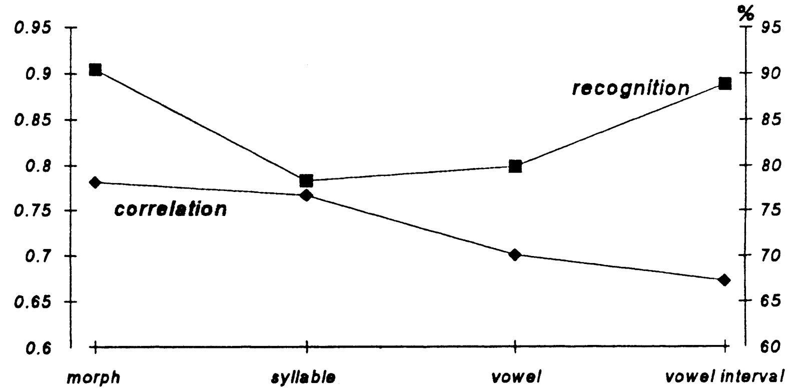

Figure 2: Correlation and recognition results of four different subunits of words using duration and the balance comparison.

Figure 3 displays results for duration, mean energy and fundamental frequency of vowels. It can be seen that duration predicts the location of stress best.

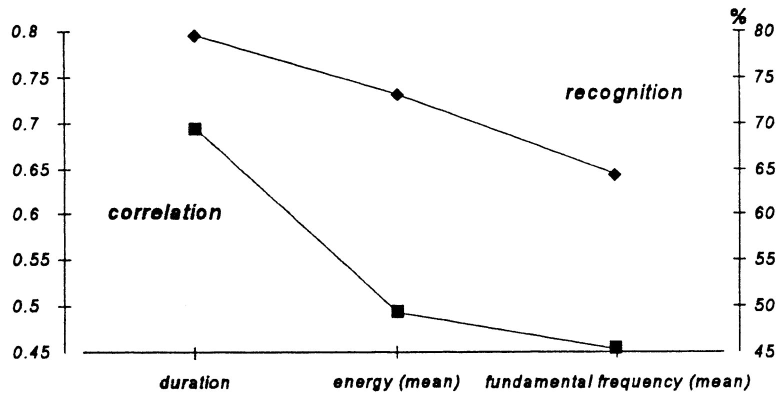

Figure 3: Correlation and recognition results of three different properties using vowels and the balance comparison.

Figure 4 shows results for f_0 using the parameters vowel and balance.

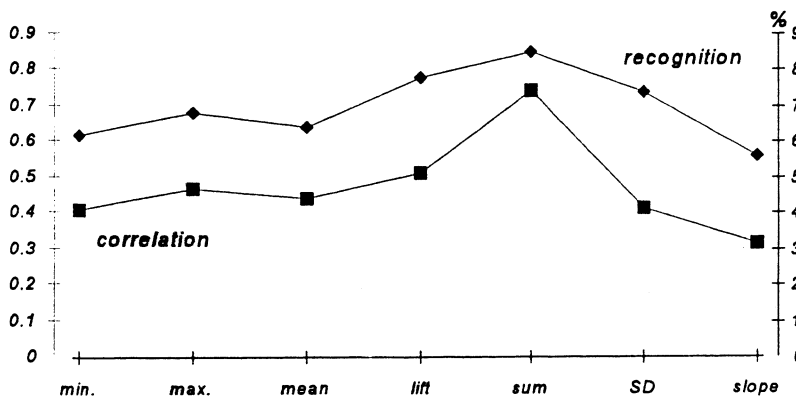

Figure 4: Correlation and recognition results comparing seven measures of fundamental frequency of words using vowels and the balance comparison.

Using fundamental frequency as a predictor of stress in German words yields recognition rates up to 85.539 % (sum of f_0). Also, the lift and SD measure are useful for stress detection.

The results obtained using different measures of energy are shown in Figure 5.

