

Polynomial Modeling of Child and Adult Intonation in German Spontaneous Speech

Language and Speech

54(2) 199–223

© The Author(s) 2011

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0023830910397495

las.sagepub.com

**Laura E. de Ruiter**

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Abstract

In a data set of 291 spontaneous utterances from German 5-year-olds, 7-year-olds and adults, nuclear pitch contours were labeled manually using the GToBI annotation system. Ten different contour types were identified. The fundamental frequency (F_0) of these contours was modeled using third-order orthogonal polynomials, following an approach similar to the one Grabe, Kochanski, and Coleman (2007) used for English. Statistical analyses showed that all but one contour pair differed significantly from each other in at least one of the four coefficients. This demonstrates that polynomial modeling can provide quantitative empirical support for phonological labels in unscripted speech, and for languages other than English. Furthermore, polynomial expressions can be used to derive the alignment of tonal targets relative to the syllable structure, making polynomial modeling more accessible to the phonological research community. Finally, within-contour comparisons of the three age groups showed that for children, the magnitude of the higher coefficients is lower, suggesting that they are not yet able to modulate their pitch as fast as adults.

Keywords

children, German, intonation, polynomial modeling, tonal alignment

Introduction

One of the tasks of intonation research is finding out which types of accents are used in a particular linguistic context, such as questions or different focus structures. The conventional way to go about this is to manually assign descriptive labels to the speech data that were produced in these contexts. Currently these labels are typically drawn from autosegmental-metrical (Ladd, 1996) descriptions of intonational grammars (e.g., Pierrehumbert & Hirschberg, 1990), in which intonation patterns are described in terms of sequences of high (H) and low (L) tones. It is assumed that for a given language intonation patterns can be described using a limited set of labels. Information on how these labels should be used is provided in labeling guidelines (e.g., Beckman & Ayers, 1997, for

Corresponding author:

Laura de Ruiter, Bielefeld University, – Rektorat – , P.O. Box 10 01 31, 33501 Bielefeld, Germany

Email: laura.deruiter@uni-bielefeld.de

English; Benzmüller & Grice, 1997, for German). The correct application of the labels is sometimes checked by having multiple labelers annotate the same data and comparing their consistency (e.g., Grice, Reyelt, Benzmüller, Mayer, & Batliner, 1996; Pitrelli, Beckman, & Hirschberg, 1994). Results show that labelers' agreements are above chance, indicating that there is a non-accidental relationship between the speech signal and the labels.

However, these studies are very labor-intensive, and they do not always yield the level of consistency that would be desirable. This sometimes raises questions regarding the validity of the labels. Many researchers feel that it would be helpful if the phonological categories could also be validated instrumentally (Ladd, 2008, p. 12). In the case of GToBI (German Tones and Break Indices; Grice & Baumann, 2002) accent categories, there should, for example, be a measurable F0 minimum within the accented syllable of a L*+H accent. Of course, labelers use these criteria implicitly in their judgments, but the parameters in question (e.g., the position of turning points) may as well be measured directly to have more objective evidence for the correct application of the labeling criteria. However, acoustic landmarks like turning points have to be annotated as well, and the labeling process is even more time-consuming. A complicating factor is that the determination of these landmarks is not always without problems. For example, certain segments (e.g., fricatives such as /v/) cause so-called microprosodic variation in the pitch curve, making it sometimes difficult to identify the location of turning points (such as the F0 minimum in L*+H accents) unambiguously. When in doubt, labelers who have to decide on the location of a turning point may (without being aware of it) be inclined to select a location that is favorable to the prosodic label they have assigned, which potentially adds a certain amount of subjectivity. In view of these difficulties, it would be desirable to have an additional, more objective way of obtaining quantitative empirical evidence for intonation labels.

Such empirical acoustic evidence for intonation labels in English was presented by Grabe, Kochanski, and Coleman (2007). They modeled the fundamental frequency of hand-labeled accents mathematically using polynomial equations, and showed that the accents that were assigned different labels were also significantly different from each other in their mathematical descriptions. Polynomial equations are a way to describe curves, or rather continuous functions (of which F0 is one as well), in a mathematical expression constructed from variables and constants (e.g., $3x^2 + 4x + 5$). They are one of a number of approaches used in speech technology to fit F0 curves. Other curve-fitting models include the Fujisaki model (Fujisaki, 1992), MOMEL (Hirst, Di Cristo, & Espresser, 1993) or Tilt (Taylor, 2000).

The corpus used in the Grabe et al. (2007) study consisted of 714 read-out sentences, produced by 42 speakers. The nuclear accents of these sentences (i.e., the final pitch accent in the phrase and the subsequent boundary tone) were manually annotated according to the IViE (Intonational Variation in English) labeling system (Grabe, 2002), which is an autosegmental-metrical transcription system based on ToBI (Tones and Break Index; Beckman & Ayers, 1997), but developed to allow dialect-independent transcription of English intonation (for details, see also Grabe, 2004). Seven different nuclear accent types were found in the corpus and modeled with orthogonal Legendre polynomials (details are given below), resulting in a mathematical description of each accent. Statistical analyses showed that parameters of the polynomial descriptions of six of the seven accent types that had been identified by the labelers differed significantly from each other. The authors conclude that polynomial modeling can provide intonational phonologists with a tool to empirically validate linguistic descriptions of intonation (Grabe et al., 2007, p. 299).

Against this background, the present study sets out to investigate whether polynomial modeling can be applied to another language, in this case German, with similar results. In contrast to Grabe et al.'s corpus, the speech material analyzed here consists of natural rather than scripted speech.

What is more, the speakers in this study were drawn from three different age groups: five-year-olds, seven-year-olds and adults, adding more variability to the data in terms of speaking rate and average pitch. Thus the data provide a real test case for the usability of the polynomial approach.

I also present an extension to the polynomial model that shows how information about the relative alignment of tonal targets with the segmental string can be derived from the modeled curves, connecting this new approach directly with the research on tonal alignment (e.g., Arvaniti, Ladd, & Mennen, 1998; Atterer & Ladd, 2004). Finally, I explore how polynomial modeling can provide a window into intonational development.

2 Method

2.1 Data and annotation

The data set consists of 291 spontaneous utterances taken from a corpus of narrations that were elicited by means of a picture-based story telling task. The data had originally been collected in the course of a different study investigating how German speakers of different ages mark the information status of discourse referents in narrative discourse (more information on the study can be found in De Ruiter, 2010 and Herbst, 2007). The speakers were 29 five- and 26 seven-year-old children, and 31 adults, native speakers of German from North Germany. The children were recorded individually in quiet rooms at their respective kindergartens and primary schools, the adults were recorded in an anechoic chamber in the laboratory of the Linguistics Department of the University of Potsdam. The recordings were made using a condenser microphone and a DAT recorder, and the signals were digitized directly into the computer at a sampling rate of 44 kHz (16 bit format). The picture stories featured four referents with disyllabic names that have a mainly sonorant segmental make up to facilitate pitch tracking. Three of the words had a trochaic stress pattern – *Möwe* (/ˈmø:və/, ‘seagull’), *Biber* (/ˈbi:bɐ/, ‘beaver’), *Biene* (/ˈbi:nə/, ‘bee’) – one had an iambic stress pattern – *Kamel* (/kaˈme:l/, ‘camel’). The utterances that were selected for analysis were sentences in which these words occurred in final position. Two examples are provided in (1).

- (1) a. Er malt eine Biene.
‘He draws a bee.’
b. Sie kommt wieder zur Möwe.
‘She comes back to the seagull.’

The utterances were annotated and analyzed using the speech-analysis tool Praat (version 5.0.35, Boersma & Weenink, 1992–2008). The intonational phrases (IP) containing the target words were segmented at the level of the syllable. Following this, intonation of the target words and of the subsequent phrase boundary was labeled following GToBI guidelines (Benzmüller & Grice, 1997; Grice & Baumann, 2002; Grice, Baumann, & Benzmüller, 2005) based on a combination of auditory analysis and visual inspection of the F0 tracks. The labeling was first done by the author (a native speaker of German), and subsequently extensively discussed with a second researcher (also a German native speaker) experienced in GToBI annotation.¹

Six different nuclear pitch accent types and four different boundary tones were identified in the data. In addition, a substantial number of items were deaccented. Not all possible combinations of pitch accents and boundary tones occurred, and some combinations occurred only very rarely. In order to arrive at a reasonable number of items per category (i.e., items per pitch accent + boundary tone combination), the four different boundary tones were pooled into two major groups, high and low boundary, using the simplified labels H% and L%. The boundary tones H-% (high boundary)

and H-^H% (upstepped boundary) were subsumed under H%.² The boundary tones L-% (low boundary) and !H-% (downstepped boundary) were subsumed under the label L%.³ Following Grabe et al. (2007), accent labels (e.g., H*) and boundary tone labels (e.g., L%) are separated by commas in the transcription. Table 1 displays the pitch accent + boundary tone combinations that occurred at least ten times. For brevity, these combinations will be referred to as nuclear contours. The first column of Table 1 gives the autosegmental-metrical labels, the second column provides a stylized representation of the typical F0 shape, and the third column gives the frequency of the contour in the data set. Note that the category of the downstepped contours !H*,L% was not collapsed with its non-downstepped counterpart (which was done by Grabe et al., 2007). With ten different nuclear contours, the data set contains a larger number of contours than was modeled in the Grabe et al. study, where seven different nuclear contours were analyzed. On the one hand, this is due to the fact that I also included deaccented items in the data, and on the other to the fact that the GToBI system also contains right-headed accents (e.g., L+H*), which do not exist in the IViE transcription system, where an accent like L+H* would simply be described as H*.

