Available online at www.sciencedirect.com

ScienceDirect

Journal of Phonetics 36 (2008) 406–422

Journal of Phonetics

www.elsevier.com/locate/phonetics

# Prosody in production at the onset of word use: A cross-linguistic study

Rory A. DePaolis[a,*], Marilyn M. Vihman[b], Sari Kunnari[c]

[a] *Department of Communication Sciences and Disorders, James Madison University, MSC# 4304, Harrisonburg, VA 22807, USA*
[b] *Department of Language and Linguistic Science, University of York, Vanbrugh College, York YO10 5DD, UK*
[c] *Department of Finnish, Information Studies and Logopedics, University of Oulu, PO Box 1000, 90014 Oulu, Finland*

Received 24 October 2005; received in revised form 21 January 2008; accepted 22 January 2008

---

## Abstract

An investigation of the acoustic correlates of prosody in infant disyllabic vocalizations was undertaken on the basis of data from 10 each of American- and Finnish- and five each of French- and Welsh-learning infants at the onset of word use (10–18 months). The 639 disyllables were analyzed acoustically for duration, intensity and fundamental frequency (*f*0). The infants differed in their production patterns with regard to all three acoustic cues, although they agreed in exhibiting high variability in the production of intensity and *f*0. Infants exposed to each of the four languages showed evidence of final syllable lengthening and the production of both first- and second-syllable-stressed or accented disyllables. However, Finnish and Welsh infants produced proportionally more trochaic and iambic patterns, respectively. The use of acoustic prosodic cues at the onset of word use is argued to reflect a combination of biological predispositions and response to prosodic cues that show consistency in the input signal, while a more complete integration of prosody and segmental features seems to require more lexical knowledge or experience.

---

## 1. Introduction

Perception studies of prosody in pre-linguistic infants have been highly productive in recent years. Infants have been shown to be sensitive to the rhythm of their native language at birth (Moon, Cooper, & Fifer, 1993; Ramus, 2002). American infants have been shown to be sensitive both to the predominant stress pattern of English (Jusczyk, Cutler & Redanz, 1993) and to phrase structure (Jusczyk et al., 1992). Prosody also helps infants to segment words (Echols, Crowhurst, & Childers, 1997; Jusczyk, Houston, & Newsome, 1999). Thus, given this level of sophistication in the processing of prosodic information, it might be expected that infant babble would begin to display some of the prosodic patterns of the language the infant is learning. A recent study by Engstrand, Williams, and Lacerda (2003) found this not to be the case. In fact, these authors found that listeners were not able to reliably detect language-specific differences in the babble of 12- or 18-month-old infants.

---

[*] Corresponding author.
*E-mail address:* depaolra@jmu.edu (R.A. DePaolis).

0095-4470/$ - see front matter
doi:10.1016/j.wocn.2008.01.003

Given the large discrepancy between what infants can perceive and what they produce, what are the constraints on prosodic acquisition? The adult-like production of prosody is a difficult task for the pre-linguistic infant, requiring that their perceptual sensitivity to elements of prosody be translated into fine motor adjustments affecting fundamental frequency, timing, and intensity over more than one syllable. Is the major constraint physiological, perceptual, or a combination of the two? Studies of infant production of prosodic cues have so far failed to resolve this issue, as we discuss below. This study examines the emergence of control of the acoustic cues to prosody at the onset of word use to gain insight into the mechanisms that underlie prosodic learning.

Before reviewing the acoustic parameters of prosody it is useful to consider the conceptual framework used here to characterize prosodic development in infants. Word stress or accent is a linguistic construct and thus of limited use when examining pre-linguistic vocalizations. For a pre-linguistic infant, the experience of word level prosody must be signal-based; it will take on linguistic value only gradually, with the onset of word production. In effect, in order to make an empirical study of the acquisition process at this early stage it is necessary to free oneself from the biases of the adult prosodic system. Indeed, the terms ‘iambic’, ‘trochaic’, and ‘spondaic’, used to describe weak–strong, strong–weak, and even-stress patterns in adult languages, respectively, may not be appropriate descriptors for infant vocalizations since the terms imply linguistic knowledge. It may be more appropriate to consider signal-based or acoustic iambs and trochees as more in line with the infants’ experience. This paper will clearly indicate references to the patterns as acoustic iambs and acoustic trochees when they are being considered as signal-based. All other references to trochees or iambs will be in the usual linguistic sense.

