BUCLD 37 proceedings
To be published in 2013 by Cascadilla Press
Rights forms signed by all authors

Capturing Developmental Patterns in Intonation Acquisition: A Target-oriented Parametric Approach

Britta Lintfert and Bernd Möbius

1. Introduction

Prosody links and organize semantic information, syntactic and morphological structure as well as segmental sequences into a consistent set of address frames like syllables, metrical feet, phonological word and intonational phrases (Levelt, 1989). With this prosodic frame infants learn to recognise the single segments of the speech stream during the acquisition process. The interaction between prosody and statistics in the segmentation of fluent speech suggests that prosody acts as a filter to find out possible word–like sequences in the speech stream (Shukla et al., 2007). The ability to segment fluent speech into words emerges during the first year of life, whereas the perception of phonetic representation of intonation is well developed with birth (Echols and Marti, 2004; Vihman et al., 2004). Therefore the prosodic development has been claimed to be the starting point for the production of speech as even in infants' cries different cry melodies and fundamental frequencies in the earliest communicative gestures have been noted (Gilbert and Robb, 1996; Wermke et al., 2002). Even pre-linguistic babbling depends on the input and has been claimed to show adult–like prosody (Siegel et al., 1990; Halle et al., 1991; Davis et al., 2000). But the production of intonation depends on the articulatory abilities of the child and remains difficult for infants and toddlers to produce in a stable manner (Snow and Balog, 2002; Oller et al., 2007).

In the field of prosody acquisition two different approaches are common to describe the development of intonation (Stoel-Gammon and Dunn, 1985): the *independent* and the *relational* approach. In an independent analysis of intonation (Cruttenden, 1997; Crystal, 1986) the child's productions are not compared to mature models. Intonation contours are described with reference to properties such as direction (i.e., falling or rising), accent range (i.e., amplitude of pitch change), and complexity (e.g., changes in direction measured in semitones). Based on these

[*] Affiliations of both authors: Saarland University, Computational Linguistics and Phonetics. We would like to thank Antje Schweitzer for writing the analysis scripts. This study is part of the project *Tonal and temporal development of prosodic categories of German speaking children* funded by the German Research Foundation (DFG, Grant MO 597/3-1)

measurement a descriptive illustration of the developing patterns is reported (Oller et al., 2007). Recent work also considers the relationship between intonation and emerging pragmatic skills (D'Odorico and Fasolo, 2007; Balog et al., 2009).

In contrast, in a relational analysis the child's production is compared to a mature model (i.e., the adult model). An increasing number of studies investigating the development of intonation apply the AM theory of intonation and the prosodic annotation system ToBI (Silverman et al., 1992; Hirschberg and Beckman, 1994; Pitrelli et al., 1994) to child speech (Prieto and Vanrell del Mar, 2007; Chen and Fikkert, 2007). ToBI approaches analyze intonation contours as sequences of (possibly categorical) intonation events, where each event can be decomposed into high and low pitch targets which are aligned with the syllable structure. Beyond the identification of pitch targets and their coarse alignment with the syllable structure, finer aspects of the phonetic realization of these events, such as amplitude of the pitch movements, or exact peak alignment within syllables, are not analyzed in the ToBI framework. However, the categories posited by ToBI or by its language-specific variants are developed for adult speakers. The problem in applying adult categories to child speech is the assumption that children with the beginning of meaningful speech are already capable of consistently using the different categories of intonation. Using these categories for child speech does not account for possible other categories during the acquisition of intonation based on children's limitations in production.

Against this background, to find categories of intonation even in pre-linguistic productions of child speech we have suggested a theory-neutral, automatic method for describing the shape of the F0 contour (Lintfert et al., 2010; Lintfert and Möbius, 2012). We proposed to parametrize F0 contours in the vicinity of accented syllables by PaIntE approximation (Möhler and Conkie, 1998). Groups of similar contours can then be identified by *K*-means clustering, reasoning that different clusters may be interpreted as different intonational categories. Results on adult data of child-directed speech showed a much better than chance correspondence between adult clusters and GToBI(S) categories (Lintfert et al., 2011). In this paper, we validate the idea of mapping clusters to ToBI categories on child speech at different developmental stages. We also compare the child clusters to adult target ToBI categories to show a developmental pattern of intonation contours and describe the variability of intonation contours. We extend the methodology described in (Lintfert et al., 2011) to speech produced by children aged between 1 and 8 years. We compare the intonation contours produced at different ages to the adult target form as well as we describe the variable production of intonation categories for each age. This method facilitates the description of a developmental pattern, evolving towards adult targets, as indicated by classification accuracies increasing with age.

