for the statement. The two versions are clearly distinct perceptually and in their pragmatic connotation.
- In both versions of Speaker B's comment, *Anna* may get a reinforced accent to highlight the informational contrast more strongly, and this reinforced level is categorially different from the default level in perception and pragmatic connotation, but may itself be graded according to the degree of contrast the speaker wants to convey.

The phonetic manifestations of the default and reinforced *sentence accent* levels are primarily signalled by F0 peak contours of varying height and may therefore be called *pitch accents*. The partially deaccented level has its acoustic exponents primarily in the duration domain although it may be accompanied by an F0 peak inflection of a magnitude that is well below the F0 peak declination, and, of course, also by higher energy. This accent may be called *duration accent*. On the other hand, a *pitch accent*, particularly at the *reinforced* level, is accompanied by accented-syllable lengthening, more particularly its nucleus.

#### 2.3. Accent coding by non-pitch features

The analysis of the *Kiel Corpus of Spontaneous Speech* [8] and perceptual tests following on from these production data have shown [12] that there is a third type of *sentence accent* 

that has to be recognised which is primarily related to nonpitch features, viz. acoustic energy based on phonatory and articulatory force, and may therefore be called *force accent*. It is signalled by a large increase in the duration and the acoustic energy of the accented-syllable onset, as well as by nonmodal, breathy or creaky, tight voice quality in the accented word [10, 12]. The force accent strengthens the voiceless and non-modal sections of speech; in many cases F0 cannot be analysed at all. This contrasts with the strengthening of the sonorous sections of speech by a pitch accent, and lengthening is more extreme under a force accent. Another type of force accent occurs when the syllable nucleus is also reinforced in addition to the onset, but with lax voice quality and pitch features.

### 3. Communicative functions of accents

#### 3.1. The meaning of emphasis

Following Armstrong and Ward [1] and Coustenoble and Armstrong [2], two main categories of *emphasis* need to be distinguished: (1) special prominence which a speaker gives to certain words for rational highlighting and expression of contrast to what has been said; (2) special prominence to amplify the meaning of words and to express a particularly great degree of what they imply. The former has been called *emphasis for contrast*, the latter *emphasis for intensity*.

The authors illustrate the difference with the example *There's an enormous improvement*. Contrasting the word *enormous*, for example to a preceding utterance *There was very little improvement.*, results in a focus on the word, with a high pitch fall on the only accented syllable *-nor-*. Contrast may be scaled by different ranges of the fall. The expression of intensity on the same word, on the other hand, produces an upward pitch glide with levelling out on *-nor-*, followed by a gradual descent on the subsequent syllables and a nuclear accent on *-prove-*; there is also greater energy and lengthening of the accented-syllable onset. Emphasis for intensity may be scaled by widening or narrowing the pitch range, and the other features may be varied as well for this scaling.

#### 3.2. Emphasis for contrast in pitch accents

#### 3.2.1. Peak contours

The coding of pragmatic meaning by peak contours depends on their synchronization with the vocal tract dynamics. *KIM* has established four distinctive positions for German, which equally apply to English – *early, medial, late medial* and *late* [14]. Whereas *early* implies 'conclusion of an argument', the others refer to 'opening an argument', i.e. they can be used to signal *contrast*, but with varying connotations. This can be illustrated with the German and English utterances *Er war mal mager./He used to be slim.* in the situational context of looking at old photos of acquaintances.

- With a *medial* peak contour, which falls from the maximum F0 point shortly after accented-vowel onset, the speaker singles out information from among factual possibilities.
- With a *late medial* peak contour, which has a substantial rise into the accented vowel, the speaker contrasts an observation with what s/he knows, but signals rational acceptance of this contrast. The contrast can be scaled by raising the peak point.

• With a *late* peak contour, where the rise occurs later and is preceded by a stretch of low-level F0 in the accented vowel, the speaker adds a personal affective evaluation to the contrast. This imparts the meaning of an exclamation *Well*, *I never!* The contrast and the affective component can be scaled by raising the peak point.