2.1.1 Polynomial modeling. In this section, first the polynomial approach by Grabe et al. (2007) is described. I will then give details of the way polynomial modeling has been implemented in this study, and point out where the method differs from the approach taken by Grabe et al.

In their study, Grabe et al. analyzed the region of F0 beginning 100 milliseconds before the nuclear accent of each sentence in the corpus, and extending to the end of the voiced part of the IP. All utterances in their corpus were designed to be fully voiced, but some voiced fricatives tended to be devoiced phrase-finally, so that some utterances contained unvoiced material. These parts were discarded from the analysis. In addition to measuring F0, Grabe et al. also extracted measures of loudness and aperiodicity from the signal (details can be found in Kochanski, Grabe, Coleman, & Rosner, 2005). In the course of fitting the polynomial model to the data, these measures were later used to give more weight to loud and sonorant regions, such as syllable centers, assuming that in these regions, F0 measures are more reliable and F0 movements may be more perceptually relevant (Grabe et al., 2007, p. 287). Before fitting, all F0 values were normalized by dividing them by the speaker's mean, and subtracting 1. Thus, a normalized F0 of zero corresponded to the speaker's average F0. Furthermore, Grabe et al. shifted and scaled (i.e., compressed or extended) the time axis so that the nuclear contour region spanned values between -1 and 1, which is a prerequisite for Legendre polynomials. The F0 data in the analysis region of each sentence were then modeled as a best-fit sum of Legendre polynomials (the exact details of the procedure are given in Kochanski et al., 2005). Legendre polynomials belong to the class of orthogonal polynomials. As argued by the authors, orthogonal polynomials have the advantage of minimizing the correlations among the coefficients, which would otherwise have to be taken into account in the statistical analysis. In contrast to other orthogonal polynomials, Legendre polynomials ensure that the coefficients are equally sensitive throughout the utterance (Kochanski et al., 2005). After fitting, each nuclear contour is described by a model. This model M is specified by a set of coefficients, c_i , that multiply the different Legendre polynomials before they are added together:

$$M(x) = c_0 + c_1 \cdot x + c_2 \cdot \left(\frac{1}{2}(3x^2 - 1) \right) + c_3 \cdot \left(\frac{1}{2}(5x^3 - 3x) \right) \quad (1)$$

The polynomial analysis bears similarities to a Fourier analysis in that the lower-ranking polynomials identify the more slowly varying properties of the curve, whereas the higher-ranking polynomials

Table 1. Autosegmental-metrical contour labels, F0 stylizations and frequency of occurrence (*N*) of the contours observed in the data set. The grey-shaded area indicates the lexically stressed syllable, the white areas preceding and following it represent the pre- and post-accentual syllable, respectively. Most of the stylized shapes are adapted from Grice and Baumann (2002, p. 279ff.). Following a notation used by Baumann (2006), the symbol “∅” is used to indicate lack of pitch accent, but note that it is not part of the GToBI annotation scheme

Contour label	Stylized F0	<i>N</i>
H*,L%		40
L+H*,L%		24
!H*,L%		16
H+!H*,L%		34
H+L*,L%		10
H*,H%		19
L+H*,H%		23
L*,H%		74
∅,H%		16
∅,L%		29
Other		6
TOTAL		291

identify the more rapidly changing properties (Grabe et al., 2007, p. 289). The more complex a curve is, the more polynomials are needed to fully describe it. As noted by Grabe et al., the first four polynomials ($c_0 - c_3$, displayed in Figure 1) have straightforward physical interpretations:

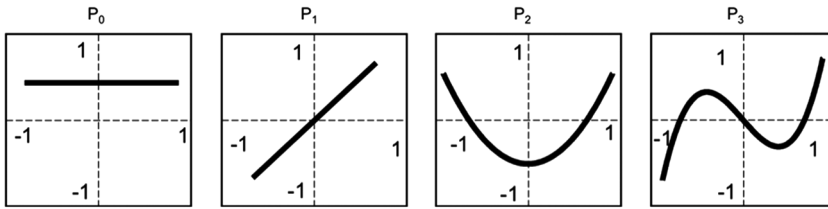


Figure 1. The first four Legendre polynomials P_0 – P_3 . Following the naming convention introduced by Grabe et al. (2007), the first coefficient (c_0) will be referred to as AVERAGE, the second coefficient (c_1) will be referred to as SLOPE, the third coefficient (c_2) as PARABOLA, and the fourth coefficient (c_3) as WAVE

- The first coefficient, c_0 , corresponds to the average F0 of the accent after pitch normalization.
- The second coefficient, c_1 , gives an indication of the overall slope (falling or rising) of the accent.
- The third coefficient, c_2 , specifies the overall curvature of the accent, which can be a broad dip or a rise in the middle of the accent.
- The fourth coefficient, c_3 , corresponds to a wave-like shape.

Having outlined the original approach taken by Grabe et al., the following paragraphs now describe in detail how the nuclear contours in the present data set were modeled.

Prior to the analysis, the F0 tracks were inspected for errors such as octave jumps and manually corrected where necessary. The analyzed domain consisted of the (voiced regions of the) lexically stressed syllable, the pre-stressed syllable, and the post-stressed syllable (see below for details on how the voiced region was determined). This is different from Grabe et al.'s domain (see above). There were two reasons to define the region differently: first, Grabe et al.'s criterion is based on adult speech, while the present data set also comprises child speech. Young children typically have a lower speech rate, which increases with age until at least age eleven (e.g., Boutsen & Hood, 1996; Sturm & Seery, 2007). This means that for the young speakers, a starting point of 100 milliseconds preceding the centre of the stressed syllable would still be way into the stressed syllable. As a consequence, the analysis region would not even cover the entirety of the stressed syllable, which is clearly undesirable. Second, unlike the IViE transcription system, the GToBI scheme also contains right-headed accents (e.g., L+H*), which means that the pitch movement on the pre-stressed syllable is considered to be an important part of the overall accent shape for some accent types. Any criterion of having the beginning of the analysis region start at some arbitrary distance from the center of the stressed syllable (e.g., 200 milliseconds) is in danger of “cutting off” parts of the pre-stressed syllable even with adult speakers. The modeled F0 curve would consequently not be a truthful representation of the F0 curve that the intonational labeling was based on, and could therefore not be used to find empirical support for the accent labels. Determining the analysis region on the basis of the syllable structure avoids these problems. An illustration of the analysis region is provided in Figure 2. Within the analysis region, F0 was measured in steps of 5 milliseconds. At each point at which F0 was measured, intensity and harmonics-to-noise ratio⁴ (Boersma, 1993) were also extracted. These two measures were later combined in a weighting parameter used in the fitting algorithm described below.

Before fitting the data, F0 and time were normalized. For F0 normalization, all F0 values were divided by the speaker's mean (which was calculated by taking the average of all first unstressed

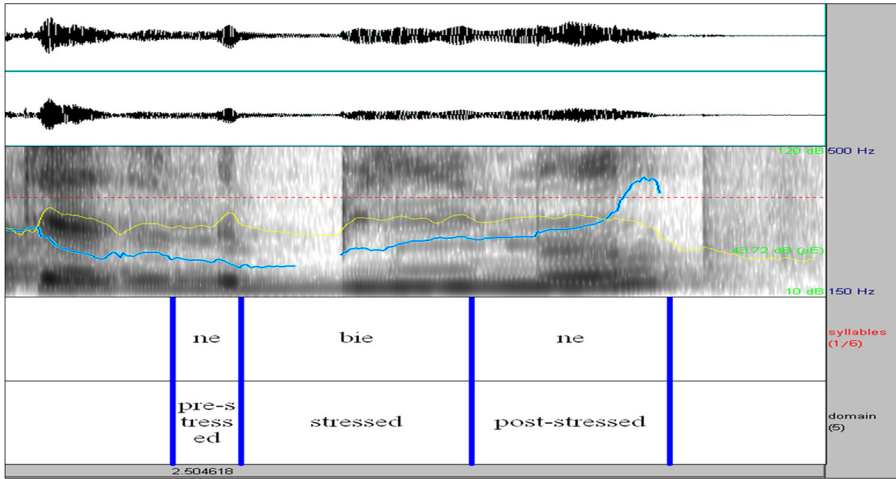


Figure 2. Amplitude envelope, spectrogram, intensity (thin light line) and F0 (thick black line) curve of an utterance by a five-year-old, *Guck, der malt ne Biene* ‘Look, he is drawing a bee’. The domain in which F0 is measured consists of the pre-stressed syllable (“ne”), the lexically stressed syllable (“bie”) and the post-stressed syllable (“ne”)

syllables in all utterances made by one speaker), and subtracting 1 from it. Thus, a value of 0.1 corresponds to an F0 that is 10% above the speaker’s mean. For time normalization, the time axis of the analysis domain was shifted and scaled (i.e., compressed or extended) to values between -1 and 1 .⁵