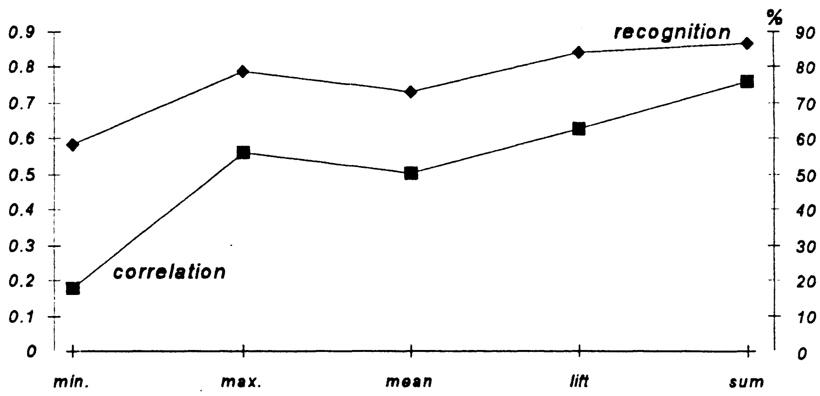

Figure 5: Correlation and recognition results comparing five measures of energy using vowels and the balance comparison.

Recognition rates of the sum of energy are even better than those for duration and fundamental frequency, also the maximum of energy can detect stress positions quite well.

Figure 6 shows the results for different methods of comparison for the property duration and the unit vowel. For this evaluation, only correlations could be computed.

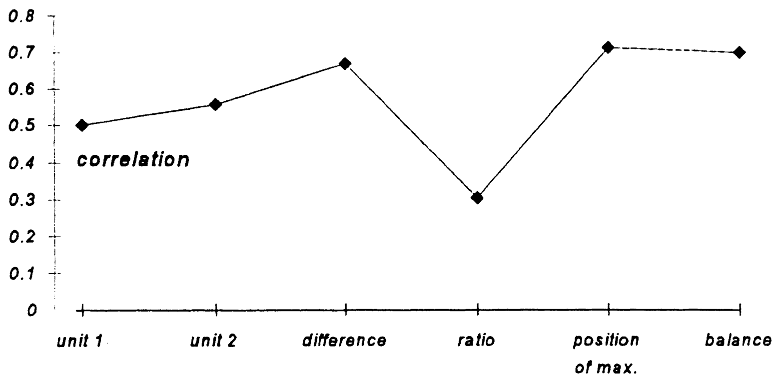

Figure 6: Correlation results for different methods of comparison units using duration and vowels.

The second unit - here the vowel - correlates more with the stress position than the first one. The best results for the measurement of stress positions can be obtained by using the positional coding, the balance value and the difference.

In order to check whether these results are valid across different linguistic and acoustical environments, some additional control measurements have been conducted: Correlations and recognition rates of five parameter combinations (positional coding of duration, mean f_0, lift of f_0, mean energy, and lift of energy) were calculated for median split subgroups of the data. The following subgroup criteria were selected: sentence position, sentence duration, focus probability, word duration, relative word duration, speaking rate, distance to last lexical stress, sex of speaker, part of speech. For none of these subgroups there was a significant change in the predictive power of duration.

6 Conclusion on Parameter Evaluation

This investigation has shown that duration-related measures (duration, sum of energy, sum of f_0) are the most reliable measures for stress location. The most useful methods of comparison are the positional and the balance measure. The most reliable units are morphs and vowel onset intervals. Altogether, the best parameter combination for the prediction of stress positions of German bisyllabic words is the balance of the sum of energy of morphs (correlation: 0.809, recognition: 93.233 %).

The fact that RMS predicts stress positions even better than duration alone, might also hold on perceptual level: Duration perception can be manipulated by energy (Turk & Sawusch 1996), louder events are perceived as being longer. Further research has to show to what extent the better results of the sum of energy corresponds to duration perception by the listener when detecting stress positions.