These different *contrast* connotations are all speaker oriented, i.e. the speaker does not ask for the listener's opinion on the assessment of the contrast [13,14].

#### 3.2.2. Valley contours

The pragmatic meaning of valley contours is also linked to their synchronization with the vocal tract dynamics. *KIM* has established two distinctive positions for German, which equally apply to English – *early* or *late*. In addition, there are signal – function differences for convex vs concave shapes of the contours in the accented syllable [5]. These differences can be illustrated by the German and English utterances *Er ist in Rome*?/*He is in Rome*? (Speaker B) in the situational context of Speaker A saying *Er ist nach Rom gefahren./He's gone to Rome*.

- With an *early* valley contour, which rises, in a convex trajectory, from a minimum F0 point before the accented-vowel onset to a high phrase-final end point, Speaker B asks a surprise echo question, without personal affective evaluation: *It sounds unexpected; confirm that I have understood it correctly.*
- With a late valley contour, which rises, in a concave trajectory, from a low F0 stretch inside the accented-vowel to a high phrase-final end point, Speaker B asks a surprise echo question with personal affective evaluation: *I can't believe it!*

These different surprise connotations are all listener oriented, i.e. the speaker asks for the listener's response to the speaker's surprise [13]. The patterns represent what Gussenhoven and Rietveld [7] investigated in Dutch as a difference between H\*H-H% and L\*H-H% in AM/ToBI notation. We are dealing here with two different categories of surprise, one factual because there is a clash with the speaker's knowledge, the other affective because the speaker expresses incredulity and disagreement. The use of a single surprise scale to assess a degree of affect in both types of utterances, which was the authors' experimental design, can therefore not capture the meaning categories coded by the two valley types. Furthermore, surprise coded by a valley contour, as in [7], is to be differentiated from surprise coded by a peak contour, as in [14]. The former is addressed to a listener to be resolved; the latter stays with the speaker and has an exclamatory value. These features apply to the West Germanic group of languages, Dutch, English, and German.

#### 3.3. Emphasis for intensity in pitch accents

The pitch accent coding of *emphasis for intensity* differs from the one for *contrast*, at least in German and English. The pitch contour is of the peak type, but the internal timing is quite different. Pitch rises to a plateau and then descends more slowly. This timing difference is enhanced by accentedsyllable onset lengthening for the F0 rise and a levelling out on the nucleus before the slow descent, e.g. in *Er ist eine Seele von Mensch./He's an absolute gem*. Furthermore, the voice quality may change to breathy voice, e.g. in *Reizend./It's lovely*. The intensification of the word meaning may even be achieved by a narrowing of the pitch range rather than an expansion. In all cases, this *emphasis for intensity* highlights the positive meaning of words by strengthening their sonorous sections.

### 3.4. Emphasis for Intensity in non-pitch accents

The typical non-pitch force accent in 2.3 codes emphasis for intensity, but adds a negative, disapproving expressive evaluation to the statement the speaker is making. The English and German examples qoted in 3.1 and 3.3 could not be produced with a non-melodic force accent because it would destroy the speaker's positive expressiveness. On the other hand, a strong negative intensity conveyed by *Er ist ein Schwein!*/ *He's an absolute bastard*! would be most effectively signalled by a non-pitch force accent with an increased non-sonorous onset and a breathy nucleus, highlighting the disgust the speaker wants to express on the person's behaviour he is talking about.

Similarly, it would be appropriate to have melodic accents for positive intensity emphasis in English *It's delicious!*, *It smells!* or in the German equivalents *Das schmeckt!*, *Das riecht!*, in each case with strong lengthening of the sonorants of the accented-syllable onsets in addition to the tonal structures decribed in 3.3. But it would be odd to have the same prosodic patterns in *It stinks!* or *Das stinkt!*, unless the speaker relishes in odours that are unpleasant for others. The latter two examples are expected to have non-melodic force accents. In *You did that beautifully!* or *Das hast du toll gemacht!*, a melodic force accent expresses appreciation, whereas a non-melodic force accent sounds sarcastic because the verbal meaning clashes with the accentual meaning, and the latter is rated higher.