Like in Grabe et al., the model was specified by a set of four coefficients, c_i , that multiply the different Legendre polynomials before they are added together (see Equation 1). For the estimation of the coefficients of the Legendre polynomials that best describe a given intonation contour, I used Polyfit (De Ruiter, 2008), a customized computer program written in C++. The program reads in normalized F0 values and a weighting parameter (described below) and calculates those Legendre coefficients that minimize the difference between the predicted polynomial and the original pitch contour as estimated by Praat’s pitch tracking algorithm.⁶ The weighting parameter w is used to de-emphasize regions that are not important for the overall shape of the contour (such as micro-perturbations that are due to the segmental structure), and to give more weight to perceptually important loud and sonorant regions such as vowel centers (Kochanski et al., 2005, p. 1043). A higher w for a certain time window forces the algorithm to model F0 values in this region with more precision. The weighting parameter used in this study is a combination of the intensity and the HNR of the signal. Intensity was normalized by dividing each value by the mean intensity of the voiced parts of the entire utterance. Unlike intensity, HNR values usually cover a wider range of values and can also be negative, in cases where there is more noise than harmonics in the signal, which is the case for instance in voiceless regions. I normalized the HNR measures using a sigmoid function, which transforms all possible values (from ∞ to $-\infty$) into values between 0 and 1. Hence negative HNR values (where there is a lot of noise in the signal) receive a low score near 0, whereas positive ones receive a score closer to 1. In Equation 2 below, H refers to the normalized HNR value (H -score) that is computed from h , the original HNR value as measured by Praat. The constant e is Euler’s number, the base of the natural logarithm. The coefficient k determines the steepness of the S-curve produced by the function. I calculated the coefficient k using the criterion

that a HNR value of 15 dB (roughly equal to 97% energy from the harmonic part, see Praat manual on “harmonicity”) receives an H -score of 0.75. The resulting coefficient k is 0.02453.

$$H(h) = \frac{1}{1 + e^{-kh}} \quad (2)$$

The weighting parameter w was then the product of normalized intensity (I) and the standardized HNR value (H):

$$w = I \cdot H \quad (3)$$

As described above, the analysis region was a three-syllable domain around the lexically stressed syllable. However, unvoiced regions (like devoiced vowels) at the beginning or the end of the domain can be problematic. When the program determines the coefficients to model the intonation contour, it mainly fits the polynomials to the voiced parts while the polynomials can take any form for unvoiced regions. This is not harmful for voiceless regions in the middle of a voiced region, if we assume that the F0 contour constitutes a smooth function; the algorithm interpolates between the voiced regions. However, for voiceless regions before pitch onset or after pitch offset, the fitting becomes unpredictable. Note that weighting alone cannot solve this problem, as a very low w would still “allow” the program to fit almost any curve. To avoid this problem, I set the domain to start at the first voiced frame within the original three-syllable domain, and to end at the last one. Following this approach, the analysis domain was adjusted for 211 out of the 291 nuclear contours. For these contours, the domain was on average shortened by 10% of the overall duration of the three-syllable domain.

3 Results

To give an impression of how the modeled curves compare to the empirical F0, Figure 3 shows the original F0 curve and the modeled curve of two nuclear contours superimposed onto each other. One can clearly see that the microprosodic effects such as the pitch depression in the voiced obstruent [b] of the word *Biene* in the left panel have not been modeled.

Before presenting the results of the statistical analysis, I will first show the average coefficient values for each accent, which make the results of the accent pair contrasts easier to interpret. These profiles are shown in Figure 4. If c_0 is negative, this indicates that the mean F0 for this accent is low, while a positive c_0 indicates a mean F0 that is higher compared to the average F0 of the speaker. A negative c_1 is a sign of a predominantly falling accent, whereas a positive second coefficient means that the slope is mainly rising. If c_2 is negative, the F0 curve is convex (dome-shaped), if it is positive, the curve is more concave (cup-shaped). Finally, c_3 describes whether the overall shape is falling-rising-falling (if it is negative), or rising-falling-rising (if it is positive).

The statistical analyses were performed using R version 2.6.2 (R Development Core Team, 2008). Instead of a Multivariate Analysis of Variance (MANOVA), which was used by Grabe et al. (2007), I tested for differences between the accents using another statistical technique, linear-mixed effect (LME) models in R (Baayen, Davidson, & Bates, 2008; Bates & Sarkar, 2007; for applications see e.g., Kuperman, Bertram, & Baayen, 2008; Plag & Baayen, 2009; Staub, Grant, Clifton, & Rayner, 2009). These models have a number of advantages over traditional analyses like

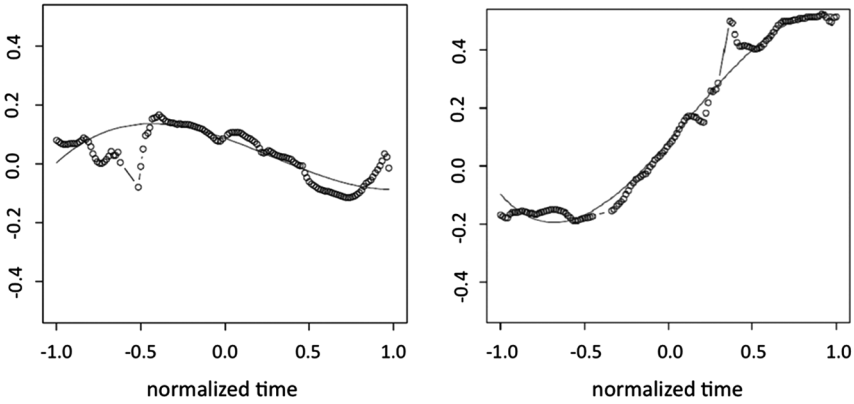


Figure 3. Examples of the (normalized) original F0 track (circles) and the curve modeled by Polyfit (continuous line). The x-axis plots the normalized time (from -1 to 1), the y-axis shows the units of the normalized F0. The panel on the left is an example of an accent labeled H*,L%, the panel on the right is an example of a contour labeled L*,H%.

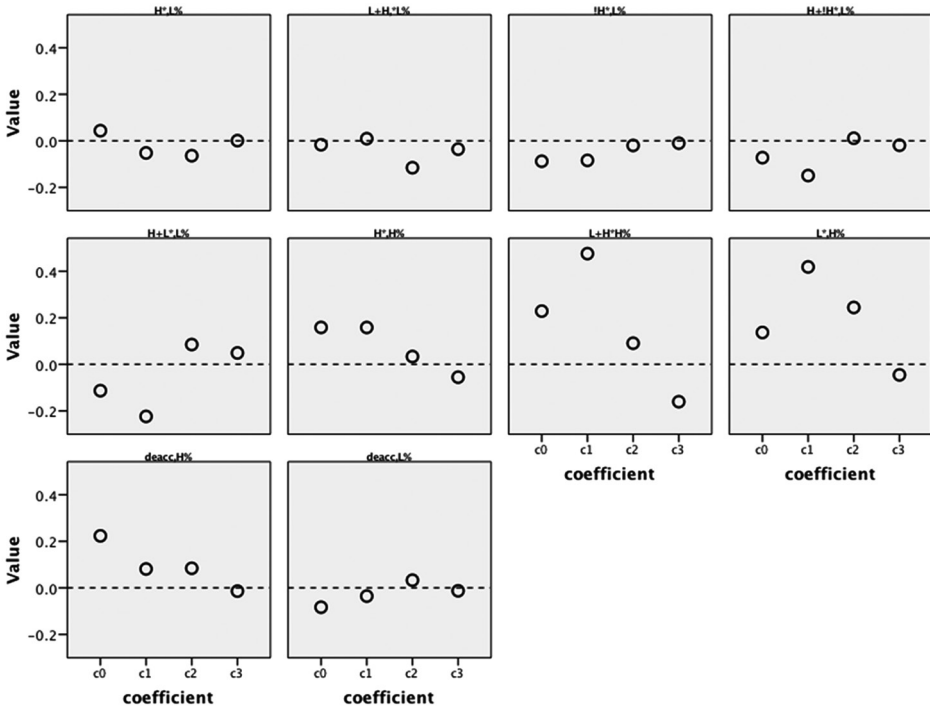


Figure 4. Four-coefficient F0 profiles for ten contours observed in the data. The values of the coefficients (c_0 - c_3) are mapped on the y-axis

ANOVA (as discussed by e.g., Baayen, Tweedie, & Schreuder, 2002; Quené & Van den Bergh, 2004, 2008). They are more robust with respect to handling missing data. This is useful for the present analysis, since the groups (i.e., nuclear contours) are not balanced with respect to sample size. Furthermore, LME models are more sensitive to differences between groups, because they can adjust for random effects such as items and subjects, where necessary. Including a random effect for each subject individually allows removal of variance that is due to systematic individual variation, making it easier to detect possible effects of the independent variable. Four LME models were constructed, one for each of the four coefficients. The models tested whether the contour label (henceforth: CONTOUR) is a significant predictor for a given coefficient. The models specified the coefficient as dependent variable, CONTOUR as fixed factor. However, as explained before, the items are not all of the same structure. The majority of items ($N = 167$) have an open lexically stressed syllable and penultimate stress (e.g., *Biber* (/ˈbiːbɐ/), but in the *Kamel* (/kaˈmeːl/) items ($N = 118$), the lexically stressed syllable is closed, and stress occurs on the final syllable of the word. Syllable structure and stress pattern of a word affect the shape of the pitch contour. For German, Möbius and Jilka (2007) found that in falling contours, the F0 peak occurs earlier in closed syllables and when the nuclear accent occurs on the last syllable of an IP. These two conditions are both met by the *Kamel* items. We may therefore expect that the contours of the *Kamel* items differ in their shape from the other contours, and consequently also in their coefficients. For this reason, I included STRESS (with the two levels *penultimate* and *ultimate*) as an additional fixed factor in the models. In a first step, full models including both predictors (CONTOUR and STRESS) and their interaction were specified. Predictors with a p -value larger than 0.1 were removed if this did not deteriorate the fit of the models (as estimated by a likelihood-ratio test). Both SUBJECT and ITEM were initially included as crossed random factors. However, in three of the four models (for AVERAGE, SLOPE, and WAVE), the variance explained by ITEM was effectively zero, and this factor was therefore removed from the model. All reported p -values were obtained by Laplace approximation.