No perception oriented explanation can be given to account for the finding that morphs are the best predictors of stress locations: Morphs are semantic units, there are little acoustical cues that mark morphological boundaries. Vowels and vowel onset intervals are more plausible perceptual units, both of them being well detectable because of a rise of energy at the onset (Pompino-Marschall 1989, Janker 1997).

The model of lexical stress suggested here is as follows: A rise of energy in the (speech-) signal is perceived as marking a new event. In speech signals this rise of energy coincides with the onset of vowels. Durations of these events are compared within groups of events, the longest of which is perceived as most prominent and thus stressed (cf. Figure 7).

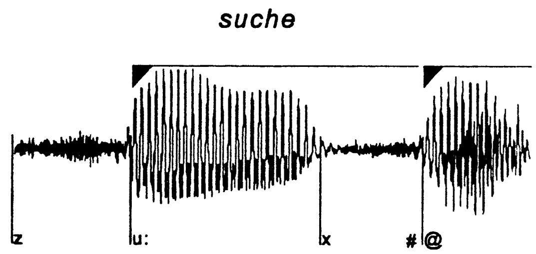

Figure 7: Vowels and vowel onset intervals.

On the signal level, morphological units show the best results for the correlation and the detection of stress location. In German, lexical stress and semantic structure correspond. On the perceptional level, vowels or vowels onset intervals (rise of energy at onset) are the more likely stress units.

What is the link between morphological and perceptually relevant units in German words? How can acoustical and semantical properties of German words be integrated in one model of lexical stress? Why is the morphological structure of German words a good predictor of lexical stress although morphological boundaries are not relevant for stress perception?

7 Morphology

Morph categories have different segmental structures: For each morph category, mean durations of morph onset, nucleus, and coda were calculated (Figure 8).

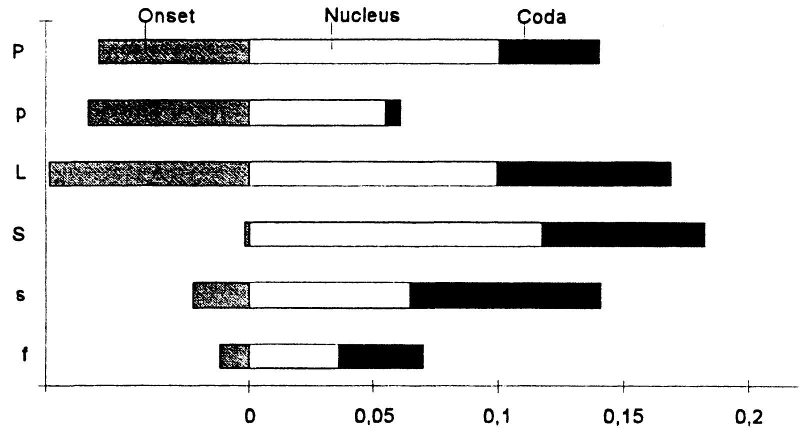

Figure 8: Onset, nucleus, and coda durations of morph types

The information of Figure 8 combined with the knowledge on morphological structures of German words can explain the correspondence of informational relevance and perceptual prominence of morphological units in German: Computing vowel and vowel onset interval durations of common word structures (L+f - *Kinder* [kInd6], L+L - *Haustür* [haUsty:6], L+s - *glücklich* [glYklIC], p+L - *bereit* [b@RaIt], P+L - *Vorzug* [fo:6tsu:k], L+S - *polar* [pola:6], the stress locations can be predicted correctly.

8 Conclusion

From the above results it can be concluded that the perception of German lexical stress is associated to durational differences of units of words. This change of duration of different units of words is due to segmental complexity: Stressed units have longer sounds and more sounds. Having more sounds has functional reasons: Morphs bearing most information (lexical morphs) are open class morphs and thus need more complex signal and segmental representations, too. Thus, German lexical stress is a function of semantic weight. During an utterance, no additional effort by the speaker is needed. Inherent signal properties of segmental compositions of morphs and the morphological structure of words evoke the stress location perceptions coded in transcription dictionaries.