# 4. A function – signal framework for emphasis

The discussion of the various types of emphasis in examples from West Germanic languages in Section 3 shows the need to devise a general function – signal framework, which lays out the semantic, pragmatic, and expressive net for what has been called *emphasis*, and relates the functional categories to properties of the acoustic signal and its phonatory and articulatory antecedents. In the first instance, we need a consistent terminology. There is no harm in continuing to label the whole field *emphasis*, but its subcategories must be named with reference to externally motivated categories of meaning.

The basic division seems to be the one proposed by Armstrong, Coustenoble, and Ward, between singling out elements of discourse by making them more salient than others, on the one hand, and intensifying the meaning contained in the elements, on the other. But I suggest to rename this fundamental dichotomy as FOCUS vs INTENSITY. The West Germanic languages handle both categories in the main by prosodic and articulatory means. But syntactic devices can intervene for FOCUS, e.g., in French, and the languages of the world may divide the load of coding between the prosodic/articulatory and morphological/syntactic domains quite differently, but still maintain the basic meaning distinction, which I would thus regard as a functional language universal. Moreover, the coding of INTENSITY by non-pitch accents for negative expression is more wide-spread than just in the West Germanic languages. It is, for example, also found in spontaneous French, e.g. Alors tu sais, c'est tordu ici., "Well, you know that's crazy here." with a strong accent d'insistance on tordu, signalled by lengthening and aspiration of [t], and a general decrease of sonority and modal voice quality in the intensified word.

INTENSITY then divides into POSITIVE vs NEGATIVE, which in German and English are coded by pitch vs non-pitch accents, strengthening sonorous or voiceless signal sections, respectively. Voice qualities and onset lengthening play an important role.

FOCUS is either speaker or listener-oriented, the former signalling INFORMATION SELECTION, OF FACTUAL CONTRAST, OF AFFECTIVE CONTRAST, the latter FACTUAL SURPRISE or AFFECTIVE SURPRISE. In the West Germanic languages, these meaning distinctions are coded by differently synchronized peak or valley contours. In other languages the distribution of formal coding features will no doubt be different, but the meaning distinctions are expected to be generally valid.

The five categories can be illustrated by the German and English questions **Wo**?/**Where**? (Speaker B) in the situational context of Speaker A saying *Wir treffen uns morgen in Hanerau*?/*We'll meet tomorrow in Auchterarder*. (small towns, not widely known, in Schleswig-Holstein, Germany, and Fife, Scotland, respectively). A *medial peak* on the question word asks for more information about the specific place in the town. A *late medial peak* not only asks for more information but also points out the insufficiency of the information already provided. A *late peak* adds an affective component of irritation. An *early high-rising valley* indicates surprise at hearing an unexpected name and asks for repetition. Finally, a *late high-rising valley* gives the surprise affective colouring.

# 5. Conclusions

The function – signal framework proposed in this paper needs further refinement and more detailed analysis by combining two methodological approaches:

- analysing large databases of spontaneous speech in a variety of scenarios with respect to the sub-categories of *emphasis* and their acoustic manifestations in the domains of prosody and vocal tract dynamics
- experimental analysis of selected *emphasis* categories in production and perception, by systematic contextualization and variation of phonetic parameters, e.g. the separation, in perception and interpretation, of the various peak and valley patterns in segmentally identical chains [7,11], using, for example, the semantic differential technique [11].

In the case of stimulus generation for perception experiments, high-quality synthesis programs will have to be applied to produce natural voice quality control over and above F0 and segmental manipulation, especially in the area of non-pitch force accents.

Moreover, this investigation of *emphasis* needs to be extended to many more languages with different prosodic structures, including the Romance and the Slavonic languages, Semitic languages, tonal accent languages, such as Swedish and Japanese, and tone languages of Africa and Asia, to arrive at typologies and universals of language and speech in the function – signal relationship of *emphasis*.

