Original Paper

Phonetica 2006;63:230–246
DOI: 10.1159/000097307

Received: July 12, 2006
Accepted: October 28, 2006

Children's Production of Word Accents in Swedish Revisited

Mitsuhiko Ota

School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, Scotland, UK

Abstract

The two word accents in Stockholm Swedish (Accent I and Accent II) are distinguished by a consistent falling pitch contour on the stressed syllable of Accent II words. The current study presents new types of evidence that this feature of word accent can be systematically found in words produced by 16- to 18-month-old Swedish-speaking children. Compared to other disyllabic productions, the stressed syllable in Accent II words have a larger F_0 decline, more sequences of high-low turning points identified by a stylized contour algorithm and an earlier F_0 peak. In addition, a negative correlation is found between the value of F_0 peak and F_0 change in the stressed syllable. Taken together, these findings indicate that children learning Swedish have internalized a subtle but lexically relevant pitch contour by a very early stage of word production.

Introduction

Human infants exhibit precocious pitch perception ability. Within a few months after birth, they can discriminate linguistic stimuli that differ in the mean fundamental frequency (F_0) of syllables [Bull et al., 1985] or in F_0 contour [Karzon and Nicholas, 1989; Nazzi et al., 1998]. By 7–8 months, their reaction to complex tonal stimuli shows fundamental characteristics of adult pitch perception, such as constancy of overtones derived from the same F_0 [Clarkson and Clifton, 1985; Montgomery and Clarkson, 1997]. These findings indicate that before children begin to produce words around 12 months, they are capable of processing linguistically relevant pitch information in a manner similar to adults. In fact, cross-linguistic research has shown that children's prelinguistic babbling already reflects the overall pitch contour characteristics of the ambient language [Hallé et al., 1991; Whalen et al., 1991].

It is still not clear, however, to what extent children at the onset of word learning are able to extract and remember distinctive pitch patterns that contrast the individual lexical items within their ambient language. A challenge that children may face in extracting pitch levels or contours that are specific to words or morphemes is that not

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/06/0634–230
$23.50/0
Accessible online at:
www.karger.com/journals/pho

Mitsuhiko Ota
School of Philosophy, Psychology and Language Sciences
University of Edinburgh, 14 Buccleuch Place
Edinburgh, EH8 9LN (UK)
Tel. +44 (0) 131 650 3949, Fax +44 (0) 131 650 6883
E-Mail mits@ling.ed.ac.uk

all aspects of pitch modulation are lexically relevant. Pitch can also signal grammatical categories, phrase boundaries, focus, and utterance-level semantics and pragmatics. These various linguistic functions of pitch (in addition to the paralinguistic ones) can obscure patterns that are assigned to specific lexical items.

One learning mechanism that a child may invoke to discover lexically relevant pitch patterns is selective attention to invariant F_0 contours that recur with the same words across different linguistic contexts. Invariance detection has been proposed as a general perceptual learning mechanism [Gibson, 1969], and recent research shows its applicability to a range of linguistic domains, including speech segmentation [Hollich et al., 2005], auditory-visual perception [Prince and Hollich, 2005] and word mapping [Gogate et al., 2006]. Cross-contextual invariance can also serve as an important cue for learners in identifying pitch properties of lexical items. A pitch contour that recurs with a whole word is a better candidate for a lexical property than one that changes its shape across different linguistic contexts. In support of this view, children tend to acquire lexical tones earlier in a language that shows fairly regular pitch marking of individual words (e.g., Chinese [Erbaugh, 1992; Hua and Dodd, 2000; Li and Thompson, 1977]) than in a language that has less consistent word-level pitch patterns due to effects of non-lexical tones and tone sandhi (e.g., Sesotho [Demuth, 1989, 1993]). Furthermore, invariant pitch patterns are usually incorporated into children's production earlier than variant ones within the same language. In Tokyo Japanese, for instance, children show systematic production of the relatively invariant pitch accent contour before the more variant phrase-initial pitch rise, whose application is conditioned by syllable structure and the lexical accent [Ota, 2003].

An interesting question is whether children are also sensitive to pitch contours that remain invariable only in some part of the word (e.g., in its stressed syllable). If identification of invariance applies to linguistic units smaller than the word, then even such a pitch pattern should be predicted to emerge early in children's word production. A language that offers a suitable testing ground for this prediction is Swedish. Swedish has two lexically contrasting pitch patterns associated with the stressed syllable, usually referred to as Accent I (or 'acute' accent) and Accent II (or 'grave' accent). The contours of these two word accents in the Stockholm variety of Swedish are illustrated in figure 1. In citation forms, the apparent difference between Accent I and II is the number of pitch peaks: Accent I has one, and Accent II has two. However, the single peak in Accent I and the second peak in Accent II in these examples are not products of lexical tones, but of a sentence accent that signals the focal point of the utterance [Bruce, 1977, 1987]. Thus, these peaks are not attested on the same words when the focus position is shifted to a different word in the utterance, as illustrated by figure 1c and d. Isolated words, such as those in figure 1a and b, automatically receive the sentence accent because they constitute an utterance by themselves. By contrast, the first peak in Accent II words is a lexically assigned property and is present regardless of the sentential or pragmatic context. Figure 2 illustrates how an Accent II word is marked by a pitch peak on the stressed syllable whether it is in focus or not. Because the realization of the rise toward the peak varies depending on the type of onset consonant, the single consistent acoustic criterion in Swedish word accent is the falling pitch contour that occurs on the rhyme of the Accent II stressed syllable [Engstrand, 1995, 1997]. The rise toward the post-stress syllable, on the other hand, is lexically variable in the sense that it only occurs when an Accent II word is in focus. Note that the scope of the consistent Accent II pitch contour is limited to the stressed syllable rather than the whole word. If

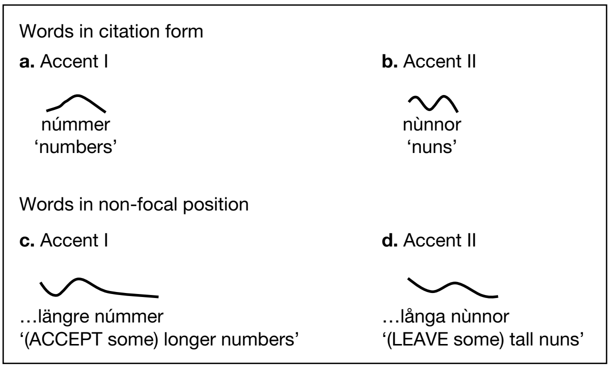

Fig. 1. Pitch contours of Accent I and II words in Stockholm Swedish [adopted from Bruce, 1977]. Examples **a** and **b** show contours in citation forms. Examples **c** and **d** show contours of words in non-focal position. Following the convention in Swedish phonology, Accent I is notationally indicated by an acute diacritic on the stressed vowel and Accent II by a grave diacritic on the stressed vowel. In **a**, *númmer* has a single pitch peak around the long consonant after the stressed vowel. In **b**, *nùnnor* has one pitch peak within or before the stressed vowel and another within the following vowel; a pitch valley occurs around the medial long consonant. In **c**, there is no pitch turning point within *númmer*. In **d**, *nùnnor* has just one peak within the stressed vowel; the pitch then declines toward the following consonant.

