THE PHONETICS OF EMPHASIS

Klaus J. Kohler & Oliver Niebuhr

Instiitute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Kiel, Germany
kjk AT ipds.uni-kiel.de ; on AT ipds.uni-kiel.de

ABSTRACT

Research is reported in a framework linking phonetic exponents to communicative functions. From the heterogeneous field of 'emphasis', two areas are selected: 'positive/negative expressive intensification' of verbal meaning, e.g. *it's delicious!* vs *it stinks!* German data are collected in controlled monologues and dialogues. On the hypothesis that 'positive emphasis' strengthens sonority, 'negative emphasis' weakens it, aspects of f0, acoustic energy, duration, voice quality are tested statistically.

1. INTRODUCTION

The term 'emphasis' in current use covers a wide spectrum of functionally different phenomena [8]:
(1) 'information selection' – special prominence for rational highlighting of certain words, e.g. *ANNA came with MANNY* (generally referred to as 'narrow focus') [3,9,10]
(a) 'factual' – simple singling out, in English or German by pitch accent on a word and simultaneous deaccentuation around it
(b) 'weighted' – degree of importance, in English or German signalled by f0 range, e.g. *I'm telling you ANNA came with Manny*
(c) 'insisting' – reinforcement, correction, contradiction, by initial consonant strengthening in addition to pitch features of (a) and (b), e.g. *no, MARY came with Manny*

(2) 'contrast to one's expectation' – degree of affective evaluation of a discrepancy between observed fact and expectation, in English or German signalled by medial to late f0 peak synchronization with the accent syllable, e.g. *he used to be SLIM* [7]

(3) 'expressive intensification' – special prominence for amplifying the verbal meaning [1,2]
(a) 'positive' – expression of pleasure, likely to be signalled by strengthening sonorous features of the accented syllable, especially nucleus lengthening, e.g. *it's deLIcious!*
(b) 'negative' – expression of dislike, by weakening sonorous features of the accented syllable, initial consonant lengthening at the expense of the nucleus, e.g. *it STINKS!* ('force accent' [5,6]).

This paper deals with the phonetic manifestation of emphasis in sense (3) in Standard German.

2. HYPOTHESES

On the general likelihood that 'positive emphasis' (**P**) strengthens, 'negative emphasis' (**N**) weakens sonorous features, specific hypotheses are:

1. ***Intonation***
In both **P** and **N,** rising-falling peak contours are expected: **P** is marked by rising into the accented vowel to a high f0 level, **N** by falling to a low f0 level in the vowel**.** Therefore,
1.1 peaks tend to be synchronized *non-early* for **P**, *non-late* for **N**;
1.2 preaccentual f0 concatenation tends to be *dipped* for **P**, *not low-dipped* for **N**;
1.3 postaccentual f0 concatenation tends to be *not low-dipped* for **P**, *low-dipped* for **N**;
1.4 The semitone range is larger for a **P** than for an **N** accent, due to a higher f0 level, and the standard deviation for the F0 values within the **P** contour is smaller, due to an f0 plateau.

2. ***Acoustic energy*** intensifies the accented syllable nucleus in **P,** the onset in **N**. Therefore,
2.1 the maximum in the vowel is reached later after the vowel onset for **P** than for **N**;
2.2 the average energy in onset consonants is smaller for **P** than for **N**.

3. ***Duration***
For comparable syllable structures in the two functional classes, the duration ratio of the initial consonant (cluster) and the voiced rhyme is smaller for **P** than for **N**.

4. ***Voice quality*** added to sonority aspect
Due to breathyness for **P**, the H1/H2 ratio is smaller for **P** than for **N**, resulting in a greater spectral tilt.

3. METHOD

3.1. Data acquisition scenario

To test the hypotheses, a speech database was necessary, containing a sufficient and easily accessible number of cases for the different functional types of emphasis. For comparative analyses of the data, control of segmental and prosodic structures of corresponding utterances was also required, and the data needed to be as natural as possible.

