Word stress correlates in spontaneous child-directed speech in German

Katrin Schneider, Bernd Möbius

Institute of Natural Language Processing, Experimental Phonetics Group,
University of Stuttgart, Germany
katrin.schneider@ims.uni-stuttgart.de, bernd.moebius@ims.uni-stuttgart.de

Abstract

In this paper we focus on the use of acoustic as well as voice quality parameters to mark word stress in German. Our aim was to identify the speech parameters parents use to indicate word stress differences to their children. Therefore, mothers and their children were recorded during a period of at least one year while they performed a special playing task using word pairs that differ only in the position of word stress. The recorded target words were analyzed acoustically and with respect to voice quality. The results presented here concern the mothers' productions of contrastive word stress, and we discuss our findings with respect to the results of previous studies investigating word stress. Our results provide further insight into the process of word stress acquisition in German.

Index Terms: production, prosody, word stress, acoustics, voice quality

1. Introduction

Jessen and colleagues [1] demonstrated that in read speech vowel duration is the most reliable correlate of word stress in German adults. They found that in identical contexts stressed vowels are always produced with a longer total duration than unstressed vowels.

The voice quality parameter spectral tilt describes the slope of the spectrum, i.e. it compares the decrease of the amplitudes of the higher frequencies with respect to the amplitudes of the lower frequencies in the spectrum. The parameters *skewness (SK)*, which corresponds to the abruptness of the glottal closure, and *rate of closure (RC)*, which corresponds to the velocity of the glottal closure, are known to be good indicators for spectral tilt. In their analysis of read speech of German adults Claßen et al. [2] reported that unstressed vowels show significantly greater *SK* values than stressed vowels. The results for the parameter *RC* show similar tendencies but they are not as reliable for word stress as the results for *SK*. The authors concluded that stressed vowels in German are produced with a faster and more abrupt closure of the vocal folds than unstressed vowels, which results in higher *RC* and *SK* values, as well as in a higher spectral tilt for unstressed vowels compared to stressed ones. In German, the higher frequencies in the spectrum of stressed vowels seem to decrease more slowly than in unstressed vowels, which results in a steeper spectral slope for unstressed vowels.

Sluijter [3] demonstrated that the parameter *completeness of closure (CC)* is a reliable correlate of word stress in English, with stressed vowels having a greater bandwidth of the first formant (glottal leakage during the vowel production) than unstressed vowels. Her interpretation of these findings was that due to the higher subglottal pressure in stressed vowels the vocal folds are not entirely closed over the phonation cycle, which results in higher CC values for stressed vowels. For the parameters *skewness (SK)* and *rate of closure (RC)* she stated that they are correlates of both word stress and sentence stress, but the results of Claßen et al. let us suppose that, at least in German, *SK* and *RC* might be reliable correlates of word stress only.

In this study, we analyzed spontaneous target word productions of three mothers in their interactions with their children, i.e. during a special playing task. The productions were recorded during several sessions over the course of at least one year. The information received from the adults' data provide the basis for the analyses of the child data because the hypothesis is that children adopt the features their parents use to indicate word stress. By the use of artificially constructed names for animal toys that are geared to the typical phoneme sequences in the production of German infants we guarantee that the input for the stress patterns of these target words only comes from their mothers. We therefore analyzed only these target words, and the results of the mothers which are presented here serve as the basis for the interpretation of the findings in the child data. Furthermore the mothers' results are compared to the findings of former studies and correlations and differences are discussed.

2. Methods

2.1. Stimuli

The data reported here are part of a larger investigation of children's acquisition of contrastive stress in German. In this larger project recordings of monolingual German speaking children and their parents were made at their homes while they played together with five pairs of animal toys. The names of these animal toys, which are the target words in this study, are bisyllabic or trisyllabic words that consist of consonant-vowel (CV) syllables and contain only speech sounds that are acquired as the first ones by German children: the vowels /a i o/ and the consonants /b d m n/. The CV syllables are phonetically similar to the reduplicated babbling and to a child's first words. According to the TAKI task design proposed by Allen [4], the names within each pair of animals differ only in the position of word stress (Table 1).