### 1.1. *Fundamental frequency*

There is evidence that infants are able to control some of the acoustic correlates of prosody at an early age, although the findings of different studies tend to be discrepant. For example, the control of pitch in imitation has been documented in infants as early as 3 months (Kessen, Levine, & Wendrich, 1979; Papousek & Papousek, 1989). On the other hand, Kent and Murray (1982) found that the predominant pattern measured across 3–9-month-old infants was a falling $f0$ contour, suggesting that infants may not yet be able to counter the natural physiological tendency for subglottal pressure (and thus $f0$) to fall across the course of an utterance. In contrast again, Davis, MacNeilage, Matyear, and Powell (2000) found no significant declination of $f0$ for four infants followed for 3–4 months after the onset of babbling (ages ranging from 0;7–1;2). The use of $f0$ was highly variable and did not differ consistently by syllable in disyllables, but it did affect adults’ judgments as to the stress patterns of disyllables.

Cross-linguistic studies offer a more powerful tool for detecting early learning in the use of prosody. Whalen, Levitt, and Wang (1991) analyzed the reduplicative babbling of five French- and five English-learning pre-linguistic infants. They found that of the 101 vocalizations judged as having rising or falling patterns across the two languages, 75% of the English vocalizations were falling whereas the French vocalizations were equally divided between the two patterns. Whalen et al. (1991) suggested that the differences could be attributed to prosodic differences between the two languages. Vihman and DePaolis (1998) also examined English and French infant vocalizations (again, five infants from each language group), sampled at the ‘4-word point’, defined as the first 30 min recording session in which the infant spontaneously produced four different identifiable words (Vihman, Macken, Miller, Simmons, & Miller, 1985) and with adult judgments as to the location of stress or accent. In contrast to Whalen et al. (1991), they found little evidence that the infants differed in their use of $f0$; however, their measure was mean $f0$ and not rising or falling $f0$ patterns. In addition, reduplicative babble is a well-practiced motoric pattern that might facilitate the overlay of prosodic patterning, while the varied two-syllable utterances in the Vihman and DePaolis study represent more of a challenge to the infant.

There is more evidence for adult-like $f0$ patterns in infants as linguistic competence or knowledge increases. English and French infants at the 25-word point (defined as the first 30 min session in which the infant spontaneously produced 25 different identifiable words) are beginning to use $f0$ to produce the stress or accent patterns consistent with the adult language (Vihman & DePaolis, 1998; Vihman, DePaolis, & Davis, 1998). In addition, at this same developmental point French and Japanese infants differ in their use of $f0$, with French infants producing more rising and Japanese infants more falling *f*0 contours, consistent with their respective adult models (Hallé, Boysson-Bardies, & Vihman, 1991). Furthermore, there is some evidence that infants use varied *f*0 patterns to signal communicative intent. In a study of 10 infants acquiring French, Marcos (1987) found a trend towards more rising *f*0 patterns in repeated requests than in initial requests, and more in initial requests than in labeling activities, although there was no change in use of pitch movements across the age range 1;2–1;10. Finally, Snow (2006) followed six groups of 10 children acquiring American English cross-sectionally between 6 and 23 months of age. He found what appeared to be a regression to narrower *f*0 range at 9–11 months and then a progressive increase in range until about 18 months, when the infants had begun producing adult-like *f*0 contours. Taking all of these studies together, there appears to be limited evidence for the control of *f*0 in the pre-linguistic period but a clear consensus that, by the time of regular production of multiword combinations, *f*0 has become decidedly adult-like.

### *1.2. Duration*

Studies have also documented limited control of aspects of timing in the first year of life. Levitt and Wang (1991) analyzed vocalizations from five French and five American infants between 4 and 13 months. They found that French infants produced more reduplicative vocalizations with longer second syllables (by 20 ms) than did the English infants (54% versus 29%, respectively). In addition, French infants produced more regularly timed non-final syllables over time (as measured by the standard deviation of the duration of these syllables), suggesting that sensitivity to the ambient language increased with time. These results bear replicating, however, since they applied to only 208 reduplicative vocalizations sampled over a long developmental window.

Davis et al. (2000) followed four pre-linguistic infants exposed to English over roughly the same age range (0;7–1;2) and compared the results to measurements from five adults. Unlike the adults, the infants showed no evidence of final syllable lengthening (fsl). Vihman (2001) found that medial stop duration in CVCV vocalizations differed between Finnish and Welsh infants (five in each group) at the 4-word point, in accordance with differences in input frequency: infants exposed to Welsh, in which the medial consonant is lengthened as part of the accentual pattern, produced longer consonants than infants exposed to Finnish, which contrasts long and short medial consonants phonologically. Evidence that Finnish infants were beginning to distinguish singleton and geminate stops in their production at the 4-word point was also reported by Kunnari, Nakai, and Vihman (2001).

