PRODUCING PHRASAL PROMINENCE IN GERMAN

Bistra Andreeva, William Barry and Ingmar Steiner

Institut für Phonetik, Universität des Saarlandes
andreeva/wbarry/steiner@coli.uni-saarland.de

ABSTRACT

This study examines the relative change in a number of acoustic parameters usually associated with the production of prominences. The production of six German sentences under different question answer conditions provide de-accented and accented versions of the same words in broad and narrow focus. Normalised energy, F_0, duration and spectral measures were found to form a stable hierarchy in their exponency of the three degrees of accentuation.

Keywords: German, phrasal accentuation, parameter interaction

1. INTRODUCTION

Prominence is a perceptual phenomenon which, in spoken language, results auditorily from the realisation of phonological structuring at two levels, lexical stress and phrasal accentuation. The acoustic dimensions which accompany greater or lesser perceptual prominence are duration, intensity, fundamental frequency (F_0), and spectral properties. The signal properties associated with these dimensions are always present in any utterance, and they are neither totally independent of one another in the production process nor in the psycho-acoustics of perception. Intensity and F_0 co-vary [8], often correlating with duration and spectral definition in production as a result of general articulatory effort. Below the duration of approximately a short syllable, duration integrates with intensity to determine perceived loudness [9]. However, they *can* be independently adjusted to a certain extent, and there is an assumption that languages differ in their exploitation of the four dimensions. E.g. intensity based dynamic accent in Slavonic languages; F_0 and duration in Germanic languages [10].

The classic experiments seeking to unravel the contribution of the different acoustic properties to the perception of prominence were those by Fry [3, 4]. However, the domain of disyllabic words chosen for the stimuli presented as single-word "utterances" confounds word-stress realisation with phrase-level accentuation. Also, the restriction to one language (British English in that case) leaves the cross-language perspective open. But systematic multi-language comparisons are almost non-existent (but see [2] for a cross-language consideration of Swedish and English, [1] for an auditory analysis of German and English and [10] for Russian, Czech, Polish and Ukrainian).

As well as working with one language, empirical studies generally focus on the importance of one dimension rather than another, with the perhaps unintentional implication that "the winner takes all" (but see [1] and [6] for more comprehensive comparisons of parameter contributions). Large-corpus analyses, which could be expected to provide solid, statistically reliable statements, either have speech-technologically orientated aims such as accent prediction and often include non-acoustic (top-down) information [5] or have muddied the waters with apparently contradictory statements rather than clearing them [cf. 7 compared to 6]. Differences in the language (varieties) used, the methods for defining the units to be analysed and the analytic procedures employed to arrive at the conclusions make comparison almost impossible.

The study presented here also focuses on only one language, German, but its prime goal is to compare the *degree* to which the four acoustic dimensions identified above change between phrasally unaccented and phrasally accented words. This separates, like [11], phrasal accentuation from lexical stress. It also addresses a question, hitherto broadly ignored, crucial to any attempt to clarify the question of cross-language differences in the production of prominence, namely whether *individuals* vary significantly in the relative extent to which they exploit the four acoustic dimensions. There are plausible phonological differences between languages (e.g. the use or not of vowel- and consonant-length oppositions, fixed vs. variable lexical-stress position, the use or not of lexical-tonal accents) which support a hypothesis of differing hierarchies in the exploitation of the acoustic dimensions, but none for such differences between speakers of the same language.

2. METHODOLOGY

2.1. Material

In order to provide a basis for the direct comparison of parameter values across different conditions of phrasal accentuation, controlled utterances were required which could be produced with de-accented and accented variants of the *same words*. Short sentences were constructed containing *two* one- or two-syllable "critical words" (CWs), one early (but not initial) and one late (but not final) in the sentence. The sentences (with the critical words underlined) are:

1. Der Mann fuhr den Wagen vor.
2. Das Bild soll nicht hässlich sein.
3. Das Kind sollte im Bett sein.
4. Der Peter kann den Film gucken.
5. Das Mädchen soll ein Bild malen.
6. Mein Vater kann Türkisch lesen.

For each sentence, a number of questions were devised to elicit a) a *broad-focus* response, b) a response with a *non-contrastive narrow-focus* on the early and c) on the late CW and d) a *contrastive focus* on the early and e) on the late CW (the contrastive focus has not been included in this study).