Animal toy 1		Animal toy 2	
brown bear	['bimo]	polar bear	[bi'mo]
big zebra	['nami]	small zebra	[na'mi]
otter	['doba]	badger	[do'ba]
big tiger	['midano]	small tiger	[mida'no]
flying eagle	['badoni]	standing eagle	[bado'ni]

Table 1: Animal toys and their names used in the TAKI task. The names are the target words of our study.

In this contrastive stress condition, stressed and unstressed vowels can be compared in identical segmental contexts. Both stress positions in the target words are possible stress patterns in German, although stress on the second or third syllable is less common than stress on the first syllable [5]. In this study, only target words produced by the adults were analyzed.

2.2. Participants

The adult data presented here are the target word productions by three mothers recorded over a period of at least one year. Detailed information about the analyzed recordings are given in Table 2.

mother	# sessions	total length	# target words
AD	9	225 min	340
BL	9	65 min	190
KS	14	225 min	510

Table 2: Participant information (KS=first author).

2.3. Acoustic measurements

All recorded utterances of the mothers were annotated on the segmental and word levels by two trained labellers. The absolute duration, F_0 contour, the first four formants (F1–F4) and root mean square (RMS) intensity of each stressed and unstressed target vowel were measured. The voice quality parameters given in Table 3 are well-established and have been used several other studies (e.g. [3],[1],[2]). In this study they were automatically estimated from their acoustic correlates based on amplitude and frequency measurements at harmonic spectral peaks [6]. A correction, denoted by a tilde (˜), was made for the effect of the formants on the amplitudes of the first and the second harmonics.

voice quality parameter	formula
completeness of closure (CC)	$B1$
incompleteness of closure (IC)	$B1/F1$
open quotient (OQ)	$H1^{\tilde{}} - H2^{\tilde{}}$
glottal opening (GO)	$H1^{\tilde{}} - A1^{\tilde{}}$
skewness (SK)	$H1^{\tilde{}} - A2^{\tilde{}}$
rate of closure (RC)	$H1^{\tilde{}} - A3^{\tilde{}}$
T4	$H1^{\tilde{}} - A4^{\tilde{}}$

Table 3: Automatically extracted voice quality parameters. $B1$ corresponds to the bandwidth of $F1$. *IC* is a $F1$-normalized version of the *CC* parameter. $H1$ denotes the first, and $H2$ denotes the second harmonic in the spectrum, and $A1$ to $A4$ correspond to the amplitudes of $F1$ to $F4$, respectively.

All acoustic measurements were made at five relative temporal positions within each target vowel (at 10%, 30%, 50%, 70% and 90% of the total vowel duration) to capture parameter changes over the duration of the speech sound instead of analyzing global mean or median values for each target vowel.

2.4. Data analysis

2.4.1. Analysis restriction

Although we analyzed adult speech data, we found, maybe due to the use of spontaneous speech, that most of the vowels in medial position in the trisyllabic words, which were always produced unstressed, were realized with a shorter duration and, furthermore, often reduced to [ə]-like productions of the underlying vowel. These shortened and reduced vowels were excluded from further analyses. Thus, only full vowels produced in stressed positions were compared to full vowels produced in unstressed positions.

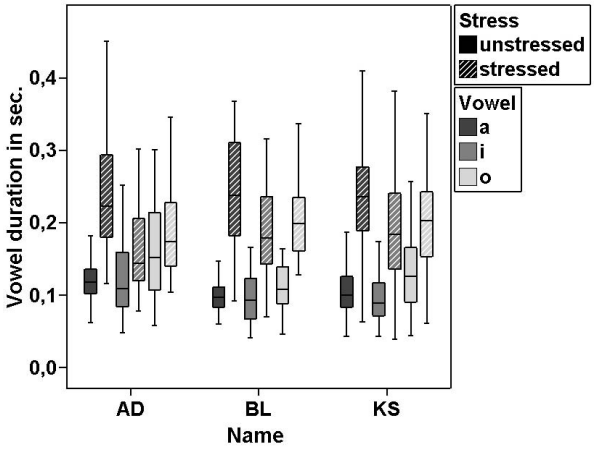

Figure 1: Vowel duration of stressed (hatched boxes) and unstressed (plain-colored) vowels for all three mothers by vowel quality (different shades of grey).