To meet these requirements, a hypothesis-driven, function-based elicitation of different types of emphasis was carried out. The basic idea is that utterances are designed and arranged in written texts to provide a linguistic and situational context frame that provokes the elicitation of the respective function on a selected key word. One of the advantages of this method is that is does not require actors or otherwise trained speakers. Instead, naive speakers can be used who only need to be controlled through minor instructions. Furthermore, starting from a written text allows controlling the segmental and prosodic make-up of the key words and their linguistic contexts.

This method has already been used successfully in a study of the realization of different pitch accent categories in German [11]. In the present investigation, the method was further elaborated to cope with the expressive nature of the speech functions. Two sets of 14 and 15 short monologues in two illustrated frames for 'positive' and for 'negative intensification', respectively, and 8 mini-dialogues contextualising 'weighting' and 'intensification', were constructed, e.g.

hm, ***lecker****!* *das* ***schmeckt!*** "yummy! it's delicious"
das ***stinkt****! zum* ***Kotzen****!* "it stinks! disgusting!"

A. *ich hab mir gestern einen neuen Laptop gekauft. zwei Jahre gebraucht für 1000 Euro. gut, oder?*
B. was hast du bezahlt? 1000 Euro? das ist zu viel! das ist ***viel*** *zu viel!*
A. "I bought myself a new laptop yesterday. two years second-hand for 1000 Euro. good, isn't it?"
B. "what did you pay? 1000 Euro? that's too much! that's far too much!"

The dialogues were read by pairs of speakers (one female, one male) sitting face to face at a table, to create a realistic communicative situation. The speaker pairs were selected with regard to two criteria: (1) They were known to the authors as having an extrovert, expressive character. (2) They knew each other very well. Both criteria were to contribute to a relaxed atmosphere during the recording session and to raise the probability of occurrence for '**P** and **N** emphasis'.

3.2. Data recording

Prior to the recording session, the speakers were instructed that they were to read dialogues from separate sheets provided for each, one after the other and as often as necessary, until both agreed that the dialogue sounded natural. They were allowed to modify the given texts slightly according to their personal tastes, e.g., by introducing or substituting words or changing the wording of a passage. None of the speakers recorded so far changed any of the typographically highlighted key words.

Each pair of speakers performed a second round of the dialogues with reversed roles. After the dialogue session, which familiarized the speakers with the recording situation and the expressive speaking style, each speaker read the 2 sets of monologues from 2 separate sheets. Again the speakers judged each other's productions with regard to their naturalness, and continued rendering the monologues until a satisfying version was reached.

The speech signals were recorded by direction microphones placed on the table in front of each subject. The recording was stereo, with a separate channel for each microphone. A complete recording session took 1-2 hours. Four pairs of speakers from North Germany (3 pairs in their 30s, 1 in their 60s) have been recorded with this experimental set-up. Audio examples as well as the presented texts are available in [12].

3.3. Data labelling

The complete corpus was prosodically labelled by Niebuhr, using the program package *xassp* [4] and the KIM-based tool PROLAB [13]. Auditory analysis of the corpus showed that the type of emphasis produced within each of the designed context frames for **P** and **N** was not homogeneous with regard to both function and phonetic manifestation. The frequency of '**P** emphasis' produced in the contexts designed for '**N** emphasis' or vice versa was rare (< 5%), but each of the two context frames elicited *more* types of emphasis than just **P** and **N**. Therefore, labelling had to be done by reference to the perceived types of emphasis, rather than the given context frames; and the PROLAB notation system was expanded accordingly, marking '**P** and **N** emphasis' besides another two types, viz. (1b) 'weighted', a traditional PROLAB category, and the new category (1c) 'insisting'. It needs to be stressed that this labelling was guided by function, not by phonetic properties.

3.4. Data analysis

In the subjects' labelled monologues and dialogues, Niebuhr carried out analyses along the parameters of the hypotheses in section 2. For hypotheses 2-4 and 1.4, physical measurements were obtained in *xassp, praat,* and *cool edit*, and *t* tested in SPSS; hypotheses 1.1-1.3 were tested on database searches for the prosodic labels by *Chi*^{2}.

4. RESULTS

There are 159 cases labelled as '**P** emphasis', 128 as '**N** emphasis'.