References

Heuft, B.; Portele, T. (1994): Zur akustischen Realisierung des Wortakzents. Elektronische Sprachsignalverarbeitung 95: 197-204.

Jessen, M.; Marasek, K.; Schneider, K.; Claßen, K. (1995): Acoustic Correlates of Word Stress. ICPhS 95, 4: 428-431.

Kohler, K.J. (ed.)(1994): Phonetisch-Akustische Datenbasis des Hochdeutschen. Kieler Arbeiten zu den PHONDAT-Projekten 1989-1992. AIPUK 26.

Janker, P.M. (1997): Evidence for the P-Center Syllable-Nucleus-Onset Correspondence Hypothesis. ZAS Papers in Linguistics 7: 94-124.

Pompino-Marschall, B. (1989): On the Psychoacoustic Nature of the P-Center Phenomenon. Journal of Phonetics 17: 175-192.

Turk, A.E.; Sawusch, J.R. (1996): The Processing of Duration and Intensity Cues to Prominence. JASA 99 (6): 3782-3790.

Domains and Properties of Lexical Stress in German

Andreas Mengel
Institut für Kommunikationswissenschaft, Technische Universität Berlin

Abstract
This investigation addresses acoustical aspects of lexical stress in German. Different properties, units of words and methods of comparison have been evaluated in order to identify speech signal properties relevant for stress perception. It is argued that duration is the key property of stress. An explanation of how duration integrates information structure and perceptual prominence is given.

1 Motivation

In trying to identify acoustical correlates of word stress, fundamental frequency (f_0), energy and duration are usually considered as stress properties. Duration and f_0 (Heuft & Portele 1994) or only duration (Jessen et al. 1995) have been found to be dominant cues in German. These results were obtained with a set of specially designed words or sentences.

There are a couple of motivations for this study. First of all its aim is to replicate the above findings by using a larger corpus. The second reason is to evaluate possible parameters that are needed to model the perception of lexical accent: If it is duration and fundamental frequency that predict stress location best, then how do speakers of German perceive stress? A third reason for this study is to understand the functional aspects behind the signal properties of German lexical stress.

2 Specific Questions

For the purpose of this study, a set of four questions were addressed, for not only the problem of what acoustical parameters are relevant for lexical stress in German may be important but also how they are measured.

a) If stress is perceived as being located at some special position in a word, then it must be possible to identify units that allow the perception of stress. It is necessary to find out what the **units** relevant for the perception of lexical stress are. Furthermore, it is important to ask which phonological, semantic, articulatory or auditory units can serve best as these units. Candidates for possible units are morphs, syllables, vowels, and vowel-onset intervals.

b) If stress can be detected in speech signals, then there must be signal properties that mark stress and that differ systematically between stressed and unstressed units. It is possible that the signal properties used for detection of stress vary but it is hypothesized that there is one predominant signal aspect that is perceived as indicating stress. Possible candidates of **signal properties** of stress are duration, f_0, and energy.

c) If special signal properties indicate stress, then these properties must be evaluated. Different ways of evaluation are possible. **Signal property measures** considered are the mean, the minimum, the maximum, the lift, the sum, the slope, and the standard deviation of the signal properties.

d) If stress can be explained as a perception process that involves the evaluation of signal properties, it is also necessary to map or to compare these signal properties to some reference measure. First of all, there could be an internal reference, a standard which measures of units are compared to. It might also be, that units of words are compared to each other. In that case, different kinds of comparison are possible. Thus, candidates for the kind of **comparison** that is involved when detecting stress are the absolute values, ratios, differences, the position of the maximum, and a balance measure of the property measures of units of words.