2.4.2. Relative stress parameters

Word stress is relative and it only emerges in polysyllabic structures. In previous studies we compared all unstressed vowels with all stressed vowels of the same vowel quality. In the present paper we additionally carried out analyses in which parameter differences between the stressed and the unstressed (full) vowel of each word were considered. The results of these analyses are taken as parameters indicating word-internal stress.

2.4.3. Statistical analysis

The SPSS program version 12.0 was used for statistical analysis. An univariate ANOVA analysis was carried out for each parameter separately with the parameter itself as the dependent variable and *word stress* as the independent variable, followed by a posthoc test (Waller-Duncan).

2.4.4. Normalization

Most of the analyzed parameters turned out to be vowel-specific as well as speaker-dependent. We therefore analyzed both raw and z-transformed parameter values. The z-transformation eliminates vowel-specific effects by replacing absolute values by their difference to the vowel-specific mean and dividing this difference by the vowel-specific standard deviation of this parameter. Furthermore, we analyzed each speaker separately.

3. Results: Word prosodic parameters

3.1. Vowel duration

Vowel duration has been shown to be the most reliable cue for word stress in the speech of German adults [1]. In our data, there were slight vowel-specific duration differences between

the three analyzed vowels /a/, /i/, and /o/.However, for all three speakers, stressed vowels are significantly longer ($p < 0.001$) than unstressed vowels (Figure 1). This effect is observed for all vowels pooled, for each vowel phoneme separately, and also for the position of the vowel phoneme in the word. We also found evidence for the lengthening of word-final syllables, but the strength of this final lengthening is speaker-dependent.

Our results show strong evidence that vowel duration is a reliable correlate of word stress in German adults, confirming the results of previous studies (e.g. [1]).

3.2. Root mean square intensity (RMS)

No significant difference between the RMS intensity values of stressed vowels compared to unstressed vowels was found for the three speakers. Furthermore, if there was a significant difference for the RMS intensity values of the vowels within a target word, it was always greater for the first vowel than for the second or the last vowel in the word, regardless of the stress position for this word. We therefore conclude, that in German the parameter RMS intensity is an indicator of syllable position within a word rather than of word stress.

4. Results: Voice quality parameters

4.1. Incompleteness of closure

Sluijter [3] found that stressed vowels show higher values of the voice quality parameter *completeness of closure (CC)* corresponding to a greater F1 bandwidth (B1), suggesting that stressed vowels have a greater glottal leakage than unstressed vowels. In our data *CC* was not correlated with word stress and, furthermore, it seemed to be vowel dependent. Therefore, we used a normalized version of this parameter, *incompleteness of closure (IC)* [7], in which the influence of the first formant on the *CC* values is minimized (see Table 3).

Our results indicate that the parameter *IC* is a good cue for word stress in German child-directed speech. All three participants demonstrate higher z-transformed *IC* values for unstressed vowels compared to stressed ones, indicating that unstressed vowels are produced with a greater glottal leakage than stressed vowels. However, looking at the raw data (Figure 2), there are slight individual differences. While AD always produces unstressed vowels with higher *IC* values than stressed vowels, BL and KS display almost no difference for /i/ vowels, and KS displays no difference for /a/ vowels. Furthermore, BL produces the /o/ vowels with the opposite tendency for this parameter.

4.2. Skewness, rate of closure, and T4

Claßen et al. [2] found that in German the parameter spectral tilt is a reliable correlate of word stress, with unstressed vowels having greater spectral tilt values than stressed vowels. They concluded that in the production of stressed vowels the vocal folds are closed faster (parameter *RC*) and more abruptly (parameter *SK*) than in unstressed vowels, which results in higher values for unstressed vowels compared to stressed ones. Therefore, spectral tilt was seen as a cue for word stress in German.

In our data we found supporting evidence for the hypotheses of Claßen [2] and Jessen [1], as well as some contradictory behavior of these parameters for the analyzed vowel qualities.

