Acoustic correlates of contrastive stress in German children

Britta Lintfert, Katrin Schneider
Institute of Natural Language Processing
University of Stuttgart, Germany
{britta.lintfert,katrin.schneider}@ims.uni-stuttgart.de

Abstract
This study examines the acoustic correlates of stress in children’s productions of bisyllabic and trisyllabic words, differing in stress placement. Acoustic results of the data of two children aged 4;11 to 6;1 are reported as a part of a more comprehensive study on the acquisition of stress in German. In contrast to recent findings that infants show an early preference in perception for rhythmic patterns in their own language, contrastive stress is supposed to be acquired quite late. Sufficient input must be available to the infants for building up representations of contrastive stress. We found that children around the age of five are able to perceive contrastive word stress but that they may differ in their production from one another. For German, vowel length is the most reliable correlate of word stress in both children’s utterances. Intensity as well as the voice quality parameters glottal opening and spectral tilt, including rate of closure and skewness, tend to be used adult-like.

1. Introduction
Perception studies show that prelinguistic infants are sensitive to the suprasegmental structure of their native language [1]. Even 6- to 12-week-old infants are able to discriminate two languages that differ in head direction and in its prosodic correlates while the other phonological properties are similar [2]. By 9 months of age, infants show a selective preference for stress patterns of their native language, and before the onset of meaningful speech, infants are sensitive to the frequency of the sound patterns in the ambient language [3]. Infants have to be sensitive to their native language’s sound organization, its phonotactic structure and its word structure in order to segment words in fluent speech [4]. It seems that syllables are important units of speech processing for infants [5] and rhythm and perceptually salient syllables are two important prosodic cues for segmentation [3]. Infants rely on the prosodic characteristics of their native language to infer its syntactic properties. Variability makes it difficult to build up phonemic representations of the sounds of the native language. The acoustic realization of phonemes varies – depending on speaker, speech rate or phonetic context – but listeners are able to assign different realizations of a given sound to the same category, a concept which is referred to as phonemic constancy.

Many infant speech perception studies suggest that infants store specific information about voices and the words which they are exposed to. By using the widespread input from their parents babies establish expectations about the typical patterns of prosodic variations of utterances in the ambient language. They obtain sufficient exemplar variance to develop language-specific generalizations about the distribution of values in the auditory dimensions of the phonetic space [6]. Each familiar word form is encoded in terms of episodic memory traces of fine-grained parametric representations of the auditory and articulatory patterns. A further encoding in terms of coarser-grained generalization of each word form is suggested by the literature on effects of phonological type frequency [5, 7].

The relationship of phonetic cues to the hierarchical prosodic and intonational representation must be learned for each individual language [8]. In a model such as exemplar theory it is assumed that speech categories are stored as clouds of perceived exemplars in long-term memory. Speech perception and production are closely connected: during perception the phonetic parameters of the input determine its placement in the phonetic space; during production a set of exemplars of the respective category is randomly selected, and the median of this subset is calculated and produced with a random error [9].

To develop an exemplar-based model for the acquisition of contrastive stress in German we analyze the effects of stress on its main acoustic correlates, i.e. on fundamental frequency, on duration and on the quality of different vowels in German. We aim to evaluate when children are able to perceive contrastive stress, when they are able to produce it and how they realize it. Our hypothesis is that children adopt the features that their parents use to indicate stress.

2. Method
2.1. Participants
The data reported here are part of a larger investigation of children’s acquisition of stress. We present the data collected during 9 months and produced by the two oldest children of our study, one boy (aged 5;4 to 6;1) and one girl (aged 4;11 to 5;2). Both children live in monolingual German-speaking environments. They had no unusual prenatal, sensory or developmental concerns, or hearing problems. Their parents were recorded too.