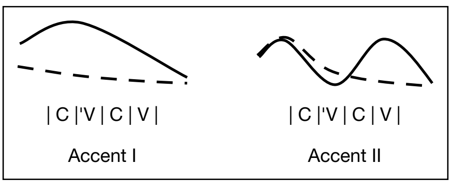

Fig. 2. Schematic illustration of Accent I and Accent II contours in focal (solid line) and non-focal (dotted line) position. The falling pitch contour on the stressed vowel in Accent II is the most invariable marker of this prosodic contrast. The precise segment-pitch alignment may differ depending on the segmental quantity of the stressed syllable [Bannert and Bredvad, 1975; Bruce, 1987]. The alignment shown here is that typical of a long stressed vowel followed by a short consonant. When the stressed vowel is short and is followed by a long consonant, the second consonant tends to coincide with the peak of Accent I and the valley of Accent II.

this local regularity is detected and used as a cue for lexical learning, the falling contour on the stressed syllable can be expected to emerge as one of the earliest characteristics in children's production of Stockholm Swedish word accent.

Evidence pertaining to this prediction has been somewhat elusive, however. In an analysis of longitudinal data collected from a single subject between 18 and 26 months of age, Peters and Strömqvist [1996] examined the proportion of stressed syllables in Accent II targets that had an F_0 fall greater than 20 Hz and the proportion of post-stress syllables that had an F_0 rise greater than 20 Hz. While 51% of the post-stress syllables met the rise criterion, only 20% of the stressed syllables satisfied the fall criterion. Kadin and Engstrand [2005] measured the F_0 change in disyllabic Accent II words produced by eleven 18-month-old (as well as eleven 24-month-old) Swedish children. Many of the 18-month-olds did indeed show a pitch fall in the stressed syllable of Accent II words. Also in keeping with the two-peak contour of Accent II illustrated in figure 2, the F_0 peak in the stressed syllable occurred earlier than what would be expected in a single-peak contour (as in Accent I words, or initially stressed words in English). However, because the 18-month-olds produced too few Accent I words, it could not be established whether the pattern was characteristic of Accent II words, or both Accent I and Accent II words produced by Swedish children of this age. The paucity of Accent I in spontaneous production data is due to the fact that most of the earliest words Swedish children attempt are bare stems, which typically have Accent II. To cope with this problem, Engstrand et al. [1991] compared 17-month-olds' productions of Accent II words with all other disyllabic vocalizations, including Accent I words and vocalizations that could not be identified as words. In 3 of the 5 children they examined, a decline in F_0 was observed in the stressed syllable in Accent II targets, but the values of F_0 change did not differ significantly from those of the remaining disyllabic vocalizations. In contrast, a significant difference was found between Accent II words and other vocalizations in the F_0 rise toward the second syllable. In sum, the findings so far do not furnish conclusive evidence that Accent II words are consistently and contrastively marked by a lexical falling contour on the stressed syllable in the production of children around or before the age of 18 months, despite clear indication that the rising contour toward the post-stress syllable – presumably a more variant feature in the input – has been internalized in children's early word production.

Before we draw the conclusion that the lexical pitch fall lags behind the stress accent rise in development, some methodological issues in previous research need to be addressed. First, it is possible that an underlying pitch fall may not be observable when the visible F_0 contour is interrupted in the potential locus of the fall, that is, the rhyme of the stressed syllable. This is particularly true in cases where the stressed syllable is closed by a long voiceless stop, as in /titta/ ('look') and /dokka/ ('doll'). Figure 3 presents an example of F_0 track generated for a target Accent II word (/titta/) produced by a Swedish-speaking child. Note the large gap in the visible F_0 information during the post-vocalic closure of the stressed syllable. Many such words are likely to have been included in previous studies. Second, differences in the shape of F_0 contours cannot be determined by measurements of F_0 value alone. In some of the studies mentioned above, evidence for the Accent II-type fall was sought in an F_0 drop in the stressed vowel (sometimes called the 'fall parameter') measured as a difference in F_0 value from the turning point [Engstrand et al., 1991] or the F_0 onset of the stressed vowel [Peters and Strömqvist, 1996] to the F_0 offset of the vowel. Although a decline in the fall parameter can reflect the crucial hallmark of Accent II, it can also be recorded in a contour that has a single pitch peak in the stressed syllable followed by a fall that extends to the following syllable, a pattern sometimes found in Accent I words. What is important, in addition to the value of F_0 change in the stressed syllable, is the shape and timing of the pitch fall. As is evident from figure 2, a typical stressed-syllable fall in Accent II can be distinguished from a potential F_0 decline in Accent I in that it commences earlier and targets a low point around the end of the stressed syllable. The timing of the decline has been examined in the study by Kadin and Engstrand [2005], which shows that the F_0 peak in the stressed syllable occurs earlier in 18-month-olds' Accent II productions than in the disyllabic words produced by children acquiring American English. However, as noted above, no comparison was made with other

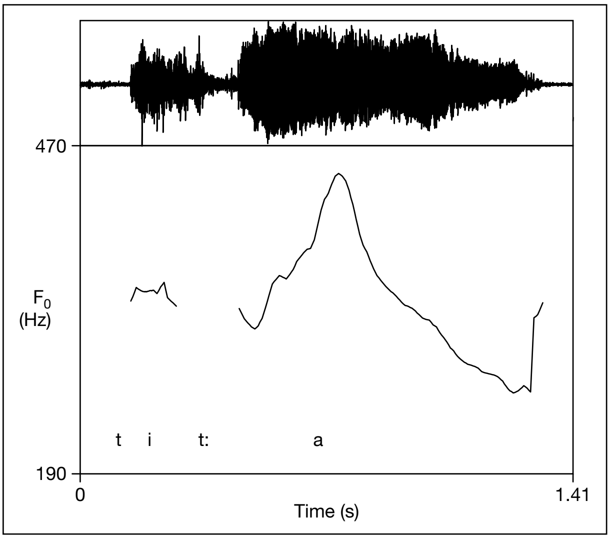

Fig. 3. Pitch contour of an Accent II word (/titta/ 'look') produced by a 17-month-old Swedish-speaking child (taken from Hanna's data described below). The upper panel, aligned with the F_0 tracking, shows the waveform.

disyllabic productions made by the Swedish children, leaving unresolved the question of whether this is a unique characteristic of Accent II words.