To provide a (potential) basis for comparing the parameter modification across sentences independently of the different segmental structuring of the critical words (and thus, if possible, to derive a speaker- and/or language-specific quantification of the accent-dependent modification), a reiterative "*dada*" version of each realisation was produced immediately after the normal-text response. This was produced in two stages: (i) a *da* or *dada* replacement of the two CWs and (ii) a *dada* replacement of the whole sentence. The latter is considered here.

2.2. Speakers, recordings and analysis

Six regionally homogeneous speakers of High German (3f, 3m) produced 6 repetitions each of the sentences and their 2 *dada* versions from a PPT presentation in response to the recorded questions in a sound-treated studio. The regional homogeneity aimed at increasing the chance of a group hierarchy in the exploitation of the acoustic dimensions (i.e. the regional sub-stratum which could have influenced the establishment of their prominence-giving mechanisms was constant). All were tertiary-educated.

The recordings were made using an AKG C420^{III}PP headset on a Tascam DA-P1 DAT recorder and transferred digitally via the optical channel to a PC using the Kay Elemetrics MultiSpeech speech-signal processing program. Segmentation, labelling with SAMPA and further processing was done using the Kiel XASSP speech signal analysis package.

Six labelling assistants were allocated different sentences (to maximize labelling consistency across conditions within each sentence) and segmentation problems were regularly discussed and decided with the authors at group level. The four acoustic dimensions were parameterised as follows:

1. Durations were calculated for all *feet* in the sentences, for the *CW*s and their component *syllables* as well as the *syllables of the feet* to which the CWs belonged. Furthermore, the duration of the phonetic sound-segments comprising the syllables were calculated. All durational measurements were normalised as a percentage of the mean duration of the corresponding unit in the sentence.

2. Since comparisons focus on changes in identical words across conditions, F_0 was calculated as the mean fundamental frequency (Hz) across the syllable nucleus (vowel or syllabic sonorant) of the lexically stressed syllable of CWs and in the unstressed syllable preceding and following it. These values were also expressed as percentages of the mean overall F_0 of the sentence.

3. The signal strength was captured in a number of ways: (i) the normalised mean intensity (dB) of the syllabic nuclei in the lexically stressed syllable of the CW, expressed as a percentage of the mean overall intensity of the sentence. (ii) The spectral balance (calculated as the energy difference from 70 Hz to 1 kHz and from 1.2 to 5 kHz]) of the syllabic nuclei in the lexically stressed syllable of the CW.

4. Spectral definition was captured with the mean frequency (and bandwidth) values for formants 1-3 of the syllabic nuclei in the lexically stressed syllable of CWs. Change as a function of accentual condition was expressed as percent difference from the broad-focus realisation in each formant separately for the vowel of the lexically stressed syllable in each CW.

3. RESULTS

The following sections contain the percent difference values for each of the acoustic parameters separately for the group of 6 speakers. The degree to which the values vary for the individual speakers can be seen in figures 1-4 in the accompanying file (Graphs.pdf).

3.1. Duration

Greater duration is generally considered to be an important feature of stressed/accented elements in German. In unstressed syllables, the long-short vowel distinction is neutralised, leaving a tense-lax quality distinction in careful speech. This phonological regularity does not, however, necessarily mean that there is a primarily durational basis to *phrasal* prominence.

In a multivariate ANOVA (speaker × sentence × degree-of-accent (DA)), there was a main effect of DA ($p < 0.001$) on duration for the critical foot (CF), CW, critical syllable (CS) and critical vowel (CV). Table 1 shows the percent difference in (normalised) duration for early and late critical units (CUs) under the conditions of de-accentuation and narrow-focus accentuation *relative to broad-focus accentuation.*

Table 1 Percentage difference in normalised durations relative to the broad focus condition (natural)

	CU1		CU2	
Unit	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
Foot	− 3.61 %	+ 8.57 %	− 6.97 %	+ 5.45 %
Word	− 5.95 %	+ 16.13 %	− 7.33 %	+ 9.41 %
Syllable	− 5.85 %	+ 17.54 %	− 6.51 %	+ 11.86 %
Vowel	− 6.98 %	+ 12.56 %	− 7.41 %	+ 3.17 %

Behind these global differences lies a great deal of structural variation – words in the foot, syllables in the words and segments in the syllables. A comparison with the *dada* renditions gives an idea of how the units change when all differences in syllable structure have been equalised (see Table 2).