Studies involving older infants, although inconsistent in their findings, suggest the emergence of adult-like duration patterns. Boysson-Bardies, Bacri, Sagart, and Poizat (1981) found an increase in syllable durations from the early to the later sessions of a child followed from 18 to 20 months of age. Since this lengthening was accompanied by an increase in the number of words or lexical expressions used in the later sessions, these authors suggested that their results may have reflected the effort of purposeful articulation, which requires sufficient motoric coordination to overlay linguistic intent on the acoustic signal. This suggests that there may be physiological constraints as late as 18–20 months, although in this case the constraint could also be ascribed to a lack of the cognitive resources needed to integrate segmental and suprasegmental targets. Robb and Tyler (1995) reported what appear to be conflicting data, however. Following seven English-learning children from 8–26 months, they found a decrease in word duration but no change in non-word duration. Robb and Tyler's (1995) data could suggest a maturational effect on word duration such as Boysson-Bardies et al. (1981) proposed, but one that begins already with the onset of word use. The fact that Boysson-Bardies et al. (1981) found an increase in overall vocalization duration while Robb and Tyler found a decrease may be due to the extent of the period covered by the studies (2 versus 18 months) and the number of infants involved in each (1 versus 7). On the other hand, the conflicting results may suggest a more complex developmental story than either study reveals.

Three cross-linguistic studies contrasting the 4- and 25-word point found evidence for adult-like duration only at the latter developmental level. Vihman and DePaolis (1998) found little evidence to distinguish French and English infants' vowel durations at the 4-word point, although these same infants were producing more typically adult-like patterns by the 25-word point. Vihman, Nakai, and DePaolis (2006) compared the duration of the initial vowel, medial consonant, and final vowel in infant (C)VCV vocalizations to comparable tokens elicited from adult speakers. They found that American English, French, and Welsh infants (five per group) were not yet producing the rhythmic properties of adult tokens at the 4-word point but were doing so by the 25-word point. Kunnari et al. (2001) found that Finnish infants are much closer to the adult models in their production of geminates versus singleton stops at the 25-word point than at the 4-word point. In addition, Snow (1994) and Konopczysnki (1993) found evidence of a consistent pattern of fsl in English and French (respectively) when the infants were either at or very close to beginning combinatorial speech. Taken together these studies suggest that the control of duration may involve several stages of learning, although there is converging evidence for limited early ambient language effects on the timing of pre-linguistic speech and clear evidence for adult-like control of timing just prior to combinatorial speech.

### 1.3. *Intensity*

Davis et al. (2000) found that in asymmetrically stressed disyllables their American infant participants used significantly less intensity on the second than on the first syllable. They also found that intensity was the only acoustic correlate of stress that appeared to resemble their adult data, suggesting that these results might reflect the physiological effect of decreasing intensity over the length of an utterance (Lieberman, 1985). Levitt (1993) examined intensity patterns in French and English infants between 7 and 11 months of age. She compared adult and infant data and found no evidence that infants' production of intensity was close to the adult model. Finally, a comparison of the English and French infants at the 4-word point by Vihman and DePaolis (1998) also found little evidence that infants in the two languages differed in their use of intensity, although by the 25-word point intensity was consistently higher in syllables judged stressed or accented by adult listeners (Vihman et al., 1998).

In summary, unlike the findings for $f0$ and duration there is little converging evidence for the adult-like use of intensity as infants begin to use words. This is partially due to the smaller number of studies that have investigated intensity. It may also reflect the difficulty in reliably measuring intensity on what amounts to a highly mobile participant group. Table-mounted microphones do a poor job of controlling microphone-to-mouth distance while body-mounted microphones are only slightly better.

### 1.4. *The acquisition of prosody*

The studies described above provide evidence for only limited control of prosody in the pre-linguistic period. They also suggest that, as the infant nears the end of the second year of life, the prosody of vocal productions become increasingly adult-like, coinciding with the increase in word production (cf. Vihman, Macken, Millar, Simmons, & Millar, 1985, who report a mean of 60% babble, with a range of 17–77%, in 9 infants acquiring English at about 16 months of age). Engstrand, Williams, and Lacerda (2003) found little remaining babble (or unintelligible language use) in their 18-month-old participants. They also found that although adult listeners could not distinguish American from Swedish infants' babble at 12 and 18 months of age, the listeners did show evidence of being able to determine the language the infant was learning at the later age when utterances considered ambiguous as between words and babble were included. Kehoe, Stoel-Gammon, and Buder (1995) also found that 18-month-olds were adult-like in the application of $f0$, intensity, and duration to indicate stress in elicited trochaic words. In contrast, Pollack, Brammer, and Hageman (1993) found that words elicited from 2-year-olds were highly variable in correct stress placement. Taken together, however, most studies support the notion that the integration of adult-like prosody into infant vocalizations occurs after the acquisition of a 'critical mass' of words.

In short, it is still unclear how early infants' vocalizations are influenced by the adult prosodic system. Some of the conflicting results may arise from the fact that studies have been based on a variety of languages over a broad range of chronological and developmental points. Studies that use age-matched samples and/or data sampled across many months inevitably suffer from high variability that is hard to ascribe to any one aspect of prosodic acquisition. For example, changes in production could be due to a loosening of physiological constraints, or to growing awareness of prosodic or segmental contrasts, or to the nature of the targets attempted as the infant begins producing words.