The reanalysis of Swedish children's early spontaneous speech production presented below is an attempt to overcome these issues with the use of alternative methods. The first step in the analysis was to control for the influence of short rhymes in the stressed syllables. This was done by measuring F_0 changes only in words that have visible F_0 tracks with sufficient duration in the stressed syllable. This procedure was meant to reduce the possibility of measuring F_0 tracks that are cut off toward the end of the stressed syllable due to lack of phonation. The second analytical strategy adopted was to identify – by means of polynomial interpolation – underlying turning points that characterize Accent II contours: a high turning point followed by a low turning point on the stressed syllable. This analysis should indicate whether children's Accent II productions can best be modelled in terms of such phonetic targets. Third, following Kadin and Engstrand [2005], the timing of the F_0 peak was measured in the stressed syllable of Accent II words, but crucially was also compared with the peak in other disyllabic vocalizations to see whether an early beginning of an F_0 decline, if any, was particular to Accent II. Finally, two correlation analyses were carried out to examine whether children's production exhibits the same types of relationship between F_0 peaks and changes we see in adults' production of Accent II words. In adult spontaneous production of Accent II words, a correlation is found between the peak and the fall of the stressed syllable, and also between the peak of the post-stress syllable and the rise from the stressed to the post-stress syllable [Engstrand, 1997]. In other words, the higher the F_0 peak in the stressed syllable, the larger the drop to follow, and the higher the F_0 peak in the post-stress syllable, the larger the rise from the stressed syllable. These are properties indicative of the presence of a low tonal target in the stressed syllable in Accent II words, and can be used to examine the marking of Accent II in children's production.

Table 1. Profile of children

Name	Didrik	Hannah	Kurt	Lina	Stig
Sex	male	female	male	female	male
Age, months and days	16.14	17.10	16.5	16.12	18.29

Method

The data consisted of spontaneous speech produced by 5 children from the Stockholm area. The recordings were originally made by a team in Sweden (including Olle Engstrand, Francisco Lacerda and Björn Lindblom) in conjunction with the Stanford Child Phonology Project (1980–1988) for a cross-linguistic study on infant vocalizations comparing babbling and early word production in American English, French, Japanese and Swedish. The 5 children are the same individuals studied by Engstrand et al. [1991], who analysed the recordings made at 17 months of age. To adjust for differences in vocabulary size, however, analyses in the current study were carried out for each subject at the 25-word point – the point that marks the first 30-min recording session during which the child spontaneously produced 25 different words. This roughly matched the 17-month point, although 3 of the children were 16 months and 1 of them 18 months old. The pseudonym, sex and chronological age of each child are given in table 1.

The recordings were of unstructured free interactions, which usually involved the child, the child's mother and a researcher. In some cases, the child's father, grandmother or other family members were also present.

Each file was first transcribed independently by 2 native speakers of Swedish using conventional orthography, and only those utterances that could be agreed on by both transcribers were used in the acoustic analysis. All productions of initially stressed disyllabic vocalizations by the children, as well as caretakers' productions of the same disyllabic words addressed to the child were extracted and digitized at a sampling rate of 22 kHz. All child samples were single-word utterances, but the adult samples were taken from single- and multiple-word utterances. F_0 plots were produced using the autocorrelation method in PRAAT [Boersma, 1993] and checked against narrow-band spectrograms. Spectrographic analysis was also carried out to mark consonant-vowel and vowel-consonant discontinuities as segmental boundaries. Samples were excluded when no reliable F_0 plots could be obtained due to overlapping, background noise, low intensity, whispering, screaming or creaky voice. Of the remainder, vocalizations that can be unambiguously identified as words were classified into Accent I or Accent II in accordance with the adult norm. In addition, unidentifiable initially stressed disyllabic productions were included in the category of 'other vocalizations' (i.e., non-Accent II disyllables) until these and Accent I words combined equalled the number of Accent II words. The balancing could not be accomplished for 2 children (Hanna and Kurt) due to an insufficient number of usable samples. The overall numbers of analysed words are shown in table 2.

Five types of measurement were made for each of the F_0 plots: (a) the stressed-syllable F_0 change, (b) the post-stress F_0 change, (c) the stressed-syllable max F_0, (d) the post-stress max F_0 and (e) the stressed-syllable F_0 peak timing. The stressed-syllable change and the post-stress change roughly correspond to the 'fall parameter' and 'rise parameter' in Engstrand et al. [1991]. The *stressed-syllable F_0 change* was measured as the F_0 difference between the visible turning point in the stressed syllable and the vowel offset. When no turning point could be identified in the stressed syllable, the F_0 of the vowel onset was used instead. The *post-stress F_0 change*, which roughly corresponds to the 'fall parameter' in Engstrand et al. [1991], was measured as the F_0 difference between the vowel offset of the stressed syllable and the turning point in the post-stress syllable. When no turning point was identifiable in the post-stress syllable, the F_0 of the vowel onset was used instead. The *stressed-syllable max F_0* and the *post-stress max F_0* were the F_0 maxima in the stressed syllable and the post-stressed syllable, respectively. These measurements were first made with all the samples in the data. A second set of measurements was made only with words with a visible F_0 track longer than 150 ms in the stressed syllable. The cut-off point was set based on a pre-analysis of the data, which showed that most stressed syllables containing either a long vowel (e.g., /so:va/ 'sleep') or a sonorant coda (e.g., /dansa/

Table 2. Disyllabic samples analysed in child speech

	Accent II		Accent I and other vocalizations	
	n	Targets included	n	Targets included
Didrik	24	<Didrik's real name> (1), *gumma* (2), *morfar* (1), *mamma* (2), *nalle* (3), *pappa* (13), *tända* (2)	24	*prosit* (1), <unidentifiable> (23)
Hanna	55	*borsta* (1), *docka* (16), *mamma* (3), *titta* (35)	22	*ajaj* (1), *tut-tut* (9), <unidentifiable> (12)
Kurt	28	*bocka* (2), *gurka* (1), *kaka* (1), <Kurt's real name> (1), *klocka* (2), *mamma* (9), *pappa* (6), *pippi* (6)	9	<unidentifiable> (9)
Lina	24	*bajsa* (2), *blomma* (1), *dansa* (1), *docka* (5), *hoppa* (1), *humla* (1), *kaka* (3), *nalle* (1), *sova* (4), *titta* (5)	24	*ajaj* (1), *hallå* (6), *hästen* (2), *katten* (1), *plåster* (1), *(po)tatis* (1), <unidentifiable> (12)
Stig	24	*docka* (8), *klocka* (1), *lampa* (4), *Lollo* (5), *mamma* (1), *nalle* (1), *pappa* (2), *pippi* (1), *trappa* (1)	24	*taxi* (6), <unidentifiable> (18)

Numbers in parentheses show the token of each word.

Gloss: *ajaj* 'ow-ow', *bajsa* 'poo', *blomma* 'flower', *bocka* 'bow', *borsta* 'brush', *dansa* 'dance', *docka* 'doll', *gumma* 'old woman', *gurka* 'cucumber', *hallå* 'hallo', *hästen* 'the horse', *hoppa* 'jump', *humla* 'bumblebee', *kaka* 'biscuit', *katten* 'the cat', *klocka* 'clock', *lampa* 'lamp', *Lollo* (character), *mamma* 'mommy', *morfar* 'grandfather', *nalle* 'teddy', *pappa* 'daddy', *pippi* 'birdie', *plåster* 'bandage', *potatis* 'potato', *prosit* 'bless you' (after a sneeze), *sova* 'sleep', *tända* 'turn on', *tanka* 'fill (with petrol)', *taxi* 'taxi', *titta* 'look', *trappa* 'stairs', *tut-tut* 'honk-honk'.