3 Data

A total of 11,644 bisyllabic word utterances (10,806 stressed on the first, 838 stressed on the second syllable) representing 782 different word types were selected (16 kHz, 16 bit mono, studio recordings, read speech, 53 speakers: 27 female, 28 male). They were extracted from the data base described in Kohler (1994). No further treatment of data has been performed as every manipulation or normalization would imply the assumption of a corresponding processing model in the human listener.

4 Parameters

In order to evaluate the suitability of the parameters described above, all possible combinations of relevant parameters have to be investigated, i.e., none of the candidates of different parameter sections (units, properties, property measures, comparison) can be evaluated in isolation. Also, although not likely, each of the combinations of the parameters are possible configurations that match well with the stress location of words. Altogether, 276 parameter configurations have to be evaluated and compared to lexical stress positions.

Units

Morphological categories were lexical morphs (L {*hund, katz-, maus,...*}), free prefixes (P {*vor, nach, bei, an,...*}), bound prefixes (p {*be-, ver-, ent-, ge-,...*}), non-native derivational suffixes (S {*-ion, -al, -ät, -ur,...*}), native derivational suffixes (s {*-lich, -ung, -bar, -heit,...*}), and inflectional suffixes (f {*-e, -t, -en, -s,...*}). For this investigation, words were divided into morphs and pseudo-morphs, no morph could have more that one vowel. Morphs without vowels were assigned to the preceding morph.

Syllables were labeled according to a morphological and a distributional principle: The morphological principle is to mark syllable boundaries in front of lexical morphs (*Taustrick* [taU.StRIk] vs. *Tauschtrick* [taUS.tRIk]). The distributional principle separates intervocalic consonant clusters into subclusters of consonants occurring at the end or at the beginning of words (*Hälfte* [hElf.te], *knipste* [knIps.t@]).

Vowels (VO) and vowel onset intervals (VI) can be identified as illustrated in Figure 1.

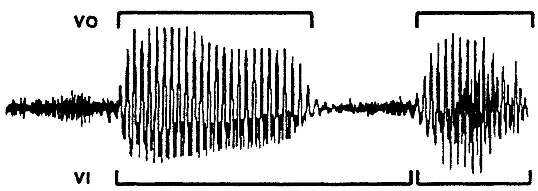

Figure 1: Vowels and vowel onset intervals.

Signal properties

Fundamental frequency was measured for voiced parts of the words. F_0 was only evaluated for sounds considered voiced. Energy was measured in RMS values of the speech signal. Both values were obtained using the function `get_f0` of the signal analysis package ESPS (Entropics). Duration was used as provided in the label data.

Signal property measures

The minimum, maximum and mean values of energy and f_0 were computed. The measure slope is the mean slope; lift is the difference of maximum and minimum of a property.

Comparisons

For this study, parameters to be compared to lexical stress positions were the absolute values of the first and the second unit, their difference, their ratio and the position of the higher value. Additionally, a balance measure was used to represent the measures of two units within a word by calculating an imaginary point of balance. Balance values range between 0 and 1 and can be used for utterances with any number of syllables.

$$M_B = \frac{1}{n-1} \sum_{i=1}^{n} \frac{v_i\,(i-1)}{\sum_{k=1}^{n} v_k}$$

(n: number of units; v_i, v_k: values of relevant units)

Two measurement criteria were chosen for evaluating the correspondence of parameter configurations and lexical stress positions of words. In order to account for the fact that the number of words stressed on the first syllable is ten times higher than those stressed on the second, the first criterion is a recognition design using all data: Does the absolute comparison of two parameters predict the lexical stress position? For the second measurement criterion ten selections of each 1,000 randomly selected word utterances (500 first-syllable stressed and 500 second-syllable stressed) are used and investigated computing Pearson correlations.

5 Results

The following figures give results across selected parameters that were compared while other parameters being held constant:

Figure 2 shows results for different units for duration and the balance measure. The duration of morphological units corresponds best to the variation of stress positions. In the recognition design, vowel onset intervals are second best, followed by vowels.