A number of studies suggest that infants have the ability to control *f*0, intensity and some aspects of timing as they begin to produce their first words. This then should be an ideal period in which to look for evidence of ambient language effects, since as the lexicon expands prosodic control may become more variable in response to the segmental demands of an emerging phonology (see, for example, Boysson-Bardies et al., 1981; Klein, 1984; Kehoe, Stoel-Gammon, & Buder, 1995; Robb & Tyler, 1995 or Jacob's variable stress productions of *cookie* and *cracker* in Menn & Matthei, 1992).

The key to understanding why there is a lack of consistent evidence that pre-linguistic infants can control elements of prosody may be the variability of those cues in the input signal. Assuming that infant speech gestures can be guided by the input, as suggested by computer modeling (Callan, Kent, Guenther, & Vorperian, 2000; Guenther, Hampson, & Johnson, 1998; Westermann & Miranda, 2004), an invariant or, at the very least, a consistent cue is necessary for the infant to register, attempt to produce, and successfully approximate it. In the terminology of Guenther et al. (1998) a "teaching signal" in the form of feedback from the auditory system is required to guide the articulatory system to model the input. English presents a variable model with respect to *f*0 and intensity since two-syllable utterances vary not only in stress assignment but also in the use of *f*0 and intensity to cue stress (Fry, 1955; Lieberman, 1960). Even worse, infant-directed speech proves to be much more variable than adult-directed speech (Vihman et al., 2006). In effect, English provides a poor teaching signal. Other languages may provide a better model. Thus, a developmentally matched cross-linguistic study using a single set of measures may enable us to determine whether the lack of consistency in infant output is due to a lack of consistency in the input, the inability of the infant to control the motor system, or some combination of the two.

The purpose of this study is to use cross-linguistic comparisons to determine whether there is evidence at the onset of word use that infants diverge by language in their use of *f*0, intensity or duration. A comparison of infants' use of acoustic cues to prosody across languages should help to determine whether the primary difficulties are inherently physiological or due to the prosodic complexity of the ambient language. Any similarities across languages will be taken to suggest the presence of physiological constraints, especially if these similarities are connected to known biological tendencies. On the other hand, differences across languages will be taken to suggest infants' sensitivity to and emergent control of ambient language effects, especially if there is some congruence with the acoustic cues of the language the infants are learning.

## 2. Method

### *2.1. Participants*

Ten infants each learning American English or Finnish and five infants each learning French or Welsh participated in this study. The children's ages ranged from 10 to 18 months. Children acquiring each of the languages were recorded on audio and video in their homes, in 30 min free play sessions with their mothers. The children acquiring English were recorded in the area around Palo Alto, California, the French children in Paris, France, the Finnish children in Oulu, Finland, and the Welsh children in the area around Bangor, Wales. Data regarding the participants' ages and the number of disyllables analyzed are presented in Table 1. The French and American data were collected with analog recording equipment (Uher 4000 reel to reel

Table 1
Participant ages and number of vocalizations

| Language | No. of infants | Mean age | Age range | No. of vocalizations |
|---|---|---|---|---|
| English | 10 | 13 | 0;10.15–1;3.0 | 190 |
| Finnish | 10 | 14 | 0;11.1–1;6.0 | 142 |
| French | 5 | 11.5 | 0;10.10–1;1.2 | 127 |
| Welsh | 5 | 13 | 1;0.4–1;2.5 | 180 |

recorder). The Welsh and Finish audio data were recorded digitally. All recordings were made with electret microphones. The French, American, and Welsh data were collected with a microphone sewn into the front of a vest worn by the infants. The Finnish data were collected with a microphone placed directly in front of the infant. More information is available in Vihman et al. (1985) for the English sample, in Boysson-Bardies and Vihman (1991) for the French, in Kunnari (2000) for the Finnish, and in Vihman (2000) for the Welsh.

### 2.2. Selection criteria

All of the disyllables were taken from the infants' 4-word point, defined as the first 30 min session in which the infant spontaneously produced four different identifiable words (Vihman et al., 1985). English, French and Finnish disyllables were extracted from audiotapes and digitized to 16 bits at a sampling rate of 22.05 kHz. The Welsh disyllables were extracted from digital recordings sampled at 44.1 kHz. Disyllables were considered candidates for inclusion if they were separated from surrounding utterances by at least 400 ms (following Branigan, 1979). Of these disyllables, those considered prosodically bound to surrounding speech were excluded. Disyllables separated by less than 400 ms were included if there were clear prosodic breaks with the surrounding speech (such as a clear inhalation to start a new breath group). Disyllables were included if they minimally contained two open (vocalic) phases separated by a closed (consonantal) phase. Words were separated from babble following Vihman and McCune (1994). Disyllables that showed excessive shifts of register, excessive vocal effort, creaky voice, or whisper were also excluded. In addition, if a disyllable was repeated multiple times, only the first three tokens were included in the analysis, although repeated instances of that disyllable later in a session were included. The reasons for excluding disyllables were highly variably across infants. For example, one infant in the English sample had 62 candidate disyllables and of that number 33 were excluded for the following reasons: noise (4), interfering talk (3), prosodically bound to surrounding speech (9), singing (2), whisper (2), unsegmentable (8), and excessive shifts of $f0$ (5).