For the parameter *skewness (SK)*, all three speakers show higher z-transformed values for unstressed vowels compared to stressed vowels. Looking at the raw data, only the vowel /i/

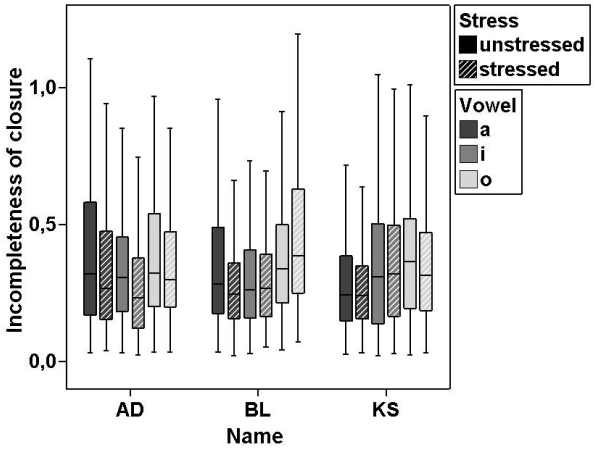

Figure 2: IC values for all three speakers and target vowels.

behaves uniformly for all participants with unstressed vowels having higher *SK* values than stressed ones. For the two other vowel qualities the results of the subjects differ. While KS always produces unstressed vowels with higher *SK* values than stressed vowels, AD and BL display the opposite behavior in the *SK* values for one of the two remaining vowel qualities.

For the parameter *rate of closure (RC)*, the z-transformed *RC* values are always higher for unstressed vowels compared to stressed ones. Our results show that for this parameter there is more consensus between the behavior of the different participants than for the parameter *SK*: BL always produces unstressed vowels with higher *RC* values than stressed vowels. AD and KS show the same behavior for the vowels /i/ and /o/, but for the vowel /a/ they display the opposite tendency in the use of the *RC* parameter.

The parameter *T4* corresponds to an extrapolation of the calculation of the parameters *GO*, *SK*, and *RC* in that the next higher formant (F4) and its amplitude are involved. This extrapolation was motivated by the results obtained in the studies by Claßen et al. [2] and Sluijter [3] for the given voice quality parameters. The slope of the spectrum seems to be a reliable correlate for word stress in German. Therefore it was necessary to extend the set of the voice quality parameters and include the influence of the fourth formant and its amplitude.

It was expected that the *T4* values for stressed and unstressed vowels are comparable to the results found for the parameters *SK* and *RC*. Our data support this hypothesis, i.e. the use of the parameter *T4* as an indicator of word stress is identical to the use of the parameter *RC* for all vowels and all participants, except that for the vowel /a/ BL did not distinguish between stressed and unstressed productions.

5. Discussion

Vowel duration seems to be the most robust correlate of word stress in German. All subjects in our study use vowel duration to differentiate between stressed and unstressed vowels. In the absence of strong reduction effects as a function of stress in full vowels, vowel duration is one of the most salient cues for syllabic stress in German.

We found strong evidence that vowel duration is not the only acoustic correlate to mark word stress in German, as already proposed by other researchers. The parameter *incom-*

pleteness of closure (IC), an F1-normalized version of the well-established *CC* parameter, turned out to be a good cue for word stress in German child-directed speech. *Incompleteness of closure* expresses the degree of glottal leakage. The higher the *IC* value, the greater the glottal leakage. All three subjects demonstrate higher z-transformed *IC* values for unstressed vowels compared to stressed ones, which corresponds to the original hypothesis of Sluijter [3], even though her own results did not confirm it. We assume that the effect we found may be caused by the increased glottal pressure with which stressed vowels are produced. Due to higher glottal pressure the air flows faster through the vocal folds, which results in a greater Bernoulli force. This is responsible for the glottis being faster and more completely closed than in unstressed vowels. In unstressed vowels there is a weaker subglottal pressure, resulting in a weaker Bernoulli force. Therefore it is more likely that a part of the glottis remains open during the entire phonation cycle in the production of unstressed vowels. However, the speaker-specific differences in the raw data for this parameter have to be further analyzed. Our hypothesis is that they might be caused by age differences between the children: the child of AD was the youngest one in this study, while BL and KS talked to older children.