2.2. Stimuli
According to the TAKI-task design proposed by Allen [10] we created four object pairs consisting of animal toys with bisyllabic or trisyllabic names. The names contain only vowels and consonants that are acquired as the first ones by German children, i.e. the vowels /a/, /i/ and /o/ and the consonants /b/, /d/, /m/ and /n/. The consonant-vowel (CV) syllables are phonetically similar to reduplicated babbling and to the child’s first words. The names within a pair of animals differ only in the position of word stress, e.g. /bi/mo/ vs. /bi/mo/ (contrastive stress). Contrastive stress is supposed to be acquired rather late in language development [11, 12] but it provides conditions for comparing stressed and unstressed vowels due to the identical segmental context. The different position of stress in the target words follow typical stress patterns in German, which we estimated from the CELEX database. We found that the majority
Table 1: Animal toys and their names with contrastive stress used in the TAKI-task.

| Animal          | Name          
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>brown bear</td>
<td>Bino ['bi:mo:]</td>
</tr>
<tr>
<td>polar bear</td>
<td>Bimo ['bi:mo:]</td>
</tr>
<tr>
<td>big zebra</td>
<td>Nami ['na:mi:']</td>
</tr>
<tr>
<td>small zebra</td>
<td>Nami ['na:mi:']</td>
</tr>
<tr>
<td>otter</td>
<td>Doba ['do:ba']</td>
</tr>
<tr>
<td>badger</td>
<td>Doba ['do:ba']</td>
</tr>
<tr>
<td>big tiger</td>
<td>Midano ['mi:dano']</td>
</tr>
<tr>
<td>small tiger</td>
<td>Midano ['mi:dano']</td>
</tr>
<tr>
<td>flying eagle</td>
<td>Badoni ['ba:doni']</td>
</tr>
<tr>
<td>standing eagle</td>
<td>Badoni ['ba:doni']</td>
</tr>
</tbody>
</table>

of German bisyllabic or trisyllabic words is stressed on the first syllable, followed by words having stress on the second syllable. Stress on the third syllable is also possible, but it is much less frequent. According to these results, we evaluated the production of contrastive stress between the first and the second and the first and the third syllable of a word. The names of the animal toys are listed in Table 1.

2.3. Recording procedure

The children were recorded every 6 to 8 weeks at their homes while playing with the animal toys. During playing the child was encouraged by one parent to use the names given to the animal toys. For the recordings a Digital Audio Tape Recorder (DAT-Recorder) Sony TCD-D100 with a sampling rate of 48 kHz (16 Bit linear) and a wireless microphone NADY LT-4 (Lavalier) E-701 (600 Ohm) were used with a mouth-to-microphone distance of approximately 15 cm.

2.4. Acoustic measurements

The recorded data were transferred to a computer. Each recording session was separately labelled by three trained labellers and analyzed. The absolute duration, F0 contour and root mean square intensity of each stressed and unstressed vowel in the target words were measured. Additional voice quality measurements were carried out using the acoustic analysis based on amplitude and frequency measurements at harmonic spectral peaks [13]. The voice quality parameters measured are open quotient (OQ), glottal opening (GO), spectral tilt (ST), skewness (SK), root mean square of each stressed and unstressed vowel in the target words. Second, the stressed vowels of the female child occurring in the word’s last syllable are significantly shorter than his unstressed vowels in the word’s first syllable. Second, the stressed vowels of the female child occurring in the word’s last syllable are significantly longer than those for /a/ and /o/ than for /i/. This supports the dependence of OQ on muscle tension because due to lip spreading and lip rounding for /i/ and /o/, respectively, higher muscle tension in the larynx is needed during production. As for the adults, no dependency on word stress was found for this parameter in the child data.

3. Results

For the statistical analyses the SPSS program version 12.0 was used. A univariate GLM analysis (generalized linear model) was carried out for each parameter followed by posthoc tests (Tukey-HSD and Waller-Duncan).

3.1. Perception of word stress

As we adopt the hypothesis of Pierrehumbert [8] that perception leads production we had to check whether the children correctly assign the names to the appropriate animal toys. Both children did this constantly at the rate of nearly 100% correct, i.e. whenever they made a mistake in their assignment they corrected themselves immediately. We conclude that perception is correct and therefore analyzed the children’s production to see how they produce contrastive stress.