'dance') had a visible F_0 track that lasted at least 150 ms. The *stressed-syllable F_0 peak timing* was measured as the duration between the onset and the F_0 maximum of the stressed vowel.

In addition, a stylized contour analysis was carried out using the MOMEL algorithm developed by Hirst and Espesser [1993]. MOMEL looks for the continuous series of second-degree polynomials that best fits the visible F_0 contour, marking the inflection points as F_0 turning points. This method is sensitive to the scope of F_0 movements, so that a fall that is completed within a particular syllable can be distinguished from one that extends to the following syllable. The algorithm is also relatively impervious to microprosodic F_0 changes, and it can project turning points where no actual F_0 data are available due to discontinuities in the visible contours. It should be noted, however, that the interpolated turning points are hypothetical targets, which do not necessarily coincide with the actual point of change in the acoustic data. As such, the results of this analysis need to be interpreted only in light of the actual F_0 data. Example MOMEL outputs are given in figures 4 and 5. The size of the initial analysis window used to pick vertex candidates was set at 400 ms, the size of the second analysis window to fit a spline function to the candidates was set at 300 ms, and the maximal accepted error percentage for the polynomial approximation was 10%. The window size was adjusted occasionally when the outputs revealed apparently erroneous turning points due to changes in speech rate.

Results

Caretakers' Speech

Table 3 shows the mean values and standard deviations of stress-syllable F_0 change and post-stress F_0 change in the caretakers' production of disyllabic words addressed to the children. Most of the words were produced with a sentence accent

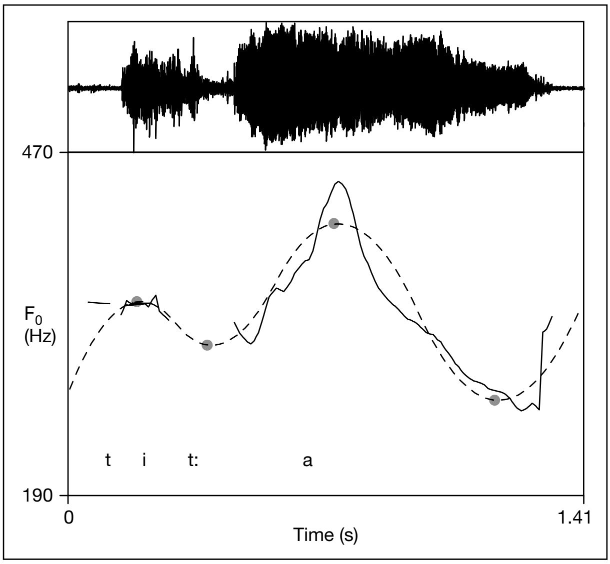

Fig. 4. F_0 curve and its stylized model generated by MOMEL for Hanna's production of an Accent II word, /titta/ 'look'. The dots on the smooth curve represent the hypothetical turning points generated by the algorithm. The upper panel shows the waveform aligned with the F_0 tracking.

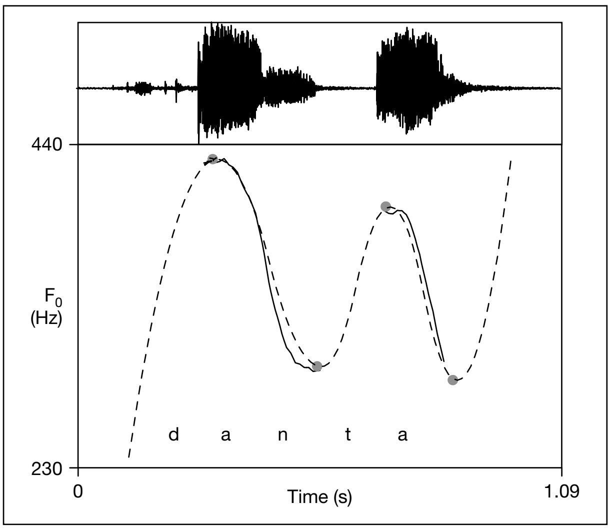

Fig. 5. F_0 curve and its stylized model generated by MOMEL for Lina's production of an Accent II word, /dansa/ (pronounced [danta]) 'dance'. The dots on the smooth curve represent the hypothetical turning points generated by the algorithm. The upper panel shows the waveform aligned with the F_0 tracking.

(88% of Accent I words and 68% of Accent II words). This means that most adult models of the Accent II words that children produce have a post-stress rise. This is not surprising, as caretakers may be aware of words that are familiar to the child, and place them more frequently in the position of focus. However, about a third of them do not, which means that children are exposed to the alternating Accent II contours depicted in figure 2. As the numbers in table 3 show, this alternation gives rise to a substantial variance in the values of the stress-syllable F_0 change and post-stress F_0 change. This is

Table 3. Mean values and standard deviations (SD) of stressed-syllable F_0 change and post-stress F_0 change in caretakers' speech addressed to children (Hz)

	Stressed-syllable		Post-stress		n
	mean	SD	mean	SD	
Accent I (with focus)	+27.5	64.8	−76.6	83.6	80
Accent I (without focus)	−13.0	6.2	−14.5	12.9	11
Accent II (with focus)	−32.5	16.8	+130.8	66.0	81
Accent II (without focus)	−20.6	14.3	−29.5	39.8	38

Table 4. Values (Hz) of stressed-syllable F_0 change

	Accent II targets			Other vocalizations		
	mean	SD	n	mean	SD	n
Didrik	−10.2	21.5	24	−1.7	25.0	24
Hanna	−15.5	47.6	55	−13.2	36.3	22
Kurt	−5.9	26.3	28	−20.2	39.6	9
Lina	−29.7	35.2	24	−22.4	26.5	24
Stig	−25.6	16.8	24	+1.2	41.3	24
Grand mean	−17.4	10.1	5	−11.3	10.7	5

Table 5. Values (Hz) of post-stress syllable F_0 change

	Accent II targets			Other vocalizations		
	mean	SD	n	mean	SD	n
Didrik	+24.5	35.7	24	+16.0	25.6	24
Hanna	+21.6	62.5	55	−5.1	69.0	22
Kurt	+85.6	46.7	28	−8.7	62.7	9
Lina	+78.4	73.5	24	+10.0	50.4	24
Stig	+65.2	42.5	24	+51.5	67.6	24
Grand mean	+55.1	30.1	5	+12.7	24.0	5

particularly true for Accent II words; when they are outside focus, the post-stress F_0 change shows a decline of about 30 Hz, but when they are within focus, the post-stress F_0 change has a very large rise that often exceeds 100 Hz. Such an exaggerated post-stress rise in Accent II appears to be a feature of child-directed speech in Swedish, as has been suggested by Engstrand et al. [1991]. On the other hand, the stressed-syllable values for Accent II words in and outside focus position are similar and confirm the relative phonetic invariance of the stressed-syllable fall.