### 2.3. Acoustic measurements

Measurements on 639 isolated disyllabic vocalizations were based upon the segmented portion of the vowels and included the intensity difference in decibels between the rms voltage of syllable 1 and the rms voltage of syllable 2, semitone difference of the semitones between the $f0$ averages of syllable 2 and syllable 1, and vowel duration ratio (syllable 2 divided by syllable 1). Duration measurements were made using narrow and wide band spectrograms, pitch and intensity trace, and both overall and expanded time displays of amplitude. The guidelines used for segmentation of each consonant–vowel sequence are included in Appendix I; they are based in part on a survey of previous work segmenting speech (for example, Konefal, Fokes, & Bond, 1982; Oller & Smith, 1977; Peterson & Lehiste, 1960; Robb & Saxman, 1990; Snow, 1994; Umeda, 1975).

The rms amplitude and mean $f0$ values that were used to compute decibels and semitone differences, respectively, were based upon the entire segmented vowel. The computation of the decibel difference between syllables was performed using Eq. (1) below:

$$\mathrm{dB} = 20\log_{10}\frac{V_{\mathrm{rms2}}}{V_{\mathrm{rms1}}}. \tag{1}$$

In both intensity and $f0$ computations a negative value indicates higher intensity or $f0$ on the first syllable. Eq. (2) was used to compute the semitone difference in $f0$ for each vowel.

$$\mathrm{Semitone} = 39.86\log_{10}\frac{f0_2}{f0_1}. \tag{2}$$

All $f0$ measurements were made using either a peak picking or autocorrelation algorithm checked against a narrow-band spectrogram for anomalies such as incorrect identification of F1 as $f0$. Mean $f0$ measures were used instead of qualitative $f0$ contours (i.e. rising, falling or flat) in order to compare differences across syllables quantitatively.

Intra-measurer reliability was assessed by the first author for 10% of the French and English disyllables chosen randomly (5% of the total number of utterances across all languages). The French and English measurements were chosen since they were made over 10 years previous to this study and there was little possibility that the utterances would be remembered. Using the guidelines in Appendix I, average measurement differences were 13.7 ms for duration, 7.6 Hz for *f*0, and .4 dB for intensity.

## 3. Results

Descriptive statistics for the three acoustic parameters of duration, fundamental frequency, and intensity are presented for each language in Table 2. Words and babble are separated in this table. The most notable difference between the languages is that the duration ratios are smaller for Finnish and Welsh than for either American English or French. Variability is also much smaller for the Welsh and Finnish duration ratios. Additionally, the Finnish infants separate from the other three languages with higher intensity and *f*0 on the first syllable. The Welsh infants also use much higher intensity on the second syllable in words than babble. The only similarity across all languages is the great extent of variability for intensity and *f*0. The English and French mean *f*0 and intensity differences between syllable one and two are close to zero, presumably reflecting what would be perceived as equal stress or accent on the two syllables.

A multivariate analysis of variance was performed with the independent variables Language and Word, and the dependent variables duration ratio, semitone difference, and intensity difference. Since the vocalizations from each child were not independent, child was nested within language and word, and the estimate of variation due to child was not used for testing purposes. There was not a significant interaction between language and word, $F(3,555) = 1.849$, $p = .137$, $\eta^2 = .01$. There was a significant effect of language for the intensity difference, $F(3,555) = 2.960$, $p = .032$, with a partial $\eta^2 = .016$ and duration ratio, $F(3,555) = 4.640$, $p = .003$, with a partial $\eta^2 = .024$. There was no Language effect for semitone difference, $F(3,555) = 1.901$, $p = .128$, with a partial $\eta^2 = .010$. There was no effect of word for intensity difference, $F(1,555) = .188$, $p = .665$, with a partial $\eta^2 < .001$, duration ratio, $F(1,555) = .1031$, $p = .310$, with a partial $\eta^2 = .002$ or semitone difference, $F(1,555) = .029$, $p = .864$, with a partial $\eta^2 < .001$. Post Hoc Scheffé multiple comparisons yielded a significant difference between duration ratios for English and Finnish ($p = .001$) and French and Finnish ($p = .002$). The intensity differences were English and Finnish ($p = .006$) and Welsh and Finnish ($p < .001$). Surprisingly, semitone difference was significant between French and Finnish ($p = .042$) and marginally between English and Finnish ($p = .077$) in spite of the lack of a significant main effect. This is consistent with small partial $\eta^2$ across all tests, indicating that there is considerable variability that is not