Table 1: **Number of analyzed accents and subjects for each age group and mean age in months**

Age	Mean	Subjects									Accent
group	age (mo)	S1	S2	S3	S4	S5	S6	S7	S8	$\sum$	tokens
1;0	13.4	✓	✓	✓						3	354
2;0	20.0	✓	✓	✓	✓					4	1087
3;0	35.2	✓	✓	✓	✓	✓				5	1848
4;0	47.6				✓	✓	✓			3	2490
5;0	59.1					✓	✓			2	147
6;0	71.3						✓	✓		2	689
7;0	83.1						✓			1	290
8;0	93.7						✓		✓	2	539

2. Method

We analysed longitudinal data from six children over about three years of recordings per child. For two children (S7 and S8) we have only recordings over one year (see Table 1). The recordings are part of the Stuttgart Child Language Corpus (Lintfert, 2009) and took place at the children's homes in familiar play situations with their mothers while looking at picture books or playing with toys. Thus the data represent spontaneous speech productions. German child-directed (CDS) (4320 accented syllables) of the mothers of the children were also recorded and analyzed. All recordings were manually annotated on the segment, syllable and word level. The children's utterances were manually annotated with respect to perceived prominence. Syllables classified as prominent were then additionally coded according to GToBI(S) pitch accent categories. Note that this coding only served as a reference for comparing the children's production of F0 contours with adult targets. It does not imply an interpretation of child speech in terms of adult categories. The parametrization was performed for the prominent, and thus potentially accented, syllables only. GToBI(S) is an adaptation of ToBI to German and provides 5 basic types of pitch accents with different discourse interpretations: L*H, H*L, L*HL, HH*L, and H*M. These contours can also be described as rise, fall, rise-fall, early peak, and stylized contour, respectively.

Inter-observer reliability was assessed on 10% of the annotated data. Inter-observer agreement on the segmental and syllable levels was 94.5%, 88.3% on the word level, and 77.8% on the prosodic level.

2.1. PaIntE parametrization

PaIntE stands for "Parametrized Intonation Events" (Möhler and Conkie, 1998) and was originally developed for F0 modeling in speech synthesis. PaIntE approximates stretches of F0 by a phonetically motivated function which is the sum of a rising and a falling sigmoid with a fixed time delay (see Figure 1). The approximation window represents three syllables, where the accented syllable is indicated by the asterisk (σ*)). The parametrization uses six parameters, viz.

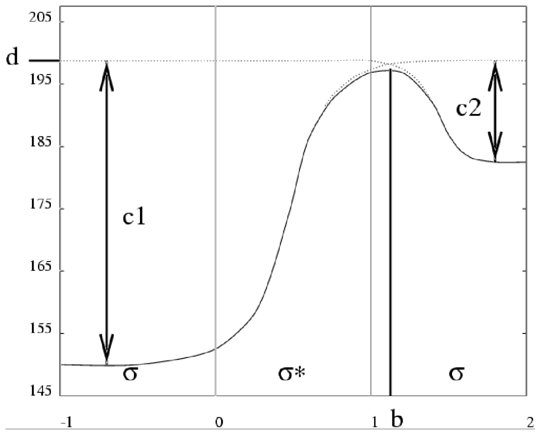

Figure 1: **Schematic of the PaIntE approximation function, reproduced from Möhler and Conkie (1998). Peak height is determined by parameter** *d***, amplitudes of rise and fall correspond to parameters** *c1* **and** *c2***, respectively, and peak alignment depends on the** *b* **parameter.**