3.2. Parameters independent of word stress

We found that fundamental frequency (F0) is vowel-specific but not depending on word stress for both children. This characteristic has already been reported for adults [16] and is now also found to be correct in childrens’ productions.

The open quotient (OQ) is considered to be a correlate of sentence accent and shows an additional modulation of muscular tension [16]. In the analyses of this parameter we found that both children showed significant higher OQ values for /i/ and /o/ than for /a/. This supports the dependence of OQ on muscle tension because due to lip spreading and lip rounding for /i/ and /o/, respectively, higher muscle tension in the larynx is needed during production. As for the adults, no dependency on word stress was found for this parameter in the child data.

3.3. Vowel length

Vowel length is seen as the most reliable cue for stress in German adults. In our data the length of the vowels differed strongly between the different vowel phonemes. Therefore, for our analyses, we transferred the absolute length values into z-scores to normalize the values.

For both children stressed vowels are significantly longer than unstressed vowels (see Figure 1). This dependence is found for all vowels pooled as well as for each vowel phoneme in our data.

Furthermore, vowel length depends on the combination of vowel position in the word and stress. Here, the results of the two children differed slightly. First, whereas the unstressed vowels of the female child are equal in length independent of their position in the word, the unstressed vowels of the male child occurring in the word’s last syllable are significantly shorter than his unstressed vowels in the word’s first syllable. Second, the stressed vowels of the female child occurring in the word’s last syllable are significantly longer than her stressed vowels in the first syllable of the word. She produces final lengthening to a great extent, which results in very long stressed vowels if they are located in the word’s last syllable. In contrast, the stressed vowels of the male child occurring in the word’s first syllable are significantly shorter than his stressed vowels in the word’s last syllable. This is also evidence for final lengthening, but he does not use it to the same extent as the girl does. The results are strong evidence for the fact that vowel length is a reliable correlate of word stress in German children.

3.4. Root mean square intensity

The girl’s RMS intensity values for stressed vowels are significantly higher than those for unstressed vowels ($p = 0.045$). This was not found for the RMS intensity values of the boy. There was a tendency in the presumed direction but it was not statistically significant. Therefore, RMS intensity is a correlate of stress for the female but not yet for the male child.

3.5. Glottal opening and completeness of closure

Glottal opening (GO) was shown by Jessen et al. [14] to be a correlate of stress in German adults. They found that stressed vowels have lower GO values than unstressed ones. In our analyses, both children show the same tendency for /a/ and /o/, but not for /i/.
Sluijter [16] found that stressed vowels show higher CC values, i.e. they have a greater glottal leakage than unstressed vowels. This was not confirmed in our data. Both children showed greater CC values for the unstressed vowels than for the stressed ones, which is contradictory to the adult use of CC.

3.6. Spectral tilt, skewness and rate of closure
According to Classen et al. [15] spectral tilt (ST) is a correlate of stress in German adults. Unstressed vowels show higher ST values than stressed ones. They found the same results for the parameters skewness (SK) and rate of closure (RC), but these were not as reliable as those for ST.

In our study, the female child shows the tendency to use all three parameters adult-like, but this was not significant. We conclude that for this child these parameters are not yet actively used.

In contrast, the male child did show a dependence on stress for ST as well as for RC. As shown in Figure 2, his ST values for unstressed vowels are significantly higher than those for stressed vowels. This holds for the vowels in total as well as for every analyzed vowel phoneme. His RC values are also significantly higher for unstressed than for stressed vowels. Only his SK values show no dependency on stress, although they tend to behave in an adult-like manner. We conclude that the male child has already adopted the parameter settings which are observed in German adults.

4. Discussion
The children's data analyzed in this study provide clear evidence for the fact that these children are able to perceive and to produce contrastive stress. According to the hypothesis that perception leads production we looked for acoustic cues in the speech data of these children that are used to signal word stress.