Especially the force accent data suggest that the phonetic manifestations signalling the two categories of *emphasis for negative* and *positive intensity*, represent another instance of a direct function – signal link. It is an interesting research goal to investigate this link further from both form and function perspectives in production as well as in perception across a variety of languages. Particularly the integrated investigation of pitch, phonation, articulation and their timing in the signal-ling of the communicative functions of force accents can give

new insights and advance phonetic science beyond the impoverished prosodic modelling within laboratory phonology.

# 6. References

AIPUK = Arbeitsberichte, Institut für Phonetik und digitale Sprachverarbeitung, Universität Kiel

- [1] Armstrong, L. E.; Ward, I. C., 1926. A Handbook of English Intonation. Cambridge: Heffer.
- [2] Coustenoble, H. N.; Armstrong, L. E., 1934. *Studies in French Intonation*. Cambridge: Heffer.
- [3] Chen, A., 2005. Universal and language-specific perception of paralinguistic intonational meaning. PhD thesis Nijmegen. Utrecht: LOT.
- [4] Chen, A.; Gussenhoven, C.; Rietveld, T., 2002. Languagespecific uses of the effort code. In B. Bel and I. Marlien (eds.), *Proc of Speech Prosody 2002*, Aix-en-Provence: Université de Provence, 211-214.
- [5] Dombrowski, E.; Niebuhr, O., 2005. Acoustic patterns and communicative functions of phrase-final F0 rises in German: activating and restricting contours. *Phonetica* 62, 176-195.
- [6] Gussenhoven, C., 2002. Intonation and interpretation: phonetics and phonology. In B. Bel and I. Marlien (eds.), *Proc of Speech Prosody 2002*, Aix-en-Provence: Université de Provence, 47-57.
- [7] Gussenhoven, C.; Rietveld, T., 2000. The behaviour of H\* and L\* under variations in pitch range in Dutch rising contours. *Language and Speech* 43(2), 183-203.
- [8] IPDS, 1995. The Kiel Corpus of Spontaneous Speech. vols 1, 2 CD-ROM#2,3. IPDS: Kiel.
- [9] Kohler, K. J., 1997. Modelling prosody in spontaneous speech. In Y. Sagisaka, N. Campbell, and N. Higuchi (eds.), *Computing Prosody. Computational models for* processing spontaneous speech. New York: Springer, 187-210.
- [10] Kohler, K. J., 2003. Neglected categories in the modelling of prosody: pitch timing and non-pitch accents. In *Proc. 15<sup>th</sup> ICPhS*, Barcelona. 2925-2928.
- [11] Kohler, K. J., 2005. Timing and communicative functions of pitch contours, *Phonetica* 62, 88-105.
- [12] Kohler, K. J., 2005. Form and function of non-pitch accents. *AIPUK* 35a, 97-123.
- [13] Kohler, K. J., 2005. Pragmatic and attitudinal meanings of pitch patterns in German syntactically marked questions. *AIPUK* 35a, 125-142.
- [14] Kohler, K. J., 2006. Paradigms of experimental prosodic analysis – from measurement to function. In S. Sudhoff et al (eds), *Methods in Empirical Prosody Research*. Berlin, New York: de Gruyter (= Language, Context, and Cognition, 3).
- [15] Ladd, D. R., 1996. *Intonational phonology*. Cambridge: CUP.
- [16] Ladd, D. R.; Morton, R., 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25, 313-342.
- [17] Liberman, M.; Pierrehumbert, J., 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.), *Language sound structure*. Cambridge, MA: MIT Press, 157-233.
- [18] Pierrehumbert, J., 1980. *The phonology and phonetics of English intonation*. PhD thesis, MIT.
- [19] <u>www.ipds.uni-kiel.de/kjk/publikationen/audiobsp.en.html</u> Sound illustrations of examples discussed in this paper.