the height of the F0 peak (parameter *d*), the temporal position of the peak in the syllable (*b*), and the amplitudes (*c1*, *c2*) and the steepness (*a1*, *a2*) (not shown) of the rising and falling sigmoid. The time axis is normalized to the lengths of the syllables, e.g., the peak is at the beginning of the accented syllable if *b*=0, and at its end if *b*=1. In contrast to other F0 parametrization or stylization approaches, PaIntE attempts to directly model properties of F0 contours that have been claimed to be linguistically meaningful (Schweitzer, 2011). For instance, parameters *c1* and *c2* are intended to capture the amplitude of the pitch movement. Parameter *b* quantifies the alignment of the peak with the syllable structure. Pitch movement excursion and peak alignment are tonal correlates of prominence and pitch accent type, respectively.

2.2. Normalization

The PaIntE parameters are based on different scales. *c1* and *c2* have a wide scale in Hertz and therefore a larger variance than parameter *b* which has a narrow range between -1 and 2.The z–transformation (Lobanov, 1971) eliminates these specific scaling effects by replacing absolute values by their difference to the class-specific mean and dividing this difference by the class-specific standard deviation of this parameter (Eq. 1).

$$z = \frac{x - mean(x)}{sd(x)} \tag{1}$$

2.3. Cluster analysis

K-means clustering is a hard clustering method which partitions the data into *k* clusters. The number of clusters *k* has to be specified beforehand. Each cluster is defined by its centroid: each observation belongs to the cluster with the nearest centroid. For the experiments presented here, we used R's (R Development Core Team, 2010) `kmeans` function, which by default implements the Hartigan-Wong method (Hartigan and Wong, 1979). We ran `kmeans` with 100 random starts, varying the number of clusters from 2 to 9, to cluster the data. We used all six PaIntE parameters as attributes, which were z-scored to eliminate speaker-specific and age-specific effects of pitch range and key and to match them with respect to scaling, which ensures that all parameters have approximately equal importance in clustering. The labeled accents were not used as attributes for clustering.

We used up to 9 clusters as we assume that each pitch-accent category can have more than one cluster depending on the alignment of the peak in the syllable (Lintfert et al., 2010).

For an independent analysis comparable to Balog and Snow (2007) we can derive properties such as range, direction, and complexity from the cluster centers but we can also comparing the clusters to the ToBI approaches. Following the relational approach, Schweitzer (2011) has shown that the PaIntE parameter distributions of the GToBI(S) accents indeed capture the defining properties of German pitch accents. Since the clustering identifies groups of similar contours, it stands to reason that the clusters might represent typical instances of specific GToBI(S) events. Mapping the clusters to GToBI categories in adult speech (Schweitzer and Möbius, 2009; Lintfert et al., 2011) we have shown that the clusters indeed represent intonational categories. Using the clustering for analyzing prosody acquisition it allows to find prototypical realizations in the child's productions independent of the established adult GToBI(S) categories, i.e., it allows to identify "categories" at each developmental stage, and to compare these "categories" to mature categories in a second step. To this end, each cluster was assigned the GToBI(S) accent which occurred most frequently in the cluster (*predicted category*). We than trained classifiers to predict GToBI categories with reasonable accuracy and evaluated for how many accents their manually annotated *true category* did indeed correspond to the category which had been assigned to their cluster (Lintfert et al., 2011). The percentage of accent tokens with matching predicted/true categories when clusters based on PaIntE are used for classification of new data can be interpreted as *classification accuracy*: the score indicating the percentage of correct decisions and correspondence correlating with accuracy.

The accuracy that can be reached if the accents are classified as belonging to the most frequent GToBI(S) category is indicated as *baseline accuracy*.