In accordance with results obtained by previous studies on adults' productions we found that in the children's productions neither fundamental frequency nor open quotient were correlates of word stress in German.

Vowel length seems to be the most reliable cue for word stress in German, even in children at the age analyzed here. Both children use vowel length to differentiate between stressed and unstressed vowels. There are slight differences between the two children, especially in the use of final lengthening of stressed vowels. There may be two interpretations of this distinct behavior: first, the parents of the female child may also use vowel length to a large extent for marking word stress, whereas the parents of the male child do not; second, it might be a way of playing with this parameter in order to find out how it can be used with minimal effort to achieve the desired result. For clarification, further data of both children and their parents have to be analyzed.

Root mean square intensity is also known to be a correlate of word stress in German adults. This finding was confirmed for the female but not for the male child here. Again, the same two interpretations as for vowel length are possible, i.e. the parents of one child use this parameter in a different way than the parents of the other child, or the children are exploring this parameter. This needs to be further investigated.

For the /a/ and /i/ vowels of both children glottal opening depends on stress, as for adults. For the male child this dependency reaches significance, but not for the female child. In both children the GO values of the vowel /o/ behave contradictory to the adult parameter values. Maybe the /o/ vowels are produced differently than the two other vowels. In both children the /o/ vowels are always the longest ones independent of stress. So it might not be possible for the children to use GO as a stress cue, because due to the sometimes extreme length of the /o/ vowels, especially in stressed position, the children are not able to maintain the stronger subglottal pressure typical for stressed vowels over this longer period of time. However, the general tendency supports the hypothesis that children will acquire GO as a stress cue in German later on.

The completeness of closure values suggest that the unstressed vowels have a greater bandwidth of the first formant than the stressed vowels. This is in contrast to the findings of Sluijter [16] for Dutch adults and the tendency found by Clasen et al. [15] for German adults. Sluijter justified her results with the assumption that the subglottal pressure built up for stressed
vowels also results in greater glottal leakage. We conclude that children are not able to control their subglottal pressure the same way adults do. So, in contrast to adults, children produce stressed vowels with complete closure resulting in higher CC values for unstressed vowels than for stressed ones. Analyzing this results, CC is a correlate of word stress in German children but in the opposite way of how adults use it. This has to be further investigated in order to find out whether these children change their use of CC and when this takes place.

The parameters skewness (SK) and rate of closure (RC) describe the vibration behavior of the vocal folds. For German, Claßen et al.[15] found that the closure of the vocal folds is much more abrupt (parameter SK) and faster (parameter RC) in stressed than in unstressed vowels, and that these are therefore reliable cues for stress in German. Our child data support this conclusion: the SK as well as the RC values show a tendency for adult-like behavior in both children. Concerning spectral tilt, for which Claßen et al. found that the closure of the vocal folds is slower and smoother for unstressed than for stressed vowels in German, we found clear evidence in our data that children at the age of five already use this parameter as adults do (the boy’s data) or show a tendency for adult-like use (the girl’s data). This supports the interpretation that spectral tilt is indeed a reliable correlate of word stress in German.

5. Conclusion

The child data presented in this study suggest that children acquiring German as their native language use different strategies to produce word stress. The most reliable correlate of word stress in German children around the age of five is vowel length. Other stress cues reported for adults, including RMS intensity, spectral tilt, skewness, rate of closure, glottal opening and completeness of closure, already show tendencies of an adult-like use of these parameters, but they are not yet stable.

To interpret the differences between the two children in the production of word stress, further data of these children and their parents have to be analyzed and the findings have to be evaluated by the data of more children. This is work in progress.

6. Acknowledgments

We thank Nicolai Hezinger, Stefan Kombrink and Manuel Kountz for labelling the speech data, as well as the children and their parents for participating in the recordings. This experiment was carried out as part of the project Ein exemplartheoretisches Modell zum Erwerb der akustischen Korrelate der Betonung funded by the German Research Foundation (DFG, Grant MO579/1-1).

7. References