F_0 Measurements of Children's Speech

Tables 4 and 5 present the values of stressed-syllable and post-stress F_0 change. A one-tailed paired t test shows that the post-stress syllable F_0 change was significantly

Table 6. Values (Hz) of stressed-syllable F_0 change (when visible F_0 in the stressed syllable is 150 ms or longer)

	Accent II targets			Other vocalizations		
	mean	SD	n	mean	SD	n
Didrik	−9.3	25.3	15	−1.7	25.0	24
Hanna	−31.7	38.1	9	−9.8	35.7	20
Kurt	−20.2	39.6	16	+7.6	21.3	9
Lina	−42.2	28.4	12	−25.7	28.5	16
Stig	−12.7	10.9	10	+8.9	39.7	14
Grand mean	−23.2	13.7	5	−4.1	14.2	5

Table 7. Values (Hz) of post-stress syllable F_0 change (when visible F_0 in the stressed syllable is 150 ms or longer)

	Accent II targets			Other vocalizations		
	mean	SD	n	mean	SD	n
Didrik	+18.1	34.4	15	+16.0	25.6	24
Hanna	+37.3	79.3	9	−3.0	68.7	20
Kurt	+67.3	42.9	16	−8.7	62.7	9
Lina	+91.7	66.9	12	+11.8	51.7	16
Stig	+72.4	49.1	24	+30.7	41.9	24
Grand mean	+57.4	29.4	5	+9.4	15.7	5

larger in disyllabic Accent II productions than in other disyllabic vocalizations [$t(4) = 2.53$, $p < 0.05$]. However, the difference for the stressed-syllable F_0 change was not significant [$t(4) = 0.93$, n.s.]. In all children, there was a greater F_0 increase from the offset of the stressed syllable to the post-stress syllable in Accent II targets than in other vocalizations. There was also a decline in the stressed-syllable F_0 change in Accent II targets, but this was not always larger than the stressed-syllable F_0 change in other vocalizations. These results are consistent with those of Engstrand et al. [1991], who found a significant difference between Accent II words and other vocalizations in the F_0 change toward the post-stress syllable, but not within the stressed syllable.

However, the outcome of the comparison was different when the analysis was limited to stressed syllables with a long rhyme. Tables 6 and 7 show measurements for disyllables with a stressed syllable containing a visible F_0 track that was 150 ms or longer. One-tailed t tests revealed a significant difference in both the stressed-syllable F_0 change [$t(4) = 5.64$, $p < 0.005$] and the post-stress F_0 change [$t(4) = 3.39$, $p < 0.05$]. In comparison to other vocalizations, Accent II target words had a larger F_0 drop in the stressed syllable and a larger F_0 rise toward the post-stress syllable. These results suggest that, when there is sufficient sonorant material in the stressed syllable to carry a visible pitch contour, Accent II words produced by 16- to 18-month-olds do differ from other vocalizations in that there is a distinct F_0 drop in the stress syllable in addition to an F_0 rise toward the post-stress syllable.

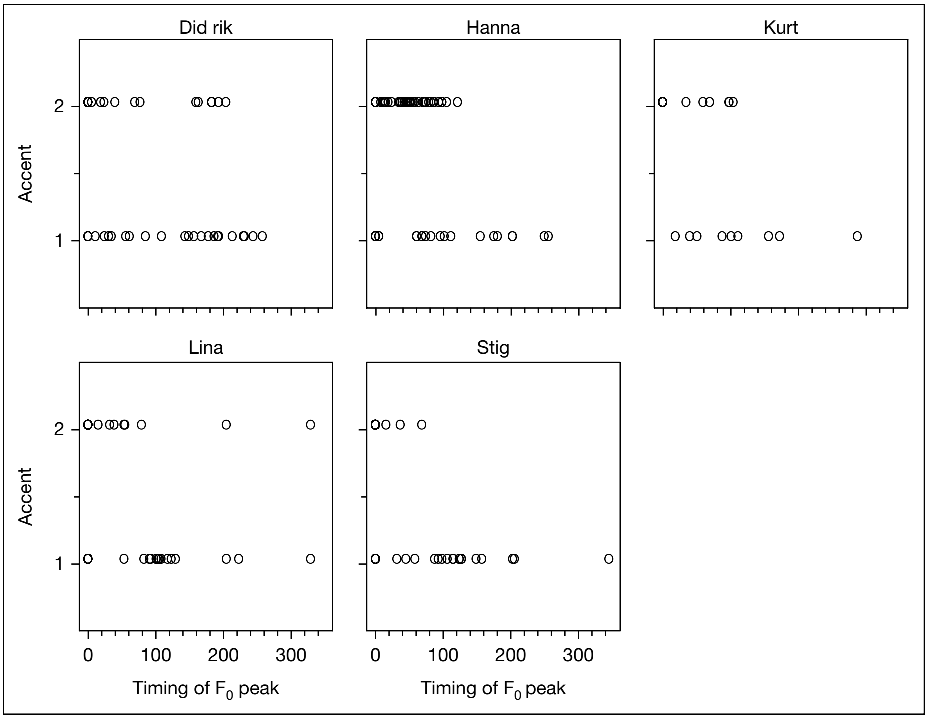

Fig. 6. Timing of the F_0 peak in the stressed vowel. Each circle represents the distance between the onset of the stressed vowel and the point of F_0 maximum in the vowel (measured in milliseconds). On the y axis, '2' indicates Accent II and '1' all other vocalizations.

Timing of F_0 Peaks

It was predicted that the falling contour that specifically targets the stressed syllable in Accent II should occur earlier in comparison to the decline, if any, in other vocalizations. Figure 6 plots the F_0 peaks recorded in the stressed syllable. Not surprisingly, the peaks are distributed quite widely in the 'other vocalizations' category (1 on the y axis), which is more heterogeneous by definition. By contrast, the peaks in Accent II words (2 on the y axis) are narrower in distribution and tended to occur near the onset of the stressed vowel. The mean distances of the F_0 peaks from the vowel onset in Accent II targets, summarized in table 8, ranged from 4.9 to 47.3 ms. As the average duration of the stressed vowel was about 170 ms, the F_0 in Accent II typically peaked within the first quarter of the vowel. A one-tailed paired t test showed that the mean F_0 peaks were significantly earlier in Accent II words than in other vocalizations [$t(4) = 7.17$, $p < 0.001$]. These results corroborate Kadin and Engstrand's [2005] finding that 18-month-olds' Accent II words have a peak soon after the onset of the stressed vowel. More importantly, the comparison between Accent II targets and other vocalizations shows that the early timing can be attributed to Accent II rather than a general feature of any disyllabic vocalization produced by young Swedish-speaking children. This offers more robust evidence that F_0 characteristics particular to Accent II have been internalized by the learners.