3. Results

3.1. Comparing PaIntE to independent approaches

According to Balog and Snow (2007), high maturity in falling accents is indicated if the accent range is greater than 4 semitones, in rising accents if the accent range is greater than 3 semitones. These properties can be directly derived from parameters *c1* and *c2*, using *d* as a reference. Second, accent direction can be compared to that in mature productions by comparing parameters *c1* and *c2*, i.e., by the relation between rise and fall amplitudes. An overall fall is given for *c2*>*c1* and an overall rise for *c1*>*c2*. Parameters *c1* and *c2* can also capture complexity to some extent. A simple heuristic could be that if both *c1* and *c2* exceed a certain threshold, say, 1 semitone, the contour can be characterized as rise-fall, indicating greater complexity than rise-only or fall-only contours (Lintfert et al., 2010).

Please note that for an independent analysis, the cluster analysis is not immediately necessary. Clustering groups similar realizations together. In each such group of similar realizations, the centroid can be interpreted as the “prototypical” realization. Thus, range, accent direction or complexity can either be regarded separately for each instance or they can be regarded for the centroids only, with each centroid representing a cluster of similar realizations.

3.2. Development of intonational categories

If the clustering serves to identify “categories” in intonation contours, then we would expect that there is a correspondence between the manually annotated GToBI(S) category and the clusters even for the earliest child data. To analyze whether the categories described by GToBI(S) and the produced categories of the children correspond or not, each cluster was assigned the GToBI(S) accent which occurred most frequently in the cluster determine the GToBI(S) category. We then evaluated for how many accents their manually annotated “true” category did indeed correspond to the category which had been assigned to their cluster. The classification accuracies (mean and standard deviation) as well as the baseline accuracies for all groups are calculated (Table 2)

At about one year, there is no difference between classification accuracy (64.8) and baseline (61.6). The results thus indicate not a better than chance correspondence between clusters and GToBI(S) categories. Nearly the same holds for the children at about two years with a classification accuracy of 61.0 and a similar high baseline accuracy of 58.4. The picture change at about three years of age. The difference between classification and baseline increases and, more importantly, the baseline decreases but remains at chance level until the age of about four. The baseline falls below chance level at about five years of age. Until the age of about four years, we do not found a good correspondence between the produced clusters and GToBI(S) categories as until this age the classification accuracies are worse than the baseline accuracy. The baseline accuracy is the accuracy that can be reached if one simply classified all accents as belonging to the most

Table 2: **Classification and baseline accuracies for the correspondence between the clusters and GToBI(S) categories.**

age	classification accuracy		baseline
	mean	*sd*	accuracy
1;0	64.8	1.9	61.6
2;0	61.0	1.8	58.4
3;0	59.5	3.6	50.8
4;0	64.9	1.7	54.6
5;0	66.9	3.8	42.6
6;0	65.6	1.8	46.0
7;0	59.2	4.9	42.6
8;0	64.6	1.1	40.5

frequent GToBI(S) category. To conclude, between one and 3 years of age no significant difference between classification accuracy and baseline can be observed, the correspondence between clusters and GToBI categories is poor (near chance-level). First increasing difference between classification accuracy and baseline can be observed at about 4 years. A correspondence between clusters and GToBI categories well above chance level can be observed for children between 5 and 8 years, but still with some instability of realizations. The data obtained for children older than four years indeed showed a much better than chance correspondence between clusters and GToBI(S) categories. Altogether a spread between classification accuracy and baseline with increasing age can be observed The results give evidence that children younger than 5 years are not capable of consistently using the different intonation categories as they are not able to produced the intonation contours in a stable manner.

Based on these results we conclude that from about 5 years of age onward the children produced intonation categories that can be annotated consistently with GToBI(S) categories. This does not appear to be well motivated at an earlier age: we found a great variability in the production of intonational contours before the age of five probably constrained by the motor abilities of the developing articulatory system. This variability mainly depends on the pitch alignment in the syllable and the steepness of the rise and fall (Lintfert et al., 2010). An adequate description of the intonation contours produced before the age of five is difficult using only GToBI(S) categories.

3.3. Mapping adult targets and child categories

To assess whether the cluster-to-category assignment is indeed similar for the child data comparable to the adult GToBI(S) targets, we applied the clustering obtained on child-directed speech data (Lintfert and Möbius, 2012) to the child data.