4.1. Intonation

All these tokens of emphasis have peak contours which may be differently synchronized with the accented vowel onset: *early* **E**, *medial* **M**, *late medial* **LM**, *late* **L** [7] (n=287). Concatenation with the preceding and following contour may be *flat* **0.** – *slightly dipped* **1.** – *low-dipped* **2.** Table 1 gives the frequency distributions. Preaccentual concatenation requires a preceding accent, so the number of cases reduces to n$_{pre}$=125.

Table 1: Frequencies of the 4 peak synchronizations and the 3 pre/postaccentual concatenations in **P** and **N**; postaccentual in italics

	E	M	LM	L	0.	1.	2.
P	1	114	44	0	20	37	6
					27	*66*	*66*
N	31	77	13	7	31	28	3
					16	*51*	*61*

Three *Chi*^{2} tests on the synchronization and the two concatenation data sets for homogeneity of distribution across the two emphasis categories were carried out. Synchronization is highly significantly different for **P** and **N** in accordance with hypothesis 1.1 (chi2$_{3,0.001}$=16.27 < 56.46). Preaccentual concatenation shows a significant trend in line with hypothesis 1.2 (chi2$_{2,0.05}$=5.99 > 4.61). Postaccentual concatenation shows no difference between the two emphasis categories. So hypothesis 1.3 is rejected.

Subsets of the total sample were formed by selecting pairs of phrases from the **P** and **N** elicitation lists, having the same number of accents and being either identical segmentally, or showing similar syllable numbers and structures (8 pairs). The occurrences of labelled **P** and **N** in the key words of these pairs were summed across the 8 speakers, yielding **P**=35 and **N**=25. The semitone ranges and standard deviations according to hypothesis 1.4 were then obtained for the complete phrases in each of these data sets; *t* tests for independent samples showed no differences, so hypothesis 1.4 is rejected.

4.2. Acoustic energy

Subsets were formed, including the identical **P/N** pairs, and phrases containing key words, labelled **P/N**, with initial fricatives /f, ʃ/ in accented syllables. This resulted in **P**=59 and **N**=53. Hypothesis 2.1 was confirmed by *t* test for independent samples, which was highly significant; cf table 2.

Table 2: Statistics of *t* test for independent samples of time (in ms) between accented vowel onset and energy maximum in the **P** and **N** subsets. T and *df* were corrected for heterogeneous variances (revealed by *F* test)

	n	mean	sd	T	*df*	p
P	59	127	76	-7.46	75.22	<.0001
N	53	47	28			

Hypothesis 2.2 was tested with a *t* test for independent samples based on key words with initial fricatives /ʃ/ in accented syllables. There is no significant difference between **P** and **N**, so hypothesis 2.2 is rejected.

4.3. Duration

The labelled **P** and **N** key words were ordered in classes of syllable structures, viz. consonant/cluster onset, long/short vowel nucleus, and voiced/voiceless/no coda. Only the structure 'single consonant (C) + long vowel/diphthong (V) + any coda' had a sufficient number of instances for statistical analysis. The C and V durations in this structure were measured and C/V calculated for the **P** and **N** sets. The results of *t* tests for independent samples show highly significant differences between the two sets for C/V, as well as C, V separately, which points to a bidirectional duration change. See table 3. Hypothesis 3 can be accepted.

Table 3: Statistics of *t* tests for independent samples of C and V durations (in ms) and C/V ratios in the accented syllables of the **P** and **N** subsets. For C/V and C, T and *df* were corrected for heterogeneous variances (revealed by *F* tests)

		n	mean	sd	T	*df*	p
C/V	**P**	47	0.48	0.17	5.53	13.76	<.0001
	N	14	1.28	0.53			
C	**P**	47	134	42	3.22	16.05	0.005
	N	14	196	67			
V	**P**	47	294	79	-5.99	59	<.0001
	N	14	161	43			

4.4. Voice Quality

The labelled **P** and **N** key words containing /a:/ or /a/ were selected, and F1, F2 and H1/H2 were taken from LPC and DFT spectra, respectively, centrally in the vowel. The formants do not differ between the two sets, but the H1/H2 ratio is significantly different, pointing to breathier voice in **P** than **N**; cf. table 4. Hypothesis 4 can be accepted.