The results show that in all models, CONTOUR was a significant predictor for the four coefficients. With the exception of four nuclear contour pairs, all contours differed from each other in at least one coefficient. The four contour pairs for which no statistically significant differences were observed were !H*,L% and Ø,L%, !H*,L% and L+H*,L%, and !H*,L% and H+L*,L%. There was no significant main effect of STRESS, but a significant interaction with CONTOUR in three of the four models (AVERAGE, SLOPE, and PARABOLA). For three nuclear contours (H+!H*,L%, L+H*,H% and L*,H%), items with ultimate stress showed a different pattern than items with penultimate stress. However, this did not have an influence on whether these nuclear contours differed significantly from other nuclear contours or not. Tables 2 to 5 present the results of the analyses.⁷

In what follows, I will describe the statistical results for the ten different nuclear contours following the order in the half-matrices in Tables 2 to 5 (i.e., going in columns from left to right). In keeping with Grabe et al. (2007), I assume that one significant difference between two nuclear contours is sufficient evidence that the two nuclear contour patterns that were labeled differently do have different mean F0 manifestations. Note, however, that also in the present data, Tables 2 to 5 show that in most cases the nuclear contour pair comparisons brought about more than one statistically significant difference. Table 6 gives an overview of the number of significant differences that were found for a given nuclear contour pair.

In order to keep the descriptions within reasonable length, not all 55 contour pair comparisons will be mentioned. For expository purposes, the descriptions will be restricted to a few comparisons for each contour. For ease of orientation, in each of the following paragraphs the contour under consideration will be set in boldface.

Table 5.

WAVE	H*,L%	L+H*,L%	!H*,L%	H+!H*,L%	H+L*,L%	H*,H%	L+H*,H%	L*,H%	∅,H%	∅,L%
∅,L%										
∅,H%										
L*,H%										
L+H*,H%										
H*,H%										
H+L*,L%										
H+!H*,L%										
!H*,L%										
L+H*,L%										
H*,L%										

Table 6. Number of significant differences between the nuclear contour pairs. Light-grey shading indicates one significant difference, dark-grey indicates more than one significant difference

	H*,L%	L+H*,L%	!H*,L%	H+!H*,L%	H+L*,L%	H*,H%	L+H*,H%	L*,H%	∅,H%	∅,L%
∅,L%	2	1	0	1	1	2	3	3	1	
∅,H%	1	2	2	3	2	1	2	2		
L*,H%	2	3	3	3	4	2	4			
L+H*,H%	4	4	4	4	3	2				
H*,H%	3	2	2	2	3					
H+L*,L%	1	3	0	2						
H+!H*,L%	2	2	1							
!H*,L%	1	0								
L+H*,L%	0									
H*,L%										

The first nuclear contour in the matrices, **H*,L%**, had a predominantly falling SLOPE (negative c_1) and a dome-shaped PARABOLA (negative c_2). With this shape, H*,L% contours differed from more cup-shaped contours like L+H*,H% or L*,H%. Furthermore, H*,L% contours were different from H+!H*,L% contours with respect to SLOPE. Both contours were predominantly falling, but the falling portion was larger in H+!H*,L% contours. From !H*,L% contours H*,L% were distinguished only in terms of AVERAGE (c_0), which was significantly lower in !H*,L% contours. However, no significant differences emerged with L+H*,L% contours.

L+H*,L% contours had a negative (i.e., dome-shaped) PARABOLA. This was sufficient to distinguish it from all other contours (except H*,L%), which were either cup-shaped (e.g., L*,H% contours), or not as strongly dome-shaped as L+H*,L% contours (e.g., H+!H*,L% contours).

Apart from the difference in AVERAGE, **!H*,L%** contours were in shape very comparable to H*,L% contours (see Figure 4), and contrasted with other contours in similar ways as H*,L% contours. With their mainly falling SLOPE, !H*,L% contours differed for example from H*,H% contours, which were predominantly rising. However, unlike H*,L% contours, !H*,L% contours were not found to be significantly different from deaccented items that were followed by a low boundary tone (∅,L%), or from contours labeled H+L*,L%.

H+!H*,L% contours were characterized by a predominantly falling SLOPE. With this, they differed from deaccented items that were followed by a low boundary tone, which were also falling, but had much shallower SLOPE than **H+!H*,L%** contours. SLOPE distinguished **H+!H*,L%** contours also from contours like **L*,H%**, **L+H*,H%** and **H*,H%**, which had a mainly rising SLOPE. Furthermore, **H+!H*,L%** contours were different from **H+L*,L%** contours both in terms of WAVE and in terms of AVERAGE, with the latter being much lower in **H+L*,L%** contours. **H+!H*,L%** contours were one of the contours for which there was a significant interaction with STRESS: in words with ultimate stress, **H+!H*,L%** contours were less strongly dome-shaped than in words with penultimate stress.

H+L*,L% contours, in turn, had a more steeply falling SLOPE than $\emptyset,L%$ contours. As in **H+!H*,L%** contours, the falling SLOPE of **H+L*,L%** contours distinguished them from all contours with a rising SLOPE (e.g., **L*,H%**).

Moving on to the first contour that had a predominantly rising SLOPE, **H*,H%**, we see that its SLOPE was more steeply rising than that of $\emptyset,H%$ contours. **H*,H%** contours differed from **L*,H%** contours in terms of PARABOLA, which was more narrowly cup-shaped in **L*,H%** contours than in **H*,H%** contours. Differences between **H*,H%** contours and **L+H*,H%** contours emerged for SLOPE, which was much steeper in **L+H*,H%** contours.

L+H*,H% contours differed from **L*,H%** contours among other things with respect to PARABOLA. Both contours had a cup shape, but it was more narrow in **L*,H%** contours than it was in **L+H*,H%** contours. There was furthermore a significant interaction with STRESS. In items with ultimate stress, these contours had a somewhat lower AVERAGE and a steeper overall SLOPE.

Like most other contours, the last “genuine” nuclear contour **L*,H%** was different from deaccented contours in both SLOPE and PARABOLA.⁹ In addition, **L*,H%** contours showed a significant interaction with STRESS in the two coefficients SLOPE and PARABOLA. In words with ultimate stress, the SLOPE was more steeply rising and more narrowly cup-shaped than in words with penultimate stress.

The two deaccented contours, finally, differed from each other in AVERAGE, which was higher in those contours that ended in a high boundary tone ($\emptyset,H%$) than in contours ending in a low boundary tone ($\emptyset,L%$).

The preceding analyses have shown that the majority of nuclear contours differ significantly from each other in at least one coefficient. However, with this type of analysis (especially when dealing with many categories), the internal structure of the data often remains rather opaque. A technique that can visualize data structures in an intuitive and easy-to-interpret format is classification trees. Classification trees predict the membership of cases (items) in the classes of a categorical dependent variable from their values on one or more predictor variables, and display the outcome of this process in a tree-like format. The structure of the tree reflects the structure of the data in the sense that similar cases end up in the same branch of the tree, and cases that bear less resemblance with each other end up in branches that are further apart. In our case, we would like to predict the (label of a) nuclear contour from the four variables c_0 , c_1 , c_2 and c_3 . To do this, I used the CART (Classification And Regression Tree) algorithm (Breiman, Friedman, Olshen, & Stone, 1984) as implemented in the `rpart` function in R. In a first step, the algorithm looks at all the predictor variables and selects the one that is most useful for splitting the data into two subsets which are each more homogeneous than the original data set. For each of these two subsets, the algorithm then creates two new subsets and so on. The resulting tree for this data set is shown Figure 5.

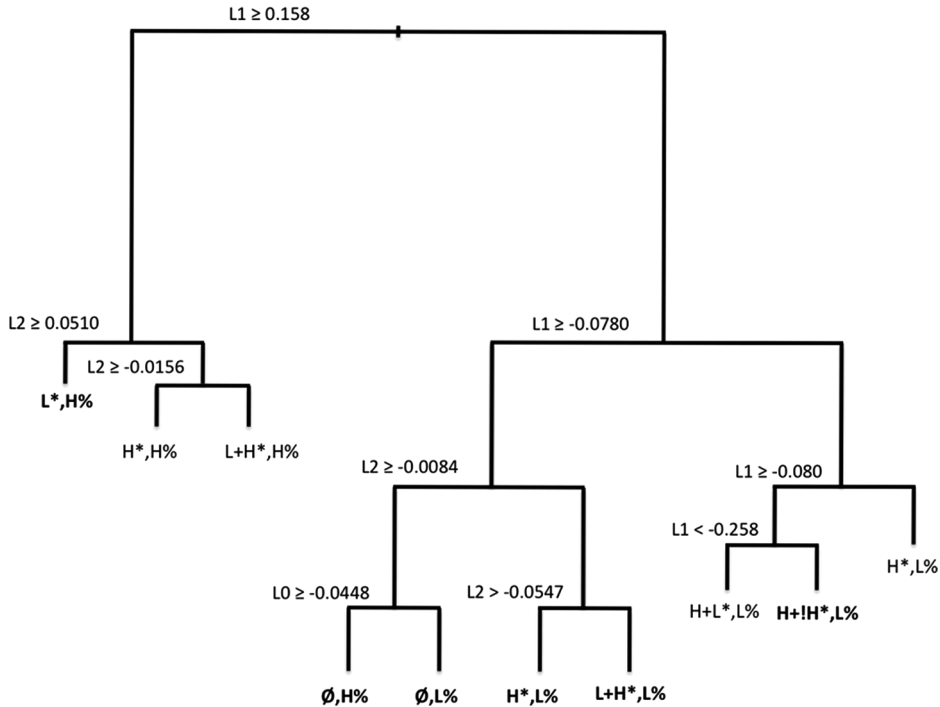


Figure 5. CART tree for CONTOUR as predicted by the four polynomial coefficients (c_0, c_1, c_2, c_3). The expression at each split gives information about the decision rule (e.g., “if c_1 is equal to or larger than 0.158, follow the left branch, otherwise the right” at the first split). Displayed is the initial tree; branches that remained after cost-complexity pruning are set in boldface