Table 3: **Classification and baseline accuracies for mapping adult targets and children categories.**

age	classification accuracy		baseline
	mean	*sd*	accuracy
1;0	55.9	3.4	61.8
2;0	51.6	2.7	58.4
3;0	67.4	2.9	63.4
4;0	59.7	2.7	54.7
5;0	64.5	3.5	42.9
6;0	62.6	5.3	64.1
7;0	60.2	4.3	42.8
8;0	61.2	3.9	40.5

For each age we assigned each datapoint of the child data to the nearest cluster center obtained on CDS data. We then evaluated in how many cases the category assigned to the cluster based on CDS data matched the manually annotated "true" GToBI(S) category.

Between one and two years of age the classification accuracy is even worse than the baseline accuracy, i.e., the accuracy that can be reached if one simply classifies all accents as belonging to the most frequently produced GToBI(S) category (Table 3). The categories produced by the children between the age of 3 and 4 years are nearly the same as the baseline accuracy and also do not correspond well with the categories produced in child-directed speech. Moreover, the accuracies reached for mapped clusters are worse than those without mapping of adult target categories. These results indicate that with the beginning of speech-like productions children are not yet capable of consistently using the categories as posited by intonation theory. At the age of about 4 years this picture changes, as indicated by the increase of classification accuracy based on the mapped clusters and the decrease of baseline accuracy (Figure 2). The correspondence between the produced clusters and the GToBI categories produced by adults speakers are well above chance level. A clear tendency towards adult-like intonation categories can be observed at these ages. Based on these results we conclude that from about 5 years of age the children produced categories comparable to those in adult intonation.

4. Discussion

This study aimed to verify that our proposed methodology, viz. the parametrization of F0 contours in combination with a clustering technique, is suitable for identifying intonational "categories". The results of Section 3.2 showed that until the age of about 5 years, there is almost no difference between the clas-

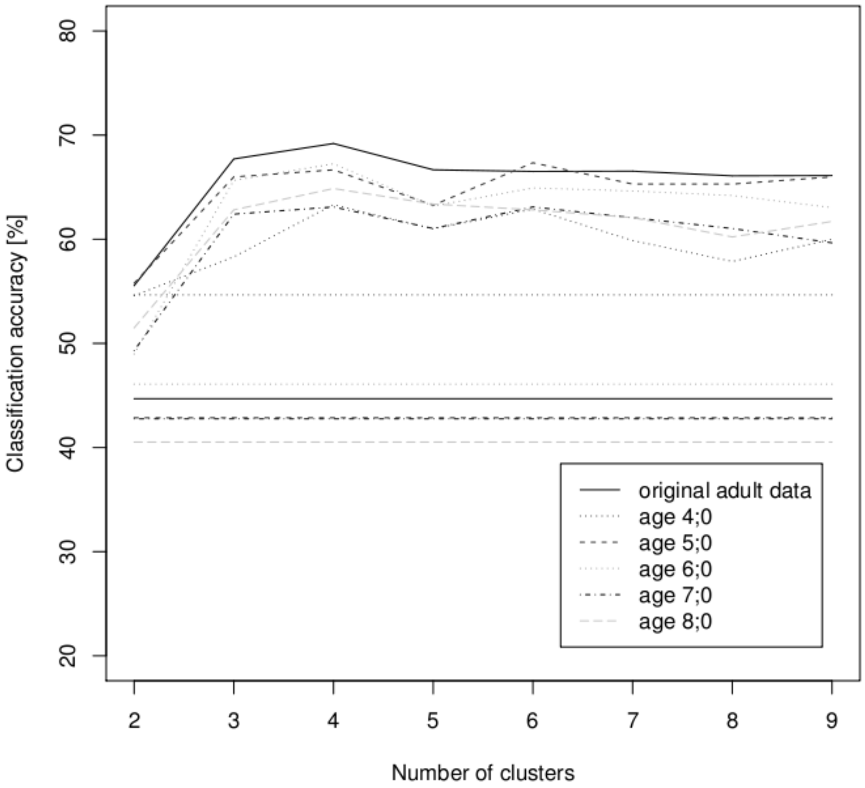

Figure 2: **Classification accuracies and baselines for 2 to 9 clusters, based on child-directed clustering data (solid line) and child data between 4 and 8 years of age. Note that the baselines for 5 and 7 years are similar.**