Table 8. Timing of the F_0 peak in the stressed vowel (ms after the vowel onset)

	Accent II targets			Other vocalizations		
	mean	SD	n	mean	SD	n
Didrik	47.3	73.3	24	117.7	86.6	24
Hanna	34.6	30.3	55	88.6	82.3	22
Kurt	12.9	30.3	28	113.7	82.6	9
Lina	31.3	77.7	24	76.9	85.7	24
Stig	4.9	15.5	24	86.0	87.0	24
Grand mean	26.2	17.1	5	96.6	18.0	5

Table 9. Proportions of stress falls and post-stress rises assigned by MOMEL in Accent II target words and other disyllabic vocalizations

	Stress fall, %		Post-stress rise, %	
	Accent II targets	other vocalizations	Accent II targets	other vocalizations
Didrik	58 (14/24)	25 (6/24)	79 (19/24)	63 (15/24)
Hanna	60 (33/55)	18 (4/22)	62 (34/55)	27 (6/22)
Kurt	54 (15/28)	11 (1/9)	92 (22/24)	33 (3/9)
Lina	79 (19/24)	11 (3/24)	79 (19/24)	42 (10/24)
Stig	96 (23/24)	13 (3/24)	100 (24/24)	46 (11/24)
Grand mean	69 (SD = 17.7)	16 (SD = 6.0)	82 (SD = 14.5)	42 (SD = 13.8)

Modelled F_0 Contours

Each disyllabic production was checked for two types of stylized turning point sequences identified by the MOMEL algorithm. A *stress fall* was defined as a sequence of a high turning point followed by a low turning point, both occurring before the post-stress nucleus. A *post-stress rise* was defined as a sequence of a low turning point followed by a high turning point, the latter falling on the post-stress nucleus. An adultlike production of Accent II words in a one-word utterance would contain both a stress fall and a post-stress rise.

The results are given in table 9. In all 5 children, there were more cases of stress fall (a high-low turning point sequence) and more cases of post-stress rise (a low-high turning point sequence) generated for the Accent II target words than for other vocalizations. One-tailed paired t tests comparing the mean proportions between Accent II and other vocalizations revealed a significant difference both in stress fall [$t(4) = 5.76$, $p < 0.005$] and in post-stress rise [$t(4) = 5.26$, $p < 0.005$]. The difference found in the proportion of stress falls identified by MOMEL indicates that children's productions of Accent II words exhibit a better mathematical fit than other vocalizations to the pitch shape that is characteristic of Accent II words, a contour that features a high and low turning point sequence in the stressed syllable.

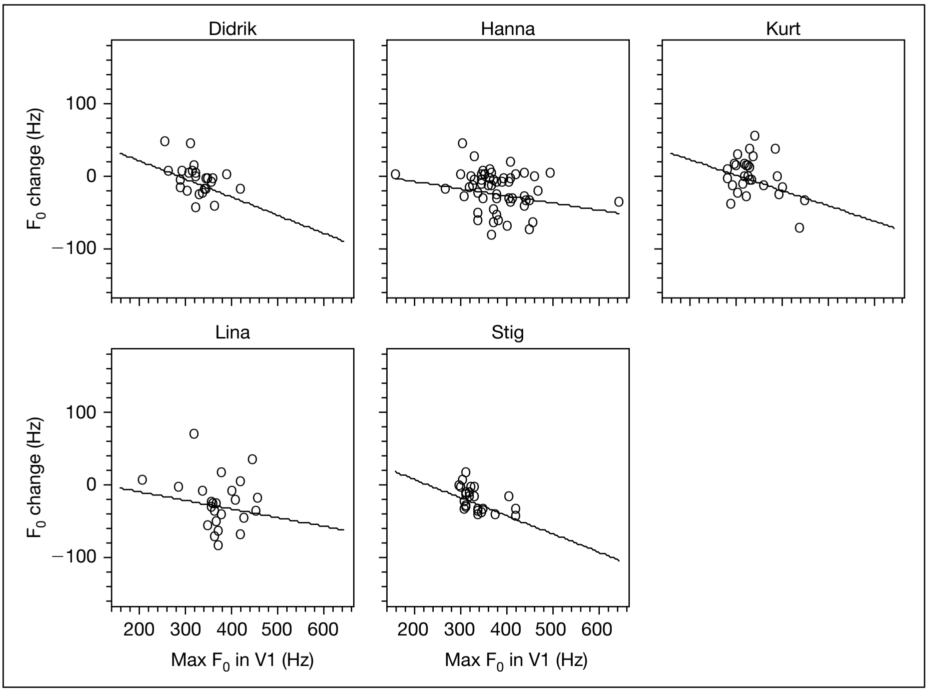

Fig. 7. Maximum F_0 in V1 vs. stressed-syllable F_0 change (Accent II words).

Correlation Analysis

If children's production of Accent II words exhibits the relative stability of the low F_0 point in the stressed syllable – a characteristic in adult speech [Engstrand, 1997] – we should expect to find a negative correlation between the peak F_0 value of the stressed syllable and the stress-syllable F_0 change, and also a positive correlation between the peak F_0 of the post-stress syllable and the post-stress F_0 change. In other words, the stressed syllable change should show a larger drop when the F_0 peak is higher, and the post-stress change should show a larger rise when the F_0 peak of the post-stress syllable is higher.

In figure 7, the stressed syllable F_0 change in the Accent II data is plotted as a function of the F_0 maximum in the stressed syllable. A significant negative correlation was found between the two variables in 4 of the 5 children [Didrik: $r = -0.428$, $n = 24$, $p < 0.05$; Hanna: $r = -0.250$, $n = 55$, $p < 0.05$; Kurt: $r = -0.357$, $n = 28$, $p < 0.05$; Stig: $r = -0.535$, $n = 24$, $p < 0.005$; all one-tailed]. In figure 8, the post-stress F_0 change in Accent II words is plotted as a function of the post-stress F_0 maximum. A significant positive correlation was found between the two variables in all children (Didrik: $r = 0.553$, $n = 24$, $p < 0.005$; Hanna: $r = 0.237$, $n = 55$, $p < 0.05$; Kurt: $r = 0.558$, $n = 28$, $p < 0.001$; Lina: $r = 0.612$, $n = 24$, $p < 0.001$; Stig: $r = 0.823$, $n = 24$, $p < 0.001$; all one-tailed). In accordance with the predictions, the results show that in these children's production of Accent II words, the higher the peak in the stressed syllable, the greater the fall toward the end of the syllable, and the higher the peak of the post-stress syllable, the greater the rise from the end of the stressed-syllable.