The tree captures the differences in shape between the nuclear contours in an elegant way. For example, predominantly rising nuclear contours are found in the left part of the tree, whereas nuclear contours that are mostly falling are located in the right part. This is achieved by the first splitting criterion, which is SLOPE: nuclear contours with a high positive c_1 (i.e., a predominantly rising slope) are sent to the left, nuclear contours with a mostly negative c_1 to the right. Another illustrative example is how PARABOLA (c_2) is used to sort $H^*,L\%$ and $L+H^*,L\%$ into different leaves: a given nuclear contour is more likely to be $L+H^*,L\%$ if its c_2 is smaller than -0.054 (i.e., if it is more strongly dome-shaped). Note at this point that this does not necessarily mean that the difference between the two nuclear contours in this coefficient is statistically significant; the criterion informs us that for the algorithm, the “best bet” for CONTOUR is $L+H^*,L\%$, if the nuclear contour has a c_2 smaller than -0.054 .

4 Extending the model: adding alignment parameters

In Appendix C of their article, Grabe et al. (2007, p. 305ff.) show in a constructed example how polynomial modeling reflects differences in fine phonetic detail such as alignment of F0 peaks with the segmental string in differences in the coefficients. They demonstrate that a change in alignment will always result in modifications in SLOPE and PARABOLA. We have seen this effect also in

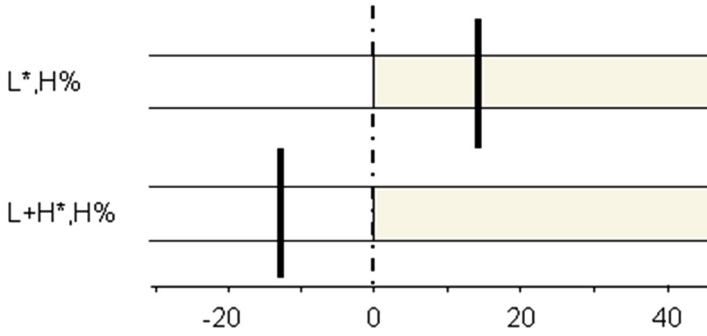


Figure 6. Graphical representation of the relative position of the F0 minimum (vertical black lines) for the two nuclear contours L+H*,H% and L*,H%, expressed as percentage of the stressed syllable duration (x-axis). The grey-shaded area marks the lexically stressed syllable, the vertical dotted line indicates the syllable onset

the present data: the nuclear contours L+H*,H% and L*,H% (which should differ in the position of the F0 minimum) differed in PARABOLA. However, while the general physical interpretation of the first four coefficients is straightforward, “translating” their values into alignment information is perhaps less obvious. This led me to investigate a different way of how one could get at alignment information that is more easily interpretable.

Each nuclear contour in the data is described by a polynomial function (see Equation 1). From this function, the position of local minima and maxima of the curve can be derived. In the present study, this was done by first using the polynomial functions calculated by Polyfit to create polynomial objects in Praat, and then applying customized Praat scripts to compute the local minima and maxima of these objects. Owing to the standardization of the scale, this position will be between -1 and 1 . For example, in the left panel of Figure 3 above, the F0 peak is located approximately around -0.5 . By measuring the absolute position and duration of the lexically stressed syllable and mapping it onto the normalized time scale, it is possible to define the location of the turning points relative to the syllable structure. Here, the location is expressed in percentage of the syllable’s duration (e.g., the maximum may occur at 45% into the lexically stressed syllable).

I calculated the alignment of the F0 minimum in the nuclear contours L+H*,H% and L*,H%, and tested the differences between the two nuclear contours statistically using LME models again, with the position of the F0 minimum (POSMIN) as dependent variable, CONTOUR as independent variable, and ITEM and SUBJECT as crossed random factors. Since the position of the lexically stressed syllable can have an effect on the shape of these contours, STRESS was included as an additional factor. The fitting procedure for fixed and random effects was the same as described before.

CONTOUR turned out to be a significant predictor for POSMIN: in L*,H% contours, the F0 minimum was aligned significantly later than in L+H*,H% contours (16.6% into the stressed syllable vs. -12.7% preceding the stressed syllable, $p < .001$). STRESS was not a significant predictor for POSMIN. Still, in ultimate stress words, the F0 minimum was aligned slightly earlier in L*+H words, although the interaction between CONTOUR and STRESS was not significant. The differences in alignment are illustrated in Figure 6.

Before we come back to these results in the general discussion, I will present another potential use for polynomial modeling of nuclear contours. With its capacity to capture fine phonetic detail, the polynomial approach not only allows for “comparisons of intonational systems across dialects

and languages” (Grabe et al., 2007, p. 298), it may also provide a window into developmental changes in nuclear contour realization. The next section investigates how the three different age groups have realized nuclear contours in these data.

5 Investigating age-related differences with polynomials

The field of child intonation research has become increasingly popular over the past years (for a review, see Snow & Balog, 2002). However, the development of pitch accent realization has not received much attention. The few studies that have been conducted on the phonetics of child intonation report that rising contours in particular may pose some difficulties for children. Snow (1998; see also Snow, 2001) asked four-year-olds to imitate falling and rising contours modeled by an adult speaker. The sentences used for imitation were short utterances like *The cat has a BOTTLE* (falling contour) or *Did you take your BOTTLE?* (rising contour). He found that children had the tendency to substitute the modeled rising contour with a falling one. When they produced rising contours, they did so with longer word durations coupled with narrower pitch range. This means that their rate of pitch change (or slope) was slower than that of the adult model. In contrast to that, they did not have any problems with imitating falling contours. Similar findings are reported by Loeb and Allen (1993), who also used an imitation paradigm.

However, these studies were more concerned with global pitch trends rather than specific pitch accents or nuclear contour types. The present data contain a number of different nuclear contour types, which allows for a more fine-grained analysis of possible age differences in nuclear contour realization. If children’s phonetic realization of nuclear contours develops with age, we should expect to find differences between the age groups in one or more of the coefficients that describe their nuclear contours. There is to my knowledge up to now only one study that used polynomial modeling to investigate pitch accent realization in children, but it used a different model and had a different objective: Ota (2006) analyzed the productions of Swedish toddlers in order to find out whether they reliably distinguish between the two Swedish word accents.¹⁰ Thus Ota was not interested in a developmental trend, but in whether the children were able to produce a specific contrast. His analysis involved the stylization of the children’s F₀ contours using the MOMEL algorithm (Hirst et al., 1993). MOMEL looks for a continuous series of quadratic second-degree polynomials (which are not orthogonal) that offer the best fit to the visible F₀ curve, and marks the inflections as turning points. To determine whether the children realized the contrast between the two accents, Ota checked whether the children’s productions contained the sequences of high and low turning points (as identified by the algorithm) that were appropriate for the two accent types. Indeed, he found that the children produced the F₀ pattern typical for Accent II words, indicating that they have acquired this phonological category. Ota’s study did not compare the children’s production to that of adult speakers, and the author himself raises the possibility that the children’s phonetic realization of Accent II contours differs from adult realizations (Ota, 2006, p. 244). In the present study, I am particularly interested in this potential difference. We will look at how children may differ from adults in the specific realization of several nuclear contours, that is, if they show differences in the coefficients that describe their contours. Polynomial descriptions offer an ideal way for studying phonetic differences *within* phonologically defined nuclear contour types, as differences in the phonetic realizations of one nuclear contour category will be reflected in differences in the coefficients.¹¹ Note that because MOMEL describes contours in a *series* of polynomial expressions rather than in one like Polyfit, it would be difficult to compare productions with each other. In the MOMEL approach, some F₀ contours may be described by a series of two polynomial expressions, while others may be described by three. It is unclear how one could determine the degree of (non-)similarity of the contours statistically. In addition,

Table 7. Frequency of nuclear contour types by age. The percentages indicate the relative frequency of this nuclear contour type for each age group. Nuclear contours that were produced at least seven times in each age group are given in bold

Nuclear contour	5 years		7 years		Adults	
	N	%	N	%	N	%
H*,L%	8	10.4	25	21.6	7	7.1
L+H*,L%	8	10.4	9	7.8	7	7.1
!H*,L%	9	11.7	5	4.3	2	2.0
H+!H*,L%	14	18.2	12	10.3	8	8.2
H+L*,L%	1	1.3	0	0.0	9	9.2
H*,H%	4	5.2	9	7.8	6	6.1
L+H*,H%	4	5.2	8	6.9	11	11.2
L*,H%	12	15.6	28	24.1	34	34.7
∅,H%	1	1.3	9	7.8	6	6.1
∅,L%	15	19.5	9	7.8	5	5.1
Other	1	1.3	2	1.7	3	3.1

non-orthogonal polynomials (as the ones implemented in MOMEL) have the disadvantage that their coefficients are not independent from each other and often highly correlated (Grabe et al., 2007, p. 287). In contrast, Polyfit-modeled nuclear contours are described in one set of minimally correlated Legendre polynomials, which can be straightforwardly subjected to analyses between age groups. This is exemplified in the following paragraphs.