Table 2
Means and standard deviations for intensity difference, semitone difference, and duration ratio for all four languages

| | dB ($\mu$) | dB (SD) | Dur. ($\mu$) | Dur. (SD) | *f*0 ($\mu$) | *f*0 (SD) |
|---|---|---|---|---|---|---|
| English | | | | | | |
| Babble | .65 | 6.13 | 2.19 | 2.10 | −.1 | 3.6 |
| Word | −.81 | 4.88 | 2.30 | 2.90 | .1 | 2.9 |
| French | | | | | | |
| Babble | −.52 | 5.10 | 2.17 | 2.73 | .2 | 3.7 |
| Word | −.24 | 5.30 | 2.51 | 2.50 | .0 | 3.8 |
| Finnish | | | | | | |
| Babble | −1.48 | 4.77 | 1.46 | .80 | −.9 | 3.0 |
| Word | −2.50 | 6.30 | 1.25 | .74 | −1.1 | 3.5 |
| Welsh | | | | | | |
| Babble | .01 | 6.26 | 1.79 | 1.33 | −.1 | 3.9 |
| Word | 2.94 | 5.37 | 1.43 | .69 | .3 | 3.9 |

Data are averaged across all vocalizations and words and babble are separated. A duration ratio larger than one indicates a longer second syllable. A negative intensity difference or semitone difference indicates that the value was higher on the first syllable.

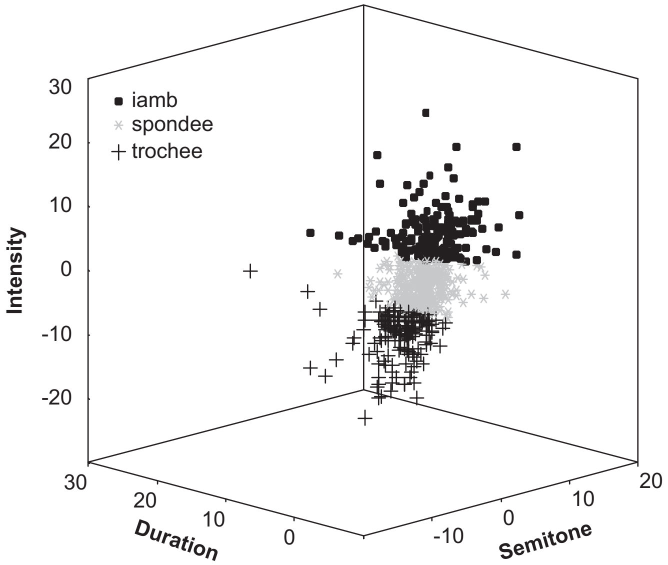

Fig. 1. Scatter plot of the acoustic classification by *K*-means cluster analysis of all disyllables.

accounted for by the independent variable language. This high variability is also consonant with a lack of consistency in the use of prosodic cues in all four languages.

The high variability and low effect sizes coupled with the French and American infants' average *f*0 difference values being close to zero suggest that infants may be producing multiple stress or accentual patterns (acoustic iambs and trochees). Since all of the languages have some type of variable prosodic patterning this seems plausible. Recent studies of infants learning English and French also support this notion (Davis et al., 2000; Vihman & DePaolis, 1998; Vihman et al., 1998, 2006). *K*-means cluster analysis (SPSS, 2005) was performed on all vocalizations in an attempt to separate iambs and trochees acoustically. No initial cluster centers were specified. In order to accommodate the infants' production of even-stressed disyllables as identified in previous studies (Davis et al., 2000; Vihman & DePaolis, 1998; Vihman et al., 1998) three categories were specified in the analysis, which was performed with 17 iterations and a final change in cluster centers of zero. The cluster analysis formed three groups that closely matched the acoustic profile of trochees, iambs and spondees in adult speech. The cluster centers for each group were: trochees (cluster centers of −7.04 dB, −3.3 semitones and a duration ratio of 1.68), iambs (cluster centers of 7.24 dB, 2.6 semitones, and a duration ratio of 2.59) and even-stressed categories (cluster centers of −.3 dB, .1 semitone, and a duration ratio of 1.64). Fig. 1 shows the distribution of acoustically classified trochees, iambs, and spondees for all of the infants.