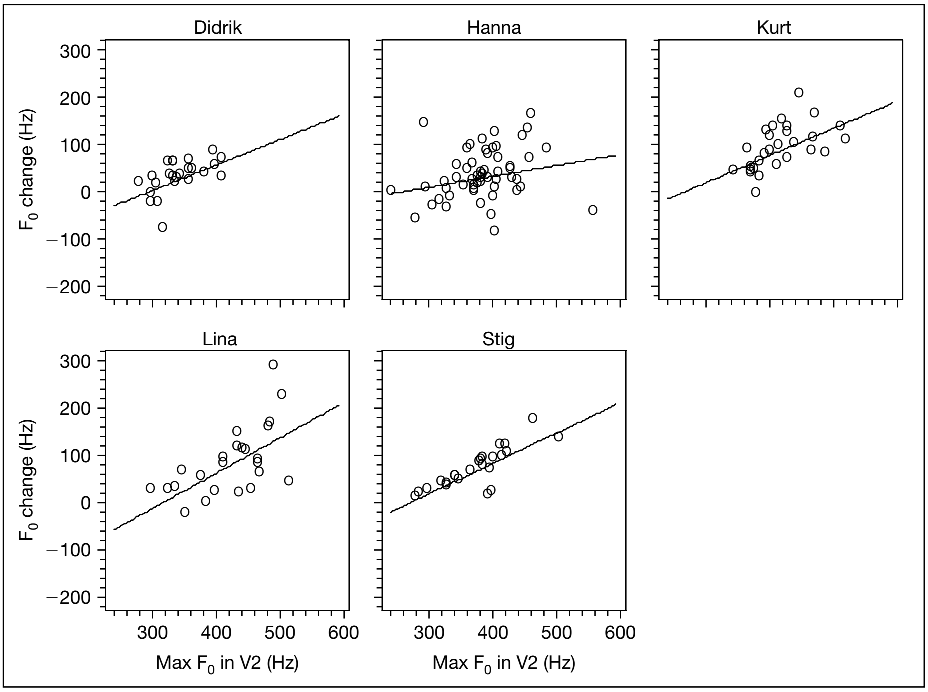

Fig. 8. Maximum F_0 in V2 vs. post-stress F_0 change (Accent II words).

However, similar correlations were also found in other initially stressed disyllabic vocalizations. In 1 child (Didrik), a significant negative correlation was found between the stressed syllable F_0 change and the F_0 maximum in the stressed syllable in non-Accent II vocalizations ($r = -0.385$, $n = 24$, $p < 0.01$; one-tailed). Furthermore, in all but 1 child, other disyllabic vocalizations showed a significant positive correlation between the post-stress syllable F_0 change and the F_0 maximum in the post-stress syllable (Didrik: $r = 0.612$, $n = 24$, $p < 0.001$; Hanna: $r = 0.583$, $n = 22$, $p < 0.005$; Kurt: $r = 0.705$, $n = 9$, $p < 0.05$; Stig: $r = 0.768$, $n = 24$, $p < 0.001$; all one-tailed). These results indicate that the tendency for the rise from the stressed syllable to be greater when the post-syllable stress peak is higher is not a unique characteristic of children's Accent II words. It is more difficult to interpret the negative correlation found between the stressed syllable F_0 change and F_0 peak in the non-Accent II vocalizations of 1 child. We cannot rule out the possibility that some of the vocalizations classified as 'others' in this child's data were actually Accent II targets that were not identified as such (an issue discussed below). It may also be the case that the larger drop in F_0 following a higher pitch peak in the stressed syllable reflects a physiological response to the increased glottal tension that accompanies raised pitch, rather than a controlled pitch modulation aiming at a low pitch target. If so, it is not surprising that the tendency should also be revealed in non-Accent II vocalizations.

In sum, children's Accent II word production reveals correlations between F_0 changes and F_0 peaks, consistent with patterns found in adult speech. However, these patterns did not contrast clearly with other disyllabic vocalizations.

Discussion and Conclusion

The analysis of child Swedish word production in this study yielded four results. First, in words with a sufficiently long stressed syllable, disyllabic Accent II targets differed from other disyllabic vocalizations in that they showed a larger F_0 decline toward the stressed syllable offset and a larger F_0 increase toward the post-stress syllable. Second, stylized pitch models produced by the MOMEL algorithm assigned more high-low turning point sequences within the stressed syllable as well as more low-high turning point sequences leading into the post-stress syllable in Accent II targets. Third, the F_0 peak in the stressed syllable occurred earlier in Accent II than in other vocalizations. Fourth, a negative correlation was found between the F_0 peak of the stressed syllable and the stressed syllable F_0 change. Taken together, these findings show that Accent II words produced by 16- to 18-month-old speakers of Stockholm Swedish exhibit a pitch pattern that uniquely characterizes the adult norm: a falling contour on the stressed syllable that targets a low point around the end of the syllable.

There are a few caveats in this conclusion, however. One limitation of the analysis is that the comparison was made between Accent II words and other vocalizations (a mixture of target Accent I words and other disyllabic vocalizations that could not be identified as word production) rather than between Accent II words and Accent I words. This methodological decision was compelled by the imbalance in the spontaneous speech data – a problem also encountered in previous studies. The outcome of this comparison indicates that the pitch contours of Accent II are in fact differentiated at least from the remaining vocalizations, but it does not fully settle the question of whether the characteristics of pitch contours uncovered here can be strictly understood as a lexical contrast between Accent I and Accent II. In addition, there is a possibility that some of the Accent I targets were misanalysed by the children as Accent II words. Overgeneralization of Accent II to Accent I words has been reported in older children [Plunkett and Strömqvist, 1992; Schmid, 1986], and it cannot be ruled out that the accent status of words in 18-month-old children is subject to a similar tendency. As the classification of target words was based on the adult norm, this could mean that some of the words counted as Accent I might in actuality have been intended as Accent II. If anything, however, this would reduce the contrast between the two categories used in the current study, and the fact that significant differences in the measurements were still found speaks to the robustness of the differentiation.

A second issue in the analysis relates to the effects of rhyme duration on the stressed-syllable F_0 fall. The results indicate, consistently with previous analyses, that the stressed-syllable fall cannot be observed when pitch-bearing segments in the rhyme are too short. By contrast, analyses of adult spontaneous speech in Stockholm Swedish show that there is a tendency for mature speakers to compensate for short rhymes by means of a faster pitch change rate [Engstrand, 1997]. It may be, therefore, that although children's lexical representations of Accent II words contain a pitch fall on the stressed syllable, their phonetic realizations are different from those of adult speakers. This may reflect a qualitative difference in realization strategies, with children leaning toward truncation of contour over rate adjustment in dealing with short pitch-bearing materials – a trade-off akin to the cross-linguistic and cross-dialectal differences found in adult languages [Bannert and Bredvad, 1975; Erikson and Alstermark, 1972; Grabe, 1998; Grabe et al., 2000]. Alternatively, it may be that young children, due to a lack of motor control skills, fail to realize the steeper pitch fall they intend to produce. At this point, it is not clear which of these possibilities (or perhaps another) is the most likely explanation of this divergence.

A third issue, already mentioned in section 4, is the status of the correlations found between the F_0 peaks and the F_0 changes leading toward and away from the putative Accent II pitch valley after the stressed syllable. While these findings are consistent with what we know about adult production of Accent II words, it is not clear whether they unambiguously separate Accent II words from other vocalizations. Addressing this question will require a better understanding of the precise mechanisms that underlie these correlations, which are still under investigation in adult speech.