sification accuracy and the baseline. Our interpretation is that the categories produced by younger children cannot be described adequately by means of GToBI(S) categories. With increasing age the accuracies increase, too, and at about 5 years of age children tend to produce categories similar to those assumed by GToBI(S), resulting in a lower baseline and higher classification accuracies. In Section 3.3 the categories produced by the children were mapped onto adult targets to identify possible GToBI(S) categories. As expected from the results, until the age of five the produced categories do not correspond well to the underlying adult targets. Children aged older than 5 years are able to produce categories similar to those of adult speakers, corresponding well with GToBI(S) categories. For younger children, however, GToBI(S) categories cannot describe the produced intonation categories adequately. We have shown that intonational categories produced by younger children (less than 4 years) cannot be adequately described by means of (adult) GToBI categories. A development toward adult-like targets can be observed from age 5 to 8, evidenced by an increasing spread between classification accuracies and baselines. Comparing the produced cluster with GToBI categories an increased (and at age 7-8) good correspondence can be observed, but the between- and within-subject variability remains high even until the age of eight.

5. Conclusion

A theory-neutral, automatic methodology for detecting intonational categories was described. The main intention for developing this method was the problem given in the field of prosody acquisition with two different approaches to describe the development of intonation. Using the independent approach no mature target is used for comparison and therefore only a description of the developing patterns can be given. In the ToBI framework adult categories are applied to children's productions, which does not account for possible other categories in the timecourse of intonation acquisition based on children's limitations in production. Therefore we developed a method based on F0 parametrization. The advantage of the proposed method is that by using the PaIntE parametrization, we can capture fine phonetic detail in children's realizations of intonation contours. Using clustering methods to further analyze the data, we can assess the variability of the children's production of intonation contours and classify the maturity of the contours depending on the children's age. We can describe the variability in peak alignment as well as amplitude of rises and falls in the children's production of intonation contours and can evaluate which categories children produce and how they differ from adult categories. This method can be applied to all stages of L1 intonation acquisition but also to adult speech. This is favorable for longitudinal studies of intonation in child speech, as we can apply the same method over the course of the study even as children go through different developmental stages from pre-linguistic utterances to multi-word utterances. An extension of the method by including temporal parameters is envisaged in future work.

The proposed method is an automatic approach and can give comparable results independent of the theoretic analysis framework. The clustering method can also include additional variables, such as those related to discourse structure. Because the method is also language-independent, it facilitates cross-language intonation studies too.

References

Balog, Heather L., Felicia D. Roberts, and David Snow (2009): Discourse and Intonation Development in the First–Word Period. Enfance 3, 293–304.

Balog, Heather L. and David Snow (2007): The adaption and application of relational and independent analyses for intonation production in young children. Journal of Phonetics 35, 118–133.

Chen, Aoju and Paula Fikkert (2007): Intonation of early two–word utterances in Dutch. In: Proceedings of 16th International Congress of Phonetic Science, 315–320.

Cruttenden, Alan (1997): Intonation. Cambridge, MA: Cambridge University Press.

Crystal, David (1986): Prosodic development. In: Fletcher, Paul and Michael Garman (eds.), Language Acquisition, Cambridge University Press, 174–198.

Davis, Barbara L., Peter F. MacNeilage, Christine Matyear, and Julia Powell (2000): Prosodic Correlates of Stress in Babbling: An Acoustical Study. Child Development 71(5), 1258–1270.

D'Odorico, Laura and Mirco Fasolo (2007): The prosody of early multi–word speech: word order and its intonational realization in the speech production of Italian children. In: Proceedings of the 16th International Congress of Phonetic Sciences, 321–325.

Echols, Catharine H. and Nathan C. Marti (2004): The Identification of Words and Their Meanings: From Perceptual Biases to Language–Specific Cues. In: Hall, D. G. and S. R. Waxman (eds.), Weaving a lexicon, Cambridge, MA: MIT Press, 41–78.