Table 7 shows how often each age group produced each nuclear contour type. The distributions are uneven, with some nuclear contours used much more often by one age group than by the others, and some nuclear contours hardly or not at all produced by one age group. (Recall that this is natural spontaneous production data.) Therefore I only looked at those nuclear contours of which there are at least seven tokens in each age group (the bold rows in Table 7). The data set contained 172 items (42 from the five-year-old group, 74 from the seven-year-old group and 56 from the adults) from 68 speakers.

For each of the four nuclear contours (H*,L%, L+H*,L%, H+!H*,L% and L*,H%), I initially specified a LME model for each of the four coefficients (c_0, c_1, c_2, c_3), with the coefficient in question as the dependent variable and age (AGE) as fixed factor. Since the stress pattern (ultimate vs. penultimate) had an effect on two of the four nuclear contour shapes investigated, STRESS was included as an additional factor for those contours. SUBJECT and ITEM were first included as random factors, but in most cases the variance explained by these factors was so low that a simple linear model without random factors (the LM function in R) was fitted instead. Owing to the small sample sizes, the statistical power was rather low. For this reason, p -values smaller than .05 will be regarded as significant, and p -values up to and including .07 will be reported as marginally significant. As expected, there were effects of STRESS on PARABOLA for H+!H*,L% contours and L*,H% contours, which will not be reported again. There were significant or marginally significant effects of AGE for at least one coefficient in three of the nuclear contours. There was furthermore an interaction of AGE and STRESS for L*,H% contours. No significant differences were found for H*,L%.

For H+!H*,L% contours, five-year-olds turned out to produce contours with a larger fourth coefficient (WAVE) than adults (-0.002 vs. -0.090 , $p < .01$). The same was true for the seven-year-olds, whose WAVE was also larger than that of adults (0.009 vs. -0.090 , $p < .05$). Note, however,

Table 8. Significant and marginally significant effects of age on polynomial coefficients for three nuclear contours. The nuclear contour concerned is listed in the left-most column, the second column gives the coefficient for which a significant effect of AGE was found, the third column gives information about the direction of the effect

Nuclear contour	Coefficient	Effect of AGE
H+!H*,L%	WAVE	adults < 7 years = 5 years
L+H*,L%	WAVE	adults < 7 years = 5 years
L*,H%	PARABOLA	adults > 5 years (adults = 7 years < 5 years)

that since we are dealing with negative numbers, in this case “larger” means that the children’s coefficients were closer to zero, whereas the adults’ were more negative. While the adults’ negative c_3 shows a falling-rising-falling movement, the children’s near-zero c_3 indicates an absence of that wave shape. The two younger age groups did not differ from each other in this respect, $p = .7$.

A marginally significant difference between adults and children in terms of WAVE was also found for L+H*,L% contours. Again, the adults’ coefficients were smaller (i.e., more negative) than the children’s coefficients (−0.097 in adults vs. −0.0088 in five-year-olds, and vs. −0.011 in seven-year-olds, both $p = .06$), but there was no significant difference between the two child groups, $p = .9$.

In L*,H% contours, there was an interaction between AGE and STRESS for PARABOLA. The five-year-olds’ realizations were more cup-shaped than those of adults in penultimate stress words (five-year-olds: 0.432, adults: 0.215, $p < .05$), but *less* cup-shaped than those of adults in words with ultimate stress (five-year-olds: 0.196, adults: 0.353, $p < .05$). Five-year-olds realized L*,H% contours that were more narrowly cup-shaped in words with penultimate stress also in comparison with seven-year-olds (five-year-olds: 0.432, seven-year-olds: 0.144, $p < .01$). Adults and seven-year-olds did not differ from each other ($p > .2$ for both stress patterns). However, it is clearly problematic for the interaction that there were only four items in the five-year-old group in the category of penultimate stress words (while there were eight items in the ultimate stress category), casting some doubt on the reliability of this outcome. In Table 8, which summarizes the findings, the results for penultimate stress words have therefore been put in parentheses.

The results of the by-age analysis will be discussed along with the other results in the next section.

6 General discussion and conclusion

I modeled the F0 of spontaneously produced hand-labeled German nuclear contours quantitatively in terms of orthogonal Legendre polynomials. Statistical analyses have shown that the majority of these nuclear contours differ from each other significantly in at least one coefficient. These results show that on average, nuclear contours that were assigned different phonological labels clearly differ in the shape of their F0 contours, paralleling the findings by Grabe et al. (2007) for English read-out speech. In what follows, I will discuss some selected examples.

As one would expect, all of the rising contours such as L*,H% and L*+H,H% were found to be different from the falling contours such as H*,L% and H+!H*,L%, either in terms of SLOPE (predominantly rising vs. predominantly falling) or in terms of PARABOLA (more cup-shaped vs. more dome-shaped). However, the coefficients also picked up on differences within the group of rising

and falling contours. For example, I found that the contour $H+L^*,L\%$ had a more steeply falling slope than $H^*,L\%$. Within the group of rising contours, $L^*,H\%$ proved different from $L+H^*,H\%$ contours for instance in PARABOLA, which was more strongly cup-shaped in $L^*,H\%$ contours.

But there were also three statistically non-significant differences. Interestingly, they all involved the downstepped counterpart of $H^*,L\%$ contours: $!H^*,L\%$ contours. The accents H^* and $!H^*$ have the same shape, but differ in their relationship with the preceding pitch peak. While the pitch maximum of H^* accents is typically about the same height as that of the preceding high pitch accent or slightly lower, $!H^*$ accents are clearly lower than the preceding pitch peak. Hence, in order to be able to distinguish $H^*,L\%$ contours from downstepped $!H^*,L\%$ contours, information about the previous high pitch accent is necessary. However, the analyses showed that the two contours differed from each other, namely in their AVERAGE. This is not surprising, given that $!H^*$ accents especially in phrase-final position would be expected to be lower in the speaker's pitch range than H^* accents. At the same time, the fact that the two contours share the same overall shape justifies combining them in one category, and this is also what Grabe et al. (2007) did in their analyses. Against this background, I collapsed the two contour categories in the present data (resulting in 55 items in the new $H^*,L\%$ category) and re-ran the analyses. As a result, two of the three previously non-significant differences with other contours became statistically significant: the difference between $H^*,L\%$ and $H+L^*,L\%$ was significant for PARABOLA, $p < .01$, with $H^*,L\%$ contours being dome-shaped, but $H+L^*,L\%$ contours being overall more cup-shaped. In a similar vein, $H^*,L\%$ contours differed from $\emptyset,L\%$ contours, whose PARABOLA was also slightly cup-shaped, $p < .05$. Especially when dealing with small to medium-sized data sets, it seems useful to combine downstepped and non-downstepped nuclear contours in order to gain more statistical power for comparisons with other nuclear contour types. However, the difference between $H^*,L\%$ and $L+H^*,L\%$ did not reach significance even after combining downstepped and non-downstepped contours. This is an interesting finding, because it backs proposals that H^* and $L+H^*$ accents should be regarded as belonging to the same category (e.g., Grabe, 1998, for German; Ladd & Schepman, 2003, for English).

As in the English data, models using the first three coefficients (c_0-c_2) were sufficient to distinguish the majority of nuclear contours in this data set from each other. The nuclear contour pair $H+!H^*,L\%$ and $H+L^*,L\%$, though, differed from each other only in the fourth coefficient, WAVE, suggesting that this coefficient is necessary to distinguish these two nuclear contours from each other. This is somewhat surprising, as one would probably also expect a difference in SLOPE (cf. Table 1). In fact, we have seen that SLOPE was the most useful splitting criterion for the CART algorithm (cf. Figure 5). The fact that the difference between these two contours was small could be seen as support for the proposal that they are actually phonetic variations of the same accent type (e.g., Grice, Baumann, & Jagdfeld, 2007; Rathcke & Harrington, in press). However, it needs to be borne in mind that $H+L^*,L\%$ nuclear contours occurred only ten times in the data. More data are needed to find out whether the two contours differ consistently from each other in their coefficients.

An important question that this study has not addressed is to what extent statistically significant differences in coefficients correspond to differences in the perception of human listeners. Some of the differences in the present data are quite large and should be perceptible. For example, at the end of the utterance the mean difference between $H^*,L\%$ contours and $H^*,H\%$ contours is 0.225 normalized F0 values. For a speaker with a mean F0 of 160 Hz, this would correspond to a difference of 36 Hz or 4.1 semitones, a difference a human listener should be able to perceive without problems. Perception experiments that systematically vary the coefficients that describe an F0 contour could shed more light on how much change in a coefficient is needed for it to be perceptible to

listeners. For these types of experiments it is advantageous that polynomial expressions can serve directly as input for speech (re-)synthesis: the coefficient values can be entered into Equation 1 above to generate the number of data points necessary for a given stretch of (to-be-synthesized) speech. After multiplying the data point values with the appropriate fundamental frequency (e.g., 160 Hz) to get from normalized F0 to absolute F0, the data points can be plotted onto a speech sound object (using for instance Praat's "Manipulation object" functionality). The speech sound can subsequently be synthesized with the PSOLA method.