Fig. 2 shows the number of acoustic trochees, iambs, and spondees identified for each language separating babble and words (Fig. 2a and b, respectively). Acoustic measures of stress assignment were used since the approach in this paper was essentially to treat the infants as signal processors with little knowledge of the ambient prosodic patterns. No attempt was made to match adult stress judgments to the infants' production patterns. Assessing acoustic patterns using adult linguistic knowledge of stress creates a logical dilemma since the acoustic patterns that represent a judged trochee in English (for example) are present due to the very judgments that are used to assess them. For a point of reference, the cluster analysis performed above categorized 71.4% of the utterances in the same way that English and French listeners judged the English and French utterances respectively (see Vihman et al., 1998 for more information on these stress judgments). The cluster analysis suggests that the acoustic equivalent of both iambs and trochees is present in early vocal productions in each language. In addition, the proportion of acoustic spondees is nearly identical in all four languages and constitutes the most common pattern. A chi-square test was carried out on each language to determine whether the proportion of iambs and trochees differed from 50%. The even-stressed

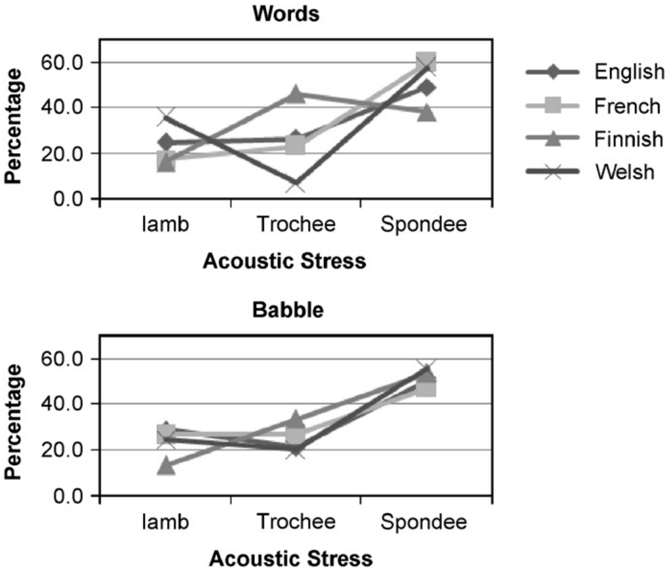

Fig. 2. Percentage of disyllables classified acoustically as iambic, trochaic, or spondaic for (a) words and (b) babble.

disyllables were not included in the analysis. Only Finnish babble ($\chi^2[1, n = 36] = 7.11$; $p < .01$) and words ($\chi^2[1, n = 39] = 9.26$; $p < .005$), and Welsh words ($\chi^2[1, n = 19] = 8.89$; $p < .005$) differed significantly from an equal production of the two stress patterns. The Welsh infants produced significantly more iambs while the Finnish infants produced significantly more trochees. Care should be taken when interpreting the Welsh finding however, since one infant produced 40 of the 61 Welsh iambs (see Appendix II for a complete tabulation of language, infant, and stress/accent patterns).

## 4. Discussion

The babbling of 30 infants exposed to four different languages at a single developmental point was examined for evidence of language-specific influences. As the languages represent a wide range of prosodic structures it is not surprising that the only similarity across all four languages was variability in the production of intensity and $f0$. Table 3 summarizes the descriptive differences and similarities. The terms 'iamb' and 'trochee' are meant to simplify interpretation of the data and are used as purely acoustic descriptors, despite the fact that they may not be ideal terms to use for all four of the languages. In the following discussion intensity and $f0$ will be addressed separately from duration.

### *4.1. Intensity and f0*

The descriptive statistics reveal that intensity and $f0$ vary widely in each language. Overall there is a moderate correlation between intensity and $f0$ for English, French and Welsh ($r = .512$, $.533$, $.522$, respectively, $p < .01$), suggesting that the infants are not yet controlling the two independently. It is possible that at the onset of word use, before the infant is able to integrate segmental planning with prosody, perceived stress or accent is equated to increased vocal effort, leading to higher combined intensity and fundamental frequency. The natural tendency for pitch and intensity to fall over the course of an utterance appears to be easily counteracted by voluntary control and exploration of $f0$ and intensity. Infants' control of pitch at an early age (Papousek & Papousek, 1989; Whalen et al., 1991) and the later difference in individual infants' preference for prosodic patterns (Vihman & DePaolis, 1998) support this notion. The Finnish correlation between $f0$ and intensity is low ($r = .35$, $p < .01$) and was significantly different from that of the American, French, and Welsh ($p < .0341$), suggesting that the Finnish infants are beginning to control $f0$ and intensity

Table 3
Interpretation of the descriptive statistics

| | English | Finnish | French | Welsh |
|---|---|---|---|---|
| Amplitude | | | | |
| Mean | Even | Trochee | Even | Iamb |
| Variability | *High* | *High* | *High* | *High* |
| *f*0 | | | | |
| Mean | Even | Trochee | Even | Iamb |
| Variability | *High* | *High* | *High* | *High* |
| Duration | | | | |
| Mean | + fsl | fsl | + fsl | fsl |
| Variability | High | Low | High | Low |