Gilbert, Harvey R. and Michael P. Robb (1996): Vocal Fundamental Frequency Characteristics of Infant hunger Cries: Birth to 12 Months. International Journal of Pediactric Otorhinolaryngol 34, 237–243.

Halle, Pierre A., Bénédict De Boysson-Bardies, and Marilyn M. Vihman (1991): Beginnings of Prosodic Organization: Intonation and Duration patterns of Disyllables produced by Japanese and French Infants. Language and Speech 34(4), 299–318.

Hartigan, J. A. and M. A. Wong (1979): A K-means clustering algorithm. Applied Statistics 28, 100–108.

Hirschberg, Julia and Mary E. Beckman (1994): The ToBI annotation conventions.

Levelt, Willem J. M. (1989): Speaking: From Intonation to Articulation. MIT Press, Cambridge, MA.

Lintfert, Britta (2009): Phonetic and phonological development of stress in German. Doctoral dissertation, Universität Stuttgart.

Lintfert, Britta and Bernd Möbius (2012): Describing the development of intonational categories using a target-oriented parametric approach. In: Proceedings of Interspeech 2012.

Lintfert, Britta, Antje Schweitzer, and Bernd Möbius (2011): A parametric approach to intonation acquisition research: validation on child-directed speech data. In: Proceedings of Interspeech 2011, 757–760.

Lintfert, Britta, Antje Schweitzer, Lukasz Wolski, and Bernd Möbius (2010): Quantifiying developmental changes of prosodic categories. In: Speech Prosody 2010 100929, 1–4.

Lobanov, Boris M. (1971): Classification of Russian Vowels Spoken by Different Speakers. Journal of the Acoustical Society of America , 606–607.

Möhler, Gregor and Alistair Conkie (1998): Parametric modelling of intonation using vector quantization. In: Proceedings of 3rd ESCA Workshop on Speech Synthesis, 311–316.

Oller, D. Kimbrough, Suneeti Nathani Iyer, Eugene H. Buder, Kyounghwa Kwon, Lesya Chorna, and Kelly Conway (2007): Diversity and contrastivity in prosodic and syllabic development. In: Proceedings of the International Congress of Phonetic Sciences, 303–308.

Pitrelli, John, Mary Beckman, and Julia Hirschberg (1994): Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP, Yokohama), 123–126.

Prieto, Pilar and Maria Vanrell del Mar (2007): Early intonational development in Catalan. In: Proceedings of the 16th International Congress of Phonetic Sciences, 309–314.

R Development Core Team (2010): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Schweitzer, Antje (2011): Production and Perception of Prosodic Events–Evidence from Corpus-based Experiments. Doctoral dissertation, Universität Stuttgart.

Schweitzer, Antje and Bernd Möbius (2009): Experiments on Automatic Prosodic Labeling. In: Proceedings of Interspeech 2009, 2515–2518.

Shukla, Mohinish, Marina Nespor, and Jacques Mehler (2007): An interaction between prosody and statistics in the segmentation of fluent speech. Cognitiv Psychology 54(1), 1–32.

Siegel, Gerald M., Martha Cooper, James L. Morgan, and Robin Brenneise-Sarshad (1990): Imitation of intonation by infants. Journal of Speech and Hearing Research 33, 9–15.

Silverman, Kim, Mary Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg (1992): ToBI: A standard for labelling English prosody. In: Proceedings of the International Conference on Spoken Language Processing, 867–870.

Snow, David and Heather L. Balog (2002): Do children produce the melody before the words? A review of developmental intonation research. Lingua 112, 1025–1058.

Stoel-Gammon, Carol and Carla Dunn (1985): Normal and Disordered Phonology in Children. Baltimore: University Park Press.

Vihman, Marilyn M., Satsuki Nakai, Rory A. DePaolis, and Pierre Halled (2004): The role of accentual pattern in early lexical representation. Journal of Memory and Language 50, 336–353.

Wermke, K., W. Mende, C. Manfredi, and P. Bruscaglioni (2002): Developmental aspects of infant's cry melody and formants. Medical Engineering & Physics 24, 501–514.