Another issue concerns the variability in the realization of the different nuclear contours. As discussed by Grabe et al. (2007), it is possible that the mean properties of all contours of one category (e.g., L+H*,H%) might be significantly different from contours of another category (e.g., H*,H%), but that the two classes might be so variable that they overlap considerably (p. 296). In the present data, the variation in coefficients produced by individual speakers was on average smaller than the mean differences in coefficients between the nuclear contours. For PARABOLA for instance, the mean adjustment for SUBJECT was 0.008, whereas the average difference between contours was 0.06, indicating that the overlap between categories is rather small.

In addition to applying the polynomial approach to German data, this article has introduced a way to derive relative alignment measures from the polynomial models, which are more easily interpretable than the plain coefficients. The position of tonal targets is expressed relative to the duration of the lexically stressed syllable. Using this parameter, the two rising nuclear contours L+H*,H% and L*,H% were shown to differ in their alignment of the F0 minimum. This result is in line with the descriptions of the accents in the GToBI literature (Benzmüller & Grice, 1997; Grice & Baumann, 2002; Grice et al., 2005): in L+H*,H% contours, the F0 minimum occurs in the pre-stressed syllable. In contrast, in L*,H% accents, it should be located in the stressed syllable, which is the case in these data as well. Although the effect of stress pattern (i.e., penultimate vs. ultimate stress) was not statistically significant, the direction of the effect suggests that tonal crowding may be present. Tonal crowding occurs when several tones are associated with the same segmental material, which can lead to small differences in alignment (Arvaniti, Ladd, & Mennen, 2006a, p. 670). In the present data, the F0 minimum in L*,H% contours occurs slightly earlier in ultimate stress items. This could be explained by the fact that both tones are associated with the final syllable, causing the L* tone to move further to the left, away from the H%. This finding would be in accordance with the results of Möbius and Jilka's (2007) corpus study for falling contours, which found that in ultimate stress words, the F0 maximum occurs earlier than in penultimate stress words owing to tonal repulsion. However, more data are needed to confirm the tendency observed in the present corpus.

Thus, relative alignment parameters derived from polynomial modeled curves are able to capture the phonetic differences between nuclear contours in an easily interpretable format. These alignment parameters always indicate the position of the minima and maxima with respect to the stressed syllable, irrespective of whether that syllable is preceded or followed by more than one syllable. In this respect this way of representing alignment can be more useful with more variable speech material than the plain coefficients. Importantly, deriving the F0 turning points from the modeled curve is a less subjective procedure than determining them manually from the original pitch track. As mentioned in the introduction, in many cases, tonal targets cannot be unambiguously identified because of discontinuities in the visible contour, which means for the labelers that they have to make (sometimes arbitrary) decisions as to where to locate the targets. The alignment parameter provides intonational phonologists with a simple tool to compare how nuclear contours differ in the alignment of their tonal targets with the segmental string, avoiding the problems

associated with manual location of pitch events. In this way, the polynomial approach can be smoothly integrated with the productive line of research on “segmental anchoring” (among others, Arvaniti et al., 1998; Arvaniti, Ladd, & Mennen, 2006b, for Greek; Atterer & Ladd, 2004, for German; Prieto, Santen, & Hirschberg, 1995, for Spanish; Silverman & Pierrehumbert, 1990, for American English).

Finally, this study has shown how polynomial modeling can be used to investigate differences in nuclear contour realization across age groups. Although the sample sizes are small, the within-category analysis of four nuclear contours has yielded some interesting findings. In ultimate stress words, five-year-olds turned out to have a less strongly cup-shaped PARABOLA than adults in L*,H% contours. This indicates that the cup shape of this contour was somewhat “flatter” than that of adults. I also found that in two nuclear contours, H+!H*,L% and L+H*,L%, for both the five- and the seven-year-olds, the absolute value of the fourth coefficient (WAVE) was low. This is a sign that their nuclear contours were not as “wiggly” as the ones produced by the adults.

The flatter PARABOLA in L*,H% contours in five-year-olds could be an indication that the children have difficulties raising the pitch towards the end of the utterance. This is in accordance with previous studies, which found that young children’s rising contours are not yet adult-like (Loeb & Allen, 1993; Snow, 1998). In particular, these studies found that children produce slower speeds of pitch change in rising contours than adults. Researchers have tried to explain this phenomenon by the increased physiological effort in rising accents (cf. Lieberman’s breath group theory, Lieberman, 1967; Snow, 1998). Xu and Sun (2002) found that adults, too, take longer to increase pitch than to decrease it. If rising contours are more difficult for adult speakers, it seems reasonable that they are even more demanding for children whose laryngeal and respiratory systems do not become adult-like before age twelve to fourteen (Stathopoulos & Sapienza, 1997). However, since there were not enough items in the corpus to compare also the other rising contours (H*,H% and L+H*,H%) across age groups, we do not know if it is rising contours in general that are difficult to produce for young children, or rather contours in which the lexically stressed syllable is low and followed by a rise, as in L*,H% contours. Previous studies that reported difficulties with rising contours did not specify the type of contours; it is therefore possible that the contours that the children had to imitate happened to be ones containing L* accents.

The findings regarding the absence of the wiggly shape in H+!H*,L% and L+H*,L% contours are new and have not been reported before. Other studies have concentrated on either the falling portion or the rising portion of a nuclear contour, but not on the overall movement of a nuclear contour, such as the fall-rise-fall movement in L+H*,L% nuclear contours. This movement is reflected in the fourth coefficient of the polynomial model. As explained before, the higher-ranking coefficients pick up the more rapidly changing properties of the F0. The fact that the children’s $|c_3|$ was so small seems to indicate that they are not (yet) able to modulate their pitch as fast as adults. The increased articulatory effort needed for raising F0 appears to pose problems for children not only in continuously rising contours, but also when more complex pitch movements are involved in realizing a contour. Here polynomial modeling may enable us to gain even better insights into children’s pitch contour realizations. More research (for example along the lines of Xu & Sun, 2002) is needed to obtain more conclusive information about the developmental trajectory of speed of pitch change control.

With regard to the age group differences found here the same caveat applies as with the differences between nuclear contour types: the differences in coefficients may not be perceptible. On the other hand, they can still be informative. In connection with the acquisition of segmental phonology the term “covert contrasts” (Scobbie, Gibbon, Hardcastle, & Fletcher, 1996) has been used to describe

situations in which children already produce a contrast between two categories (e.g., the voice onset time contrast between voiced and voiceless stops such as /t/ and /d/), but because realizations fall within one perceptual category for adults, the contrast is not recognized by the adult listeners. Analyzing children's speech acoustically can recover these imperceptible differences and can tell us something about what a speaker has acquired (Scobbie et al., 1996, p. 44). The same rationale can be applied the other way round: even though children may produce sound patterns (such as nuclear contours) that are perceptually equivalent to adult productions, there may still be systematic underlying differences, which can tell us something about the way the speech patterns mature with age.

To conclude: first, the polynomial approach can be readily applied to German natural speech and be used as a tool to obtain quantitative evidence for phonological labels. Second, the polynomial approach can be extended to include linguistic alignment information, making it even more accessible to the phonological research community. Third, by enabling the analysis of phonetic variation within nuclear contours across speaker groups, polynomial modeling can improve our understanding of language development.

Acknowledgements

I would like to thank Jan P. de Ruiter for providing me with the Polyfit program, as well as many helpful discussions and comments. Thanks also to Bettina Braun for help with the statistical analysis, and to Stefan Baumann, David Thomson and one anonymous reviewer for many insightful comments on an earlier version of this article.

Notes

- 1 One might object that it is unclear whether child speech should be labeled using an annotation system developed for adult speech. However, it is reasonable to assume that children are striving to model the intonation of adults, and it seems therefore useful to describe their speech in terms of that target model. Furthermore, previous analyses of the five-year-olds' data (Herbst, 2007) and of other parts of the corpus (De Ruiter, 2009) have already shown that the children's intonation patterns can be described using the GToBI system.
- 2 Originally, there were 32 H-% boundary tones and 45 H-^H% boundary tones.
- 3 Originally, there were 87 L-% boundary tones and 3 !H-% boundary tones.
- 4 Harmonics-to-noise ratio (HNR) quantifies the amount of additive noise in the speech signal. Additive noise can arise from airflow turbulences occurring at the vocal folds during phonation: when the vocal folds are not completely closed – as is the case in irregular phonation like creaky voice – air passes through the vocal folds and causes turbulence. This results in friction noise, which is in turn reflected in a higher noise level (Ferrand, 2002, p. 481). Hence, HNR can be used as an indicator of the periodicity of the speech signal.
- 5 The mean absolute duration of the analysis domain was 0.697 seconds ($SD = 0.221$ seconds).
- 6 The quantity that is minimized is a chi-square related merit function:

$$m = \frac{1}{N} \sum_{i=1}^N \left[\frac{data_i - pred_i}{w_i} \right]^2$$

where w_i is a weighting quantity indicating the relative contribution of data point i to the merit function. The program uses a General Linear Least Squares algorithm based on normal equations and Gauss-Jordan elimination, and is described in detail in Press, Teukolsky, Vetterling, and Flannery (1988).