Table 4: Statistics of *t* tests for independent samples of F1 and F2 (in Hz) and H1/H2 ratios in the **P** and **N** subsets. For F1, T and *df* were corrected for heterogeneous variances (revealed by *F* tests)

		n	mean	sd	T	*df*	p
F1	**P**	13	640	129	0.22	16,75	0.829
	N	22	649	74			
F2	**P**	13	1312	182	-0.07	33	0.945
	N	22	1308	146			
H1/H2	**P**	13	1.04	0.06	-2.83	33	0.008
	N	22	0.97	0.08			

5. DISCUSSION

The data acquisition procedure generated natural, albeit acted, expressive speech and may be adopted as an efficient way of eliciting different types of emphasis systematically. The linguistic and situational contextualization in mini-dialogues and in two sets of monologues did, however, not lead to unique renderings of the pre-defined emphasis functions. Listening to the recorded data showed up two types of divergences. On the one hand, other functions were implemented instead of positive and negative intensification. And in a few cases, negative intensification was used in the positive contextualization and vice versa, which the two authors observed as such, and either interpreted as inadequate renderings in the contexts or as irony, with verbal and prosodic meanings going against each other and prosodic meaning winning.

It must be admitted that in these cases there is the danger of argumentative circularity between functional assessment and phonetic manifestation, each determining the other since both classifications were carried out by the same metalinguistic observers. But the investigation has allowed to pinpoint a set of differentiating features for '**P** and **N** emphasis', which will lead to further experiments towards a complete framework linking function to phonetic exponents of 'emphasis' in stepwise, spiral-like progression. This link needs to be validated by formal perception experiments in which ordinary listeners allocate systematically manipulated stimuli to the set of 'emphasis' category labels.

The pitch, energy, and duration patterns converge in intensifying

- the nucleus of the accented syllable by lengthening, and by rising, high pitch for '**P** emphasis'
- the beginning of the accented syllable by lengthening the consonantal onset at the expense of the nucleus, followed by falling, low pitch in the nucleus for '**N** emphasis'.

So, **P** strengthens, **N** weakens sonorous features. Furthermore, **P** tends to have soft breathy voice as against tight voice phonation in **N**.

The acoustic analysis has focussed on '**P** and **N** emphasis' in the *key words* of the data collection paradigm. Since **P** and **N** were also produced elsewhere in the monologues and dialogues, further analysis of the recorded and labelled corpus has to include them. It also needs to tease out the prosodic manifestations of the other emphasis functions contained in the corpus, and compare them with '**P** and **N** emphasis', including the analysis of pauses before emphasized syllables, which can mark emphasis in general. The paradigm should also be applied to other languages. The hypothesis is that the general characterization of '**P** and **N** emphasis' will be found widely across languages and is perhaps a language universal. English certainly shows the same feature distinctions [8].

6. REFERENCES

[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] IPDS (1997). *xassp* User's Manual. *AIPUK 32,*31–115.

[5] Kohler, K. J. 2003. Neglected categories in the modelling of prosody: pitch timing and non-pitch accents. In *Proc. 15th ICPhS*, Barcelona. 2925-2928.

[6] Kohler, K. J. 2005. Form and function of non-pitch accents. *AIPUK* 35a, 97-123.

[7] Kohler, K. J. 2006. Paradigms of experimental prosodic analysis – from measurement to function. In *Methods in Empirical Prosody Research. Language, Context, and Cognition*. Berlin: de Gruyter.

[8] Kohler, K. J. 2006. What is emphasis and how is it coded? In *Proc. 3rd International Conference on Speech Prosody*. Dresden. 748-751.

[9] Ladd, D. R.; Morton, R. 1997. The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics* 25, 313-342.

[10] 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.

[11] Niebuhr, O. 2006. *Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen.* PhD thesis. Kiel University.

[12] Niebuhr, O. 2007. *Audio examples of emphasis categories in German.* www.ipds.uni-kiel.de/on/Emph07.html.

[13] Peters, B., Kohler, K. J. 2004. *Trainingsmaterialien zur prosodischen Etikettierung mit dem Kieler Intonationsmodell KIM.* www.ipds.uni-kiel.de/kjk/forschung/lautmuster.en.html.