Despite these remaining issues, the findings in this study provide new evidence that by 18 months of age, Swedish-speaking children systematically mark Accent II words with their unique feature, a falling pitch contour on the stressed syllable. It is important to underscore the implications of this finding for the role of invariance in the development of pitch-related phonology. Although a regular characteristic of Accent II word accent, the falling contour on the stressed syllable does not encompass the whole word. Also, as revealed in the analysis of caretakers' speech, the acoustically most salient pitch event (as measured by the F_0 change) in many Accent II words addressed to children is the rise toward the post-stress syllable, rather than the stress-syllable fall. The fact that the falling contour in Accent II was still incorporated in children's early word production indicates that cross-contextual invariance is used as a cue in early lexical pitch learning even when the regularity is limited to parts of the word and also when there are other more salient patterns in proximity.

On a methodological note, this study highlights the importance of exploring new ways of analysing young children's speech production. Developmental data from naturally occurring spontaneous speech have high qualitative face validity but also suffer from a number of disadvantages such as uncontrollable extraneous factors and sampling density problems. These drawbacks can be overcome to some extent by combining different analytical methods including the type of algorithmic interpolation used in this study. Judicious use of such methods may uncover other characteristics of children's speech abilities that hitherto have gone unnoticed.

Acknowledgements

Thanks are due to Marilyn Vihman for access to the Swedish recordings and useful feedback during the initial stages of the study, Bert Remijsen for assistance with PRAAT scripts, Jenny Minnema and Annelie Ädel for assistance on data analysis, Gösta Bruce and Felix Schaeffler for information on Swedish tonal alignment, Gregory Garretson for editorial help, and Olle Engstrand and an anonymous reviewer for their valuable comments and suggestions.

References

Bannert, R.; Bredvad, A.: Temporal organisation of Swedish tonal accent: the effect of vowel duration. Working Papers, Phonet. Lab., Dep. Gen. Ling., Lund Univ. *10:* 1–36 (1975).

Boersma, P.: Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proc. Univ. Amsterdam Inst. Phonet. Sci. *17:* 97–110 (1993).

Bruce, G.: Swedish word accents in sentence perspective (Gleerup, Lund 1977).

Bruce, G.: How floating is focal accent; in Gregersen, Basbøll, Nordic Prosody IV, pp. 41–49 (Odense University Press, Odense 1987).

Bull, D.; Eilers, R. E.; Oller, D. K.: Infants' discrimination of final syllable fundamental frequency in multisyllabic stimuli. J. acoust. Soc. Am. *77:* 289–295 (1985).

Clarkson, M. G.; Clifton, R. K.: Infant pitch perception: evidence for responding to pitch categories and the missing fundamental. J. acoust. Soc. Am. *77:* 1521–1528 (1985).
Demuth, K.: Problems in the acquisition of grammatical tone. Pap. Rep. Child Lang. Dev. *28:* 81–88 (1989).
Demuth, K.: Issues in the acquisition of the Sesotho tonal system. J. Child Lang. *20:* 275–301 (1993).
Engstrand, O.: Phonetic interpretation of the word accent contrast in Swedish. Phonetica *52:* 171–179 (1995).
Engstrand, O.: Phonetic interpretation of the word accent contrast in Swedish: evidence from spontaneous speech. Phonetica *54:* 61–75 (1997).
Engstrand, O.; Williams, K.; Strömqvist, S.: Acquisition of Swedish tonal word accent contrast. Actes 12ème Congr. Int. Sci. Phonét., Aix-en-Provence 1991, vol. 1, pp. 324–327.
Erbaugh, M. S.: The acquisition of Mandarin; in Slobin, The crosslinguistic study of language acquisition, pp. 373–455 (Erlbaum, Hillsdale 1992).
Erikson, Y.; Alstermark, M.: Fundamental frequency correlates of the grave word accent in Swedish: the effect of vowel duration. Q. Prog. Status Rep., Speech Transm. Lab., KTH, Stockh., No. 2–3, pp. 52–60 (1972).
Gibson, E. J.: Principles of perceptual learning and development (Appleton Century Crofts, New York 1969).
Gogate, L.; Bolzani, L. H.; Betancourt, E. A.: Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. Infancy *9:* 259–288 (2006).
Grabe, E.: Pitch accent realisation in English and German. J. Phonet. *26:* 129–144 (1998).
Grabe, E.; Post, B.; Nolan, F.; Farrar, K.: Pitch accent realisation in four varieties of British English. J. Phonet. *28:* 161–185 (2000).
Hallé, P.; de Boysson-Bardies, B.; Vihman, M. M.: Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by French and Japanese infants. Lang. Speech *34:* 299–318 (1991).
Hirst, D.; Espesser, R.: Automatic modelling of fundamental frequency using a quadratic spline function. Trav. Inst. Phonét. d'Aix *15:* 71–85 (1993).
Hollich, G. J.; Newman, R. S.; Jusczyk, P. W.: Infants use visual information to segment speech in noise. Child Dev. *76:* 598–613 (2005).
Hua, Z.; Dodd, B.: The phonological acquisition of Putonghua (Modern Standard Chinese). J. Child Lang. *27:* 3–42 (2000).
Kadin, G.; Engstrand, O.: Tonal word accents produced by Swedish 18- and 24-month-olds. Proc. FONETIK 2005, pp. 67–70.
Karzon, R. G.; Nicholas, J. G.: Syllabic pitch perception in 2- to 3-month-old infants. Percept. Psychophys. *45:* 10–14 (1989).
Li, C. N.; Thompson, S. A.: The acquisition of tone in Mandarin-speaking children. J. Child Lang. *4:* 185–199 (1977).
Montgomery, C. R.; Clarkson, M. S.: Infants' pitch perception: masking by low- and high-frequency noises. J. acoust. Soc. Am. *102:* 3665–3672 (1997).
Nazzi, T.; Floccia, C.; Bertoncini, J.: Discrimination of pitch contours by neonates. Infant Behavior Dev. *21:* 779–784 (1998).
Ota, M.: The development of lexical pitch accent systems: an autosegmental analysis. Can. J. Ling. *48:* 357–383 (2003).
Peters, A. M.; Strömqvist, S.: The role of prosody in the acquisition of grammatical morphemes; in Morgan, Demuth, Signals to syntax: bootstrapping from speech to grammar in early acquisition, pp. 215–232 (Erlbaum, Mahwah 1996).
Plunkett, K.; Strömqvist, S.: The acquisition of Scandinavian languages; in Slobin, The crosslinguistic study of language acquisition, vol. 3, pp. 457–556 (Erlbaum, Hillsdale 1992).
Prince, C. G.; Hollich, G. J.: Synching infants with models: a perceptual-level model of infant synchrony detection. J. cogn. Sys. Res. *6:* 205–228 (2005).
Schmid, B.: Comparative study of children's and adults' acquisition of tone accents in Swedish. Lang. Learning *36:* 185–210 (1986).
Whalen, D. H.; Levitt, A. G.; Wang, Q.: Intonational differences between the reduplicative babbling of French- and English-learning infants. J. Child Lang. *18:* 501–516 (1991).