- 7 There was no difference between the three age groups with respect to the mean error (a measure of distance between model and data) of the curves modeled by Polyfit. The mean error for the adult group was

- 0.12, which was not significantly different from the mean error in the five-year-old group (0.15, $p = .65$) and the mean error in the seven-year-old group (0.08, $p = .45$), as determined by means of a LME model with mean ERROR as dependent variable, AGE as fixed factor and WORD as random factor. The two younger age groups did not differ significantly from each other, either ($p = .24$).
- 8 For reasons why Bonferroni correction and other p -value adjustments are inappropriate for this type of analysis, see Gelman, Hill, and Yajima (2009), and Gelman and Tuerlinckx (2000).
 - 9 One of the reviewers remarked that it would be problematic to compare “real” nuclear contours with post-nuclear ones (i.e., deaccented syllables plus boundary tone), as the tonal movement in the latter ones depends solely on the type of boundary tone: the type of tonal movement around the accented syllable is constitutive for a nuclear contour, but this tonal movement is lacking on deaccented syllables. I fully agree with the reviewer that the tonal movement is lacking in deaccented contours. However, this is exactly what the polynomial models reveal. Unlike most other contours, deaccented contours do not “score high” on any of the four coefficients; they do not have a distinctive shape. All that can be said about them is that they have either a slight upward movement (as in the case of deaccented items followed by a high boundary tone) or a slight downward movement (as in the case of deaccented items followed by a low boundary tone), as can be seen in their SLOPE and PARABOLA.
 - 10 Swedish has two lexically contrasting pitch patterns associated with the stressed syllable, which are usually referred to as Accent I and Accent II.
 - 11 Of course, this approach presupposes that there *is* a phonology of intonation, a viewpoint that is still disputed by some researchers (e.g., Xu & Xu, 2005). A discussion of this debate is beyond the scope of this article, but the interested reader is referred to Ladd (2008) and Arvaniti and Ladd (2009) for an in-depth treatise of the issues involved.

References

- Arvaniti, A., & Ladd, R. D. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26(1), 43–74.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: The case of Greek prenuclear accents. *Journal of Phonetics*, 26, 3–25.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (2006a). Phonetic effects of focus and “tonal crowding” in intonation: Evidence from Greek polar questions. *Journal of Phonetics*, 48, 667–696.
- Arvaniti, A., Ladd, D. R., & Mennen, I. (2006b). Tonal association and tonal alignment: Evidence from Greek polar questions and contrastive statements. *Language and Speech*, 49(4), 421–450.
- Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of “segmental anchoring” of F0: Evidence from German. *Journal of Phonetics*, 32, 177–197.
- Baayen, H. R., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baayen, H. R., Tweedie, F. J., & Schreuder, R. (2002). The subjects as a simple random effect fallacy: Subject variability and morphological effects in the mental lexicon. *Brain and Language*, 81(1–3), 55–65.
- Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes. R package version 0.9975-12.
- Baumann, S. (2006). *The intonation of givenness: Evidence from German* (Linguistische Arbeiten Vol. 508). Tübingen: Niemeyer.
- Beckman, M., & Ayers, E. (1997). Guidelines for ToBI labelling, version 3. Ohio State University: The Ohio State University Research Foundation.
- Benzmüller, R., & Grice, M. (1997). Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. *Phonus*, 3, 9–34.
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Institute of Phonetic Sciences, University of Amsterdam Proceedings*, 17, 97–110.
- Boersma, P., & Weenink, D. (1992–2008). Praat: Doing phonetics by computer (Version 5.0.35). Amsterdam: Institute of Phonetics.

- Boutsen, F. R., & Hood, S. B. (1996). *Determinants of speech rate and fluency in fast and slow speaking normally fluent children*. Paper presented at the 3rd International Conference on Speech Motor Production and Fluency Disorders, June 5–8. Nijmegen, Netherlands.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and regression trees*. Belmont, CA: Wadsworth.
- Ferrand, C. T. (2002). Harmonics-to-noise ratio: An index of vocal ageing. *Journal of Voice*, 16(4), 480–487.
- Fujisaki, H. (1992). Modelling the process of fundamental frequency contour generation. In Y. Tokura, E. Vatikiotis-Bateson, & Y. Sagisaka (Eds.), *Speech perception, production and linguistic structure* (pp. 313–328). Amsterdam: IOS Press.
- Gelman, A., Hill, J., & Yajima, M. (2009, July 15). Why we (usually) don't have to worry about multiple comparisons. Available from <http://arxiv.org/abs/0907.2478v1> [2009, September 7].
- Gelman, A., & Tuerlinckx, F. (2000). Type S error rates for classical and Bayesian single and multiple comparison procedures. *Computational Statistics*, 15, 373–390.
- Grabe, E. (1998). *Comparative intonational phonology: English and German* (Vol. 7). Wageningen: Ponsen en Looien.
- Grabe, E. (2002). The IViE labelling guide [Website]. Available from <http://www.phon.ox.ac.uk/IViE/guide.html> [2009, March 6].
- Grabe, E. (2004). Intonational variation in English. In P. Gilles & J. Peters (Eds.), *Regional variation in intonation* (pp. 9–13). Tübingen: Niemeyer.
- Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, 50(3), 281–310.
- Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, 191, 267–298.
- Grice, M., Baumann, S., & Benz Müller, R. (2005). German intonation in autosegmental-metrical phonology. In J. Sun-Ah (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press.
- Grice, M., Baumann, S., & Jagdfeld, N. (2007). *Evidence for tonal identity from peak scaling under pitch span variations*. Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken.
- Grice, M., Reyelt, M., Benz Müller, R., Mayer, J., & Batliner, A. (1996). *Consistency in transcription and labelling of German intonation with GToBI*. Paper presented at the 4th International Conference on Spoken Language Processing (ICSLP), Philadelphia.
- Herbst, L. E. (2007). *German 5-year-olds' intonational marking of information status*. Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken.
- Hirst, D., Di Cristo, A., & Espesser, R. (1993). Levels of representation and levels of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 51–88). Dordrecht: Kluwer Academic Press.
- Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118(2), 1038–1054.
- Kuperman, V., Bertram, R., & Baayen, H. R. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, 23(7/8), 1089–1132.
- Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed., Vol. 119). Cambridge: Cambridge University Press.
- Ladd, D. R., & Schepman, A. (2003). “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, 31(1), 81–112.
- Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.
- Loeb, D. F., & Allen, G. D. (1993). Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research*, 36(1), 4–13.
- Möbius, B., & Jilka, M. (2007). *Effects of syllable structure and nuclear pitch accents on peak alignment: A corpus-based analysis*. Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken.

- Ota, M. (2006). Children's production of word accents in Swedish revisited. *Phonetica*, 63(4), 230–246.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge, MA: MIT Press.
- Pitrelli, J. F., Beckman, M., & Hirschberg, J. (1994). *Evaluation of prosodic transcription labelling reliability in the ToBI framework*. Paper presented at the 3rd International Conference on Spoken Language Processing (ICSLP), Yokohama.
- Plag, I., & Baayen, H. R. (2009). Suffix ordering and morphological processing. *Language*, 85(1), 109–152.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1988). *Numerical recipes in C: The art of scientific computing*. Cambridge: Cambridge University Press.
- Prieto, P., Santen, J. v., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429–451.
- Quené, H., & Van den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43(1–2), 103–121.
- Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects with binomial data. *Journal of Memory and Language*, 59(4), 413–425.
- R Development Core Team (2008). R: A language and environment for statistical computing (Version 2.6.2). Vienna: R Foundation for Statistical Computing.
- Rathcke, T., & Harrington, J. (2010). The variability of early accent peaks in Standard German. *Papers in laboratory phonology 10*, 533–556.
- De Ruiter, J. P. (2008). Polyfit: Fitting Nth order orthogonal polynomials on weighted data sets (Version 1.0). Nijmegen. Available from jan.deruiter@uni-bielefeld.de
- De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German “Vorfeld” by children and adults. *The Linguistic Review*, 26(2–3), 329–354.
- De Ruiter, L. E. (2010). Studies on intonation and information structure in child and adult German. PhD Thesis, Radboud University Nijmegen, Nijmegen.
- Scobbie, J. M., Gibbon, F., Hardcastle, W. J., & Fletcher, P. (1996). Covert contrast as a stage in the acquisition of phonetics and phonology. *QMC Working Papers in Speech and Language Sciences*, 1, 43–62.
- Silverman, K. E., & Pierrehumbert, J. B. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 72–106). Cambridge: Cambridge University Press.
- Snow, D. (1998). Children's imitations of intonation contours: Are rising tones more difficult than falling tones? *Journal of Speech, Language, and Hearing Research*, 41(3), 576–587.
- Snow, D. (2001). Imitation of intonation contours by children with normal and disordered language development. *Clinical Linguistics and Phonetics*, 15(7), 567–584.
- Snow, D., & Balog, H. L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025–1058.
- Stathopoulos, E. T., & Sapienza, C. M. (1997). Developmental changes in laryngeal and respiratory function with variations in sound pressure level. *Journal of Speech, Language, and Hearing Research*, 40(3), 595–614.
- Staub, A., Grant, M., Clifton, C. J., & Rayner, K. (2009). Phonological typicality does not influence fixation durations in normal reading. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 35(3), 806–814.
- Sturm, J. A., & Seery, C. H. (2007). Speech and articulatory rates of school-age children in conversation and narrative contexts. *Language, Speech, and Hearing Services in Schools*, 38(1), 47–59.
- Taylor, P. (2000). Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America*, 107(3), 1697–1714.
- Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America*, 111(3), 1399–1413.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159–197.