Table 2 Percentage difference in normalised durations relative to the broad focus condition (*dada*)

	CU1		CU2	
Unit	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
Foot	− 3.36 %	+ 5.01 %	− 4.05 %	+ 4.65 %
Word	− 4.63 %	+ 10.73 %	− 4.37 %	+ 8.12 %
Syllable	− 3.72 %	+ 13.28 %	− 4.75 %	+ 10.52 %
Vowel	− 9.17 %	+ 8.79 %	− 5.44 %	+ 7.45 %

It is apparent that the strongest durational effect of accentuation is found for the syllable unit. Though the durational difference for the word unit is only slightly lower, this is a reflection of the fact that 5 of the 12 CW are monosyllabic. The difference values for disyllabic words are lower.

Note that the broad-focus nuclear accent is different from the narrow-focus accent (though closer to it than the broad-focus *pre*-nuclear accent).

3.2. Fundamental frequency

Higher and/or changing F_0 as part of the intonation structure of spoken utterances is a generally accepted carrier of phrasal prominence. In the ANOVA, a significant main effect was found of DA ($p < 0.001$) on (normalised) F_0 and in the amount of pitch change over three syllables – from the syllable preceding to the syllable following the stressed syllable, henceforth "contour". Tables 3 and 4 give the percent difference for the early and late CUs under the conditions of de-accentuation and accentuation in narrow-focus *relative to accentuation in broad-focus*.

In terms of both the F_0 of the CS and the contour, the normalised F_0 differences show, with one exception, a *much greater increase* from broad to narrow focus than the decrease from broad focus accent to de-accented. The exception is in the pitch of the late,

Table 3 Percentage difference in normalised pitch relative to the broad focus condition (natural)

	CS1		CS2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
F_0	− 5.01 %	+ 25.13 %	− 14.67 %	+ 22.93 %
Contour	− 2.95 %	+ 27.93 %	− 3.48 %	+ 26.90 %

Table 4 Percentage difference in normalised pitch relative to the broad focus condition (*dada*)

	CS1		CS2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
F_0	− 3.84 %	+ 19.25 %	− 12.36 %	+ 17.20 %
Contour	− 1.57 %	+ 20.36 %	− 3.59 %	+ 20.30 %

broad focus nucleus, which is 14.67 % higher (12.36 % in *dada*) than the de-accented syllable.

3.3. Intensity

With few exceptions [6], signal intensity has been discounted as an important correlate of accentuation, though spectral balance has been found to contribute significantly [6, 8, 11]. Both measures showed a numerically small but highly significant main effect of DA ($p < 0.001$). The details for both natural and *dada* speech are given in Tables 5 and 6.

Again, the relative proximity of the late broad focus nucleus to the narrow focus accent compared to the late de-accented CS is apparent in both natural and *dada* intensity values, though only in the *dada* spectral balance value. A conspicuous difference between these two measures is the degree of divergence between them.

Table 5 Percentage difference in normalised energy relative to the broad focus condition (natural)

	CV1		CV2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
Intensity	− 0.48 %	+ 3.42 %	− 4.69 %	+ 2.89 %
Spectral balance	− 1.53 %	+ 1.59 %	− 0.06 %	+ 2.54 %

Table 6 Percentage difference in normalised energy relative to the broad focus condition (*dada*)

	CV1		CV2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
Intensity	− 0.95 %	+ 3.08 %	− 5.03 %	+ 2.31 %
Spectral balance	− 1.27 %	+ 0.63 %	− 2.63 %	+ 1.78 %

3.4. Vowel quality

Unlike English and some Slavonic languages, German does not undergo phonological vowel-quality reduction in unstressed syllables. Effort-linked phonetic reduction is, however, well documented, and is therefore of interest as an exponent of phrasal accentuation. Tables 7 and 8 give the percentage change in F_1-F_3 as a function of DA.