RMS intensity is no cue for word stress in German, neither for the general comparison between stressed and unstressed vowels nor as a stress indicator of the word stress position inside the word. The first syllable has the highest RMS intensity values followed by the RMS intensity values for the following syllables, regardless of the position of word stress. Over the course of an utterance, the subglottal pressure decreases, which results in a decrease of the RMS intensity values over an utterance as well as over the total duration of each single word. Therefore, the first syllable of a polysyllabic word has the highest RMS intensity values compared to the RMS intensity values of the following syllables.

Our results for the voice quality parameters that describe the slope of the spectrum, i.e. spectral tilt corresponding to the parameters *SK*, *RC* and the new parameter *T4*, support the findings of previous studies, but not in all details. The z-transformed values of all three parameters and all analyzed participants confirm previous findings that stressed vowels in German are produced with a faster and more abrupt closure of the vocal folds than unstressed vowels, resulting in higher z-transformed values for the unstressed vowels. Using the raw data, vowel-specific as well as speaker-specific differences emerge. These differences are greater for *SK* compared to *RC* and *T4*. One possible explanation is that there are more differences between the second formants of the vowels /a/, /i/, and /o/, than between the third or fourth formants of these vowels, and that speakers use different strategies to mark stressed vowels with the first and second formants and their amplitudes.

The differences between the results of our study and previous investigations on word stress might be due to our use of spontaneous, child-directed instead of controlled adult-directed speech data and to different measuring methods. Our results show that with respect to the parameter changes over the duration of a vowel, there are differences between speakers and between vowel qualities.

6. Conclusion

We analyzed the spontaneous productions of three mothers in conversations with their children using the Taki task design with respect to the acoustic and voice quality parameters of word stress in German. The results we obtained should serve as a basis for the analysis of the children's productions of the same words to get an insight into the process of word stress acquisition.

Our results show that vowel duration is the most reliable correlate of word stress in German adults. All speakers produce stressed vowels with longer duration than unstressed vowels.

Incompleteness of closure appeared to be a good cue for word stress with unstressed vowels produced with higher z-transformed *IC* values, i.e. they are produced with a greater glottal leakage, than stressed vowels.

RMS intensity values always decrease over the total duration of a word and are therefore not useful as an indicator for word stress.

The voice quality parameters *skewness*, *rate of closure*, and the new parameter *T4* confirm previous findings that stressed vowels in German are produced with a faster and more abrupt closure of the vocal folds than unstressed vowels. This was shown for the z-transformed values of all three parameters, although in the raw data we found vowel-specific differences.

The findings of the adults' productions concerning word stress can now be compared to the children's use of these parameters to mark word stress.

7. Acknowledgements

This study was carried out as part of the project *An exemplar-theoretic model of the acquisition of acoustic correlates of stress* funded by the German Research Foundation (DFG Grant MO 597/1).

8. References

[1] Jessen, M., Marasek, K., Schneider, K., Claßen, K., "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German", Proc. 13th ICPhS, Stockholm, 4, pp. 428–431, 1995.

[2] Claßen, K., Dogil, G., Jessen, M., Marasek, K., Wokurek, W., "Stimmqualität als Korrelat der Wortbetonung im Deutschen", Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 1998.

[3] Sluijter, A. M. C., "Phonetic correlates of stress and accent", PhD Thesis, University of Leiden, 1995.

[4] Allen, G. D., "Development of prosodic phonology in children's speech: Further evidence from the TAKI task ", In: Dressler, W. U., Pfeiffer, O. E., Rennison, J. R. (eds), Phonologica 1980, Insbrucker Beiträge zur Sprachwissenschaft, 36, pp. 9–14, 1981.

[5] Baayen, H., Piepenbrock, R., Gulikers, L., "The CELEX Lexical Database", Release 2, CD-ROM, Centre for Lexical Information, Max Planck Institute for Psychlinguistics, Nijmegen, Linguistic Data Consortium, University of Pensylvania, 1995.

[6] Wokurek, W., Pützer, M., "Automated corpus based spectral measurements of voice quality parameters", Proc. 15th ICPhS, Barcelona, pp. 2173–2176, 2003.

[7] Pützer, M., Wokurek, W., "Multiparametrische Stimmprofildifferenzierung zu männlichen und weiblichen Normalstimmen auf der Grundlage akustischer Analysen", Laryngol Rhino Otol, 85, pp. 1–8, 2006.