Acquisition of syllabic prominence in German speaking children

Britta Lintfert[1], *Bernd Möbius*[1],[2]

[1]Institute of Natural Language Processing, University of Stuttgart, Germany

britta.lintfert@ims.uni-stuttgart.de, moebius@ifk.uni-bonn.de

Abstract

An investigation of the acoustic correlates of word stress in infant polysyllabic vocalization was carried out on the basis of data from 6 German-learning infants between 7 and 36 months of age in order to describe the development of word stress in German. The data were analyzed with respect to duration, intensity, fundamental frequency (f_0), as well as vowel quality parameters describing the time and degree of opening of the glottis, the slope of the spectrum and glottal leakage. With beginning of babbling children are able to produce different stress patterns. However, the implementation and usage of the parameters contributing to marking word stress appear to be inconsistent. The children used all measured acoustic parameters for marking word stress but the usage of the parameters depended on age and on the individual child. The most important cue to mark different stress patterns is to learn to reduce the acoustic parameters for the production of unstressed vowels.

Index Terms: acquisition, prominence, syllabic stress, longitudinal data

1. Introduction

Pre-linguistic infants are sensitive to the rhythm of their native language and discover specific prosodic properties of their native language already early in age [1]. They are sensitive to the frequency of the stress patterns in their the native language [2], and they are able to encode stress information and build up cognitive representations and categorizations of these patterns. The role of stress is particularly relevant to prosodic development since early truncated productions are determined by syllable salience. It seems that even pre-linguistic infants are sensitive to aspects of stress and encode stress information as part of their proto-lexical representations [3]. These proto-lexical representations are established in the pre-linguistic period, trained during babbling and help to build up more adult-like representations which are used when children begin to establish a lexicon [4]. If the sensitivity to metrical stress is less developeda dysfunction in the acquisition of speech can occur [5]. Between the beginning of babbling to the onset of language-specific sounds, children have to discover, organize and consolidate the input of the native language on the basis of their own perceptual and articulatory capabilities [6]. For reaching adult-like production of prosody the pre-linguistic infant has to acquire fine motor adjustments affecting fundamental frequency, timing and intensity over more than one syllable. The emergence of control of the acoustic cues from first babbling to the onset of word use is essential for the prosodic learning mechanism. When word production begins to take over babbled utterances in infants' vocalizations, the integration of adult-like prosodic cues into the underlying segmental patterns has to be learnt and activated [7]. Against this background we analyzed the development of syllabic prominence in German speaking children. We describe the use of the different acoustic correlates of stress depending on their development and increasing motor control of six children (3 boys and 3 girls) between 5 and 36 months of age. In German speech the most reliable parameter to mark stress is vowel duration [8] followed by intensity, fundamental frequency, formants, and several voice quality parameters [9]. With regard to the variable use of the acoustic correlates of stress, the preferred correlates to mark stress depending on age and child are described.

2. Method

2.1. Participants and data collection

This acoustic study is based on longitudinal data from six typically developed monolingual German children (3 boys and 3 girls) between 5 and 36 months. The data are part of the Stuttgart Child Language Corpus [10]. The recordings took place at the childrens' homes in familiar play situations with their parents. The children were recorded during naturalistic interactions with their parents while looking at picture books or playing with toys. Thus, the data represent spontaneous productions of the children. However the setting was controlled to some degree as the parents were always using the same picture book during the babbling and first word production phases to motivate comparable productions from the children.

All recordings were transferred to a computer workstation, down-sampled to 16 kHz and manually annotated on the segmental, syllable and word level by two trained transcribers. Since during babbling there are no clear preferences for marking stress, the perceived prominence of the produced syllables were annotated. First, we identified the syllables with any prominence against the syllables with no salience (unstressed). The syllables with any prominence then were categorized depending on their strength as primary stressed (most prominent) or secondary stressed (less prominent).

A second person carried out reliability judgments of 10% of the annotated data with respect of vowel duration, vowel identity and marked prominence. Across all categories the interjudge agreement was 92%.