Table 7 Percentage difference in formant values relative to the broad focus condition (natural)

	CV1		CV2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
F_1	− 2.80 %	+ 5.20 %	− 2.66 %	+ 8.07 %
F_2	− 0.36 %	+ 0.28 %	− 2.42 %	+ 1.74 %
F_3	+ 0.22 %	+ 0.40 %	− 1.20 %	− 0.09 %

Table 8 Percentage difference in formant values relative to the broad focus condition (*dada*)

	CV1		CV2	
	De-accented	Narrow focus acc.	De-accented	Narrow focus acc.
F_1	− 3.65 %	+ 2.49 %	− 4.61 %	+ 5.10 %
F_2	− 0.28 %	+ 1.22 %	− 0.45 %	+ 0.98 %
F_3	+ 0.29 %	+ 0.65 %	+ 1.29 %	+ 1.42 %

4. DISCUSSION

If we consider the full range of change in parameter values (from de-accentuated to narrow focus accentuation) as the potential for the exploitation of that parameter, we find almost exactly the same hierarchy of exploitation, whether in natural sentences or *dada*-utterances, whether in early or late position:

$F_0 \geq$ Contour > Syl.dur. > F_1 > Energy > Spectr.bal.

This hierarchy confirms the frequent claims that pitch (change) and duration are more important than intensity as correlates of stress and accent. However, the position of F_1 change in the hierarchy has been identified for the first time, though Fry [4] also found spectral change to be a weaker cue (in English, a vowel reducing language) than duration in word-stress judgments. That relative F_1 change is stronger than intensity change, and that intensity change is, in turn, stronger than the changes in spectral balance was not expected.

It must be stressed that these findings reflect a "production hierarchy" based on percent difference in normalised parameters. How this metric relates to *perceptual* weight still has to be clarified. It can well be that a small percentage change in spectral balance or intensity is equivalent to a much bigger percentage F_1 change.

Another caveat is the main speaker effect found for all parameters as well as the interactions of speaker with sentence and degree-of-accent. Scrutiny of the figures in the accompanying file Graphs.pdf reveals that there are speakers who have generally higher or lower difference values, but there are also speakers who clearly use one parameter to a greater degree than another. In other words, there is some evidence even from this cursory observation that speakers within one and the same language (even regional variety) have their own individual prominence-producing strategies. How far such differences can be grouped into a small number of strategy types is unclear at this point. While such speaker differences were no impediment to the main degree-of-accent effect, they require careful further analysis.

However, the consistency of the group degree-of-accent results is encouraging as a first step towards cross-language comparisons. Whether or not the hierarchy changes from one language to another, shifts in the relative proximity of parameters within the same hierarchy are still a sign of their relative importance.

As a final note, the much discussed issue as to whether a narrow-focus nuclear accent is realized differently from a broad-focus nucleus was decisively confirmed by our data for all parameters.

Acknowledgments: This research was funded by a grant BA 737/10-1 from the German Research Council within the Research Programme SPP 1234

5. REFERENCES

[1] Benkwitz, A. 2003. *Kontrastive Untersuchungen zum Rhythmus. Britisches Englisch als Ausgangssprache – Deutsch als Zielsprache.* (Hallesche Schriften zur Sprechwissenschaft und Phonetik 14). Frankfurt: Lang.

[2] Eriksson, A., Grabe, E., Traunmüller, H. 2002. Perception of syllable prominence by listeners with and without competence in the tested language. *Proc. Speech Prosody 2002* Aix-en-Provence: Laboratoire Parole et Langage, 275-278.

[3] Fry, D.B. 1958. Experiments in the perception of stress. *Lang. & Speech* 1, 126-152.

[4] Fry, D.B. 1965. The dependence of stress judgments on vowel formant structure. *Proc. 6th ICPhS* Münster, 306-311.

[5] Hirschberg, J. 1993. Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence* 63, 305-340.

[6] Kochanski, G., Grabe, E., Coleman, J. 2005. Loudness predicts prominence: fundamental frequency lends little. *J. Acoust. Soc. Amer.* 118, 1038-1054.

[7] van Kuijk, D., Boves, L. 1999. Acoustic characteristics of lexical stress in continuous speech. *Speech Comm.* 27, 95-11.

[8] Ladefoged, P. 1967. Stress and respiratory activity. *Three Areas of Experimental Phonetics*, Oxford: OUP, 1-49.

[9] Moore, B. 2003. *An Introduction to the Psychology of Hearing* (5th edition). London etc.: Academic Press.

[10] Nikolaeva, T. 1969. *Intonacija složnogo predloženija v slavjanskich jazykach*. Moskva: Nauka.

[11] Sluijter, A., van Heuven, V. 1996. Spectral balance as an acoustic correlate of linguistic stress. *J. Acoust. Soc. Amer.* 100, 2471-2485.