Italics indicate the only acoustic correlates of prosody that are consistent across all four languages. Final syllable lengthening is not italicized since two languages exhibit it to a high degree (+ fsl) while two show only a moderate degree.

independently. The low variability in adult Finnish duration ratios coupled with fixed word stress may result in a model transparent enough for the infants to settle on a rhythmic production pattern already at this early developmental point. This in turn might enable the infant to become sensitive to *f*0 and intensity as separate cues. There are many mechanisms that could lead the infant to discover this distinction. One possibility is that the extra-long medial consonants that the Finnish infants are producing at this age (Kunnari et al., 2001; Vihman & Velleman, 2000) result in decreased subglottal pressure on the second syllable (Lieberman, 1985), leading the infant to raise pitch by increasing laryngeal tension without increasing intensity. However the infant may manage to disentangle *f*0 and intensity, it is not surprising that the vocalizations of infants learning the language with the most consistent acoustic pattern should diverge from the other three languages in this respect.

### 4.2. *Duration*

Of the three acoustic cues to prosody, duration demonstrates the clearest difference across the vocalizations of infants exposed to different languages. The data suggest that if the language the infant is learning prominently exhibits fsl, as does French, the infants exaggerate this feature and produce overly long second syllables. The data also suggest that language such as Finnish, with strong duration cues to stressed initial syllables in disyllabic words (Iivonen, 1998), lead to infant production of far fewer extra-long second syllables. Since the Finnish (and Welsh) infants exhibited considerably lower variability than the French and English infants, it can be said that these infants have settled on a more stable model of production.

### 4.3. *Prosodic acquisition at the onset of word use*

The high degree of variability of *f*0 and intensity in all four languages and of duration in English and French brings to mind the variable uses of these parameters to signal stress in adult English, in which cues can be combined on the stressed syllable or varied across syllables, causing the listener to weigh the cues perceptually to determine stress (Fry, 1958). At first glance this degree of variability seems to be lacking in the other three languages. Finnish and Welsh are considered trochaic while French is considered iambic. However, when the cues are examined from the perspective of a pre-linguistic infant, i.e., a listener lacking any access to lexical or other ‘top-down’ knowledge, a different picture emerges.

Although Finnish has trochaic stress patterns in two-syllable words it also makes use of contrastive segment length on both consonants and vowels (although consonants can contrast only word-medially). So Finnish is consistent in the application of word stress, but the placement of durational and tonal cues to stress or accent is dependent upon the moraic structure of the word as a whole and is distributed across the two syllables of a disyllabic word (Suomi, Toivanen, & Ylitalo, 2003; Suomi & Ylitalo, 2004). Finnish also has secondary stress, which can in certain instances fall on the second syllable (Iivonen, 1998).

French is often referred to as being mainly iambic due to its phrase-final accent and its even stress within phrase boundaries. However, recent studies of French have revealed a more complex story. Although primary stress occurs on the last syllable of a phrase and is signaled by duration and pitch cues, French also has secondary stress, which tends to occur on the first or second syllable of a phrase and is signaled by a pitch accent (Di Cristo, 1998; Jun & Fougeron, 2002). Even tonal cues to primary stress at the end of a phrase tend to occur across the final two syllables in the form of a pitch rise and fall. In addition, the unstressed syllable following the stressed syllable is shortened relative to preceding syllables (Di Cristo, 1998). Disyllables are thus iambic when accented (Di Cristo, 1998), but in running speech they are acoustically more ambiguous.

Welsh stress has been classified as trochaic (penultimate prominence), but the unstressed syllable is typically longer, contains a pitch increase or decrease, and has a higher amplitude envelope (but lower peak amplitude) than the stressed syllable (Williams, 1986). Welsh is thus more iambic than trochaic on purely acoustic grounds, and increased salience and stress need not coincide (Williams, 1985). It is therefore not surprising that pre-linguistic infants should respond to the iambic acoustic cues. Another aspect of Welsh stress that may affect infant production and lead to variability is the prevalence of English in the bilingual community, which results in exposure to English even for infants learning only Welsh in the home. (See Vihman, Thierry, Lum, Keren-Portnoy, & Martin, 2007, for a discussion of the effects on early word recognition of Welsh learning infants' exposure to English.)

Added to the variability in the adult implementation of prosody is emphatic and contrastive stress, which occurs in all four languages and which may involve lengthening, large pitch movements, and increased intensity at any position within an utterance. The gap between linguistic structure as classically described and the actual input signal as experienced by infants is highlighted by the fact that the prosodic cues in infant-directed speech are far more variable than those in isolated disyllables elicited from adults (