2.2. Acoustic analysis

Absolute segment duration, the F_0 contour, the first four formants (F1–F4), and root mean square (RMS) intensity for each produced vowel in polysyllabic utterances (babbles, non-words, and words) were analyzed using LPC-based analysis methods. It is well known that formant measurements are particularly difficult to perform in high-pitched voices. The large distance between adjacent harmonics can lead to under-sampled spectra and this should be kept in mind for LPC analysis, in which formant measurements are greatly influenced by the closest harmonic [11]. However, LPC analysis is the procedure used in

acoustic	Voice quality parameter	calculation by Pützer/Wokurek (2006)
$2 * F_0$	opening quotient (OQ)	$OQ\tilde{} = (H1\tilde{} - H2\tilde{})$
$F1$	glottal opening (GO)	$GO\tilde{} = (H1\tilde{} - A1\tilde{})$
$F2$	skewness (SK)	$SK\tilde{} = (H1\tilde{} - A2\tilde{})$
$F3$	rate of closure (RC)	$RC\tilde{} = (H1\tilde{} - A3\tilde{})$
$F4$	$T4$	$T4\tilde{} = (H1\tilde{} - A4\tilde{})$
–	completeness of closure (CC)	$CC = B1$
–	incompleteness of closure (IC)	$IC = B1/F1$

Table 1: Automatically extracted voice quality parameters. *H1* describes the first, and *H2* the second harmonic in the spectrum. *A1* to *A4* correspond to the amplitudes of *F1* to *F4*. *B1* corresponds to the bandwidth of *F1*.

extensive studies of acoustic characteristics of child speech To avoid formant measurement errors the performance of the automatic program was evaluated using hand-measured formant values of 80 vocalic segments from one girl aged 22 months. The differences between automatic and manually measured formant values for this girl are comparable to differences found in other acoustic studies of child speech [12, 13]. It seems that from 15 months of age on, the produced formants can be categorized into different vowels and showed as being vowel-specific [10]. Therefore, voice quality parameters were also analyzed from that age on. Based on spectral differences of the corresponding harmonics and amplitudes of the first four formants [14, 15] different voice quality parameters were calculated (see Table 1).

All acoustic parameters were extracted at five points during the vowel, at 10, 30, 50, 70, and 90% of the marked vowel duration, to capture parameter changes over the duration of the speech sound [16].

As most of the analyzed parameters turned out to be vowel-specific as well as speaker-dependent, a normalization was done 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, each speaker was analyzed separately.

For each parameter, mixed analyses of variance (ANOVA) were carried out with the analyzed parameter as the dependent variable and stress and the child's age as factors. For vowel duration as well as RMS intensity the position of the syllable in the utterance (initial or final syllable) was also used as a factor, because vowel position in the utterance can have an effect on either of these parameters [9]. The level of significance was set at $p<0.05$. In the case of significance of a factor further analyses were carried out using the Tamhane post hoc test.

3. Results and Discussion

3.1. Vowel duration

The duration of the vowel (as well as the durations of other parts of the syllable and overall syllable duration) is generally regarded as the primary cue to German word stress. Vowel duration of the stressed syllable and the closure duration of plosives preceding the vowel are significantly longer than in unstressed syllables [8, 17].

Previous results from analyses of the production of contrastive stress showed that vowel duration is the most reliable cue to mark stress in older children [9] and adults [8, 18].

For all children in this study, a strong main effect of stress on vowel duration was found to be significant. Except for two children, the interaction of stress and age was also significant, which provides evidence for a developmental pattern and variable use of vowel duration to mark stress. Stressed vowels were produced with longer duration than unstressed vowels and secondary stressed vowels, but not in the same way at all ages. The mean duration of all vowels increases until the age of 12 months and then decreases. The children mainly produced no significant differences in vowel duration between unstressed and secondary stressed vowels.

In German speech the reduction of vowel length often goes along with reduced vowel quality. Unstressed vowels were more affected by the reduction in vowel quality than stressed vowels as only full vowels can carry word stress. The amount of reduction depends on the context. Therefore to produce a perceptible difference between stressed and unstressed syllables in an utterance the children not only have to learn to shorten the vowel but also to reduce the vowel quality. To produce this difference, the underlying acoustic pattern has to be learnt. The production of stressed syllables seems to be the unmarked case as only about 30% of the syllables produced during babbling are perceived as unstressed. Stressed syllables are produced most frequently, followed by syllables perceived as less prominent but not as weak (level stress). It seems that it is not the elongation of a vowel that needs to be learnt but the other way round, the reduction of syllable length for the production of unstressed vowels. The vowels in the final syllable were mainly produced with longer duration than the vowels in initial and medial syllables. Only for one boy no such tendency was observed. This production pattern provides evidence for final syllable lengthening in five of the six children observed.

3.2. RMS-Intensity

Intensity has been reported to be a reliable cue to mark stress. In German, stressed vowels are produced with more vocal effort that result in higher intensity (loudness) than unstressed vowels. However, it is generally considered to be a weak cue in the perception of linguistic stress as the contrast between the intensity of stressed and unstressed vowels is weaker than for other cues, e.g. duration [8]. The problem with intensity to mark stress is that it can be influenced by shifting formant frequencies due to stress. For an increase in intensity an increase in the physiological correlate of intensity, the subglottal pressure, is necessary [19, 17]. It has been showed that higher frequencies are more influenced by increasing vocal effort than lower frequencies [20].

Therefore differences in intensity, measured as RMS Intensity, depend on fine-tuning of the subglottal pressure. To use different intensities to mark word stress, children have to learn to produce vowels with differences in subglottal pressure. They must reduce as well as strengthen subglottal pressure, depending on the stress pattern.

For all six children a strong main effect of stress and age could be seen, as stress and age demonstrated significant influence on intensity for all children. An age-dependent use of intensity could be observed as the interaction of stress and age was also significant for all children. In summary, it can be stated that until 3 years, unstressed vowels were mainly produced with lower intensity than stressed vowels. For the production of unstressed vowels a reduction of intensity is needed. The production of vowels perceived as secondary stressed varied depending on the child. Variable production patterns were observed for the production of secondary stressed vowels.

3.3. Fundamental frequency

The phonetic realization of word stress also involves fundamental frequency changes. In German an increase of f_0 for stressed vowels can be observed [8]. Other studies suggest that for German adult speakers f_0 is vowel-specific and they may not rely on f_0 to mark word stress [18, 17]. As sentence intonation also depends on changes in fundamental frequency, these different results may depend on the different methods of analysis.

The principle mechanism for changing the fundamental frequency is to lengthened, thinned, or tensed vocal folds through contraction of the criothyroid muscle. An addition of longitudinal tension correlates with an increase in fundamental frequency [19]. But also the vertical tension of the folds and varying subglottal pressure have an affect on fundamental frequency. For the controlled use of fundamental frequency children have to learn how to apply more or less longitudinal tension to the vocal folds via the criothyroid muscles and the interaction with subglottal pressure.

The mixed ANOVAs showed a strong main effect of stress and of the interaction between stress and age. With increasing age a tendency towards a better control of fundamental frequency could be observed. Regression analysis showed a significant decrease of fundamental frequency with increasing age. When fundamental frequency was used to mark stress then unstressed vowels were produced with lower fundamental frequency than stressed ones. Some children produced secondary stressed vowels similar to stressed ones and other children more like unstressed ones. Therefore no clear production pattern for secondary stressed vowels could be derived.

3.4. Voice quality parameters

3.4.1. Open Quotient

Open Quotient (OQ) indicates the time during which the glottis is open, defined in the time domain as a fraction of the total period. It can be described with the differences between the first two harmonic peaks with the formant influence removed. *OQ* is considered to be a correlate of sentence accent and not for word stress and is possibly caused by an additional modulation of muscular tension [20, 21].

For three children stress showed a significant influence on OQ, indicating that stressed vowels were produced with lower *OQ* values than unstressed vowels. Secondary stressed vowels demonstrated no preference. As most of the children's' utterances were one-word sentences, an influence of intonation cannot be excluded. With increasing age an increase of *OQ* values could also be observed. As *OQ* seems to be more a correlate of sentence accent than word stress, an increase in *OQ* can be a sign for a more adult-like intonation contour.

3.4.2. Glottal Opening

Glottal Opening (GO) describes the degree of opening over the entire glottal cycle. The amplitude of the first formant (A1) is influenced by the glottal aperture during the open phase. *GO* has been shown to be a correlate of stress in German adults [8, 18]. In adult speech stressed vowels are produced with lower *GO* values than unstressed ones. For the children a significant influence of *GO* could be observed, but the use of this parameter was influenced by the age of the child. If the children used this parameter to mark stress, then it was mainly to mark secondary stressed vowels, because stressed and unstressed vowels were mainly produced with the same *GO* values.

3.4.3. Spectral tilt parameters

Spectral tilt describes the slope of the spectrum by comparing the decrease of the amplitudes of the higher frequencies with respect to the amplitudes of the lower frequencies in the spectrum. The parameter *Skewness (SK)*, which corresponds to the abruptness of the glottal closure, and the parameter *Rate of closure (RC)*, which corresponds to the velocity of the glottal closure, are known to be good indicators of spectral tilt [20, 18]. The parameter *T4* corresponds to the calculation of the parameters *GO*, *SK* and *RC*. In the calculation of *T4* the next highest formant (F4) and its amplitude is involved. *SK* and *RC* depend on word stress in adult speech s[18] and in child-directed speech [9]. As unstressed vowels are expected to be produced with a less abrupt and slower closing of the vocal folds, they should have higher values than stressed vowels. The same tendency is expected for *T4*.

The spectral tilt parameters *SK*, *RC* and *T4* demonstrated a significant influence of stress for three children. As the interaction of stress and age was significant for all children, it appears that the use of these parameters depends on age and undergoes some development during speech acquisition. With increasing age a more constant use of the parameters could be observed. The main tendency for all children was that stressed vowels were produced with lower values of the spectral tilt parameters than unstressed ones. All children tended to produce unstressed syllables with stronger spectral tilt. For secondary stressed vowels, no tendency could be observed as the use of the parameter to mark secondary stress differed between the children.

3.4.4. Glottal leakage

Completeness of Closure (CC) corresponds to the energy loss in the F1 range, adding significantly to the F1 bandwidth (B1) when the glottis is not completely closed during phonation. As CC appears to be vowel specific, a normalized version of this parameter called *Incompleteness of Closure (IC)* is also used in which the influence of the first formant on the CC values is minimized. Stressed vowels show greater glottal leakage and therefore higher *Completeness of Closure (CC)* and *Incompleteness of Closure (IC)* values for adults [9].

Only in interaction with age did stress showed a significant influence on the parameters *Completeness of Closure* and *Incompleteness of Closure*. The use of these parameters to mark stress varied between children and age, as they were not used at each age to produce word stress. If these parameters were significant for word stress, then it seems that secondary stressed vowels were produced with significantly different values than stressed and unstressed vowels. In summary, the children in this study used different voice quality parameters to produce word stress. As age was mostly significant, it can be concluded that the use of these parameters is variable and not stable until the age of 36 months. The use of the parameters varied depending on the child, but the parameters describing the vibration behaviour of the vocal folds were used by all children to produce word stress. The closure of the vocal folds had been found to be more abrupt and faster in stressed than in unstressed vowels. The parameters *GO*, *IC* and *CC* were mainly used to produce secondary stressed vowels, as stressed and unstressed vowels showed no significant differences.

4. Conclusions

The data analyzed in this study provide clear evidence that with the beginning of babbling at five months of age, children are able to produce perceivable different stress patterns. However, the implementation and usage of the parameters that contribute to marking word stress appear to be inconsistent. Each acoustic parameter potentially used for marking word stress is indeed observed. But even with advancing age the use of these parameters is still variable and depends on the child. The children's data suggest that children acquiring German as their native language use different strategies to produce stress. The most reliable correlates of word stress in German children during babbling are intensity and f_0. Vowel duration as the most reliable correlate to mark word stress for adult speakers was also used by the children but, until the age of about 36 months, not in a consistent way. With increasing age the children show tendencies to use all the stress cues reported for adults, such as vowel duration, intensity, glottal opening, skewness, rate of closure and completeness of closure, in their parents–like usage pattern, but not consistently. During babbling children experiment with different acoustic parameters to produce word stress in order to find out how they can be used with minimal effort to achieve the desired result based on their articulatory capabilities. These are trained during babbling. A variable use of the possible acoustic realization of stress can be observed. With the beginning of the production of first words a reorganization of the use of the acoustic parameters has to be done; however, until the age of 36 months the parameters are still not used in an adult-like way for stress production. The children are still too inconsistent in their use of the parameters to mark word stress. With the beginning of word production the trained patterns are compared to the input and self-monitored production to build up more adult-like production patterns. During babbling such a comparison is not necessary, because no production target has to be met. Given the perception–production loop [22], the childrens' abilities to produce a new category is a function of the number and quality of input elements, and depends on the continuing development of the control and coordination of particular structures in producing stress; and therefore the babbling phase with its proto–syllables is important. As most of the syllables during babbling are perceived as stressed and not as weak, it seems that the unmarked case is the production of stressed syllables. The development of the acoustic parameters of stress can therefore be characterized as involving increased control over the articulatory movements to derive perceptually identifiable stress contrasts. The most important cue to mark perceivable stress pattern is to learn to reduce the acoustic parameters. During acquisition, children have to learn to fine tune the different acoustic parameters to mark word stress. The unique phonetic target region for each phoneme in stressed and unstressed condition has to be established.

5. 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 597/3).

6. References

[1] A. Christophe, M. Nespor, M. T. Guasti, and B. van Ooyen, "Prosodic structure and syntactic acquisition: The case of the head–direction parameter," *Developmental Science*, vol. 6, no. 2, pp. 211–220, 2003.

[2] P. W. Jusczyk, P. A. Luce, and J. Charles-Luce, "Infants Sensitivity to Phonotactic Patterns in the Native Language," *Journal of Memory and Language*, vol. 33, pp. 630–645, 1994.

[3] S. L. Curtin, T. H. Mintz, and M. H. Christiansen, "Stress changes the representational landscape: Evidence from word segmentation," *Cognition*, vol. 96, no. 3, pp. 233–262, 2005.

[4] J. F. Werker and C. T. Fennell, "From listening to sounds to listening to words: Early steps in word learning," in *Weaving a Lexicon*, G. Hall and S. Waxman, Eds. Cambridge, Mass.: MIT Press, 2004, pp. 79–109.

[5] E. de Bree, P. van Alphen, P. Fikkert, and F. Wijnen, "Metrical stress in comprehension and production of Dutch chidlren at-risk of dyslexia," in *BUCLD 32: Proceedings of the 32nd annual conference on language development*, H. Chan, H. Jacob, and E. Kapia, Eds., 2008, pp. 60–71.

[6] B. L. Davis, P. F. MacNeilage, and C. Matyear, "Aquisition of Serial Complexity in Speech Production: A Comparison of Phonetic and Phonological Approaches to First Word Production," *Phonetica*, vol. 59, pp. 75–107, 2002.

[7] R. A. DePaolis, M. M. Vihman, and S. Kunnari, "Prosody in production at the onset of word use: A cross-linguistic study," *Journal of Phonetics*, vol. 36, pp. 406–422, 2008.

[8] M. Jessen, K. Marasek, K. Schneider, and K. Claßen, "Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German," in *Proceedings of the International Congress of Phonetic Sciences 13 (Stockholm)*, vol. 4, 1995, pp. 428–431.

[9] K. Schneider and B. Möbius, "Word stress correlates in spontaneous child-directed speech in German," in *Proceedings of Interspeech 2007 (Antwerpen)*, 2007, pp. 1394–1397.

[10] B. Lintfert, "Phonetic and phonological development of stress in German," Ph.D. dissertation, University of Stuttgart, 2009.

[11] G. K. Vallabha and B. Tuller, "Systematic errors in the formant analysis of steady–state vowels," *Speech Communication*, vol. 38, pp. 141–160, 2002.

[12] S. Lee, A. Potamianos, and N. Shrikanth, "Acoustics of children's speech: Developmental changes of temporal and spectral parameters," *Journal of the Acoustical Society of America*, vol. 105, no. 3, pp. 1455–1468, March 1999.

[13] L. Ménard, J.-L. Schwartz, L.-J. Boë, and J. Aubin, "Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model," *Journal of Phonetics*, vol. 35, pp. 1–19, 2007.

[14] K. N. Stevens and H. M. Hanson, "Classification of glottal vibration from acoustic measurements," in *Vocal Fold Physiology: Voice Quality Control*, O. Fujimura and M. Hirano, Eds. Singular, San Diego, 1994, pp. 147–170.

[15] M. Pützer and W. Wokurek, "Multiparametrische Stimmprofildifferenzierung zu männlichen und weiblichen Normalstimmen auf der Grundlage akustischer Analysen," *Laryngo–Rhino–Otologie*, vol. 85, pp. 1–8, 2006.

[16] C. Shih, B. Möbius, and B. Narasimhan, "Contextual effects on consonant voicing profiles: A cross–linguistic study," in *Proceedings of the 14th International Congress of Phonetic Science (San Francisco, CA)*, vol. 2, 1999, pp. 989–992.

[17] G. Dogil and B. Williams, "The phonetic manifestation of word stress," in *Word Prosodic Systems in the Languages of Europe*, H. van der Hulst, Ed. Berlin: De Gruyter, 1999, pp. 273–334.

[18] K. Claßen, G. Dogil, M. Jessen, K. Marasek, and W. Wokurek, "Stimmqualität und Wortbetonung im Deutschen," in *Linguistische Berichte*. Westdeutscher Verlag, 1998, vol. 174, pp. 202–245.

[19] G. J. Borden, K. S. Harris, and L. J. Raphael, *Speech science primer*. Baltimore: Williams & Wilkins, 1994.

[20] A. M. C. Sluijter, "Phonetic Correlates of Stress and Accent," Ph.D. dissertation, University of Leiden, 1995.

[21] B. Lintfert and W. Wokurek, "Voice quality dimensions of pitch accents," in *Proceedings of Interspeech/Eurospeech 2005 (Lisbon)*, 2005, pp. 2409–2412.

[22] J. B. Pierrehumbert, "Phonetic Diversity, Statistical Learning and Acquisition of Phonology," *Language and Speech*, vol. 46, no. 2–3, pp. 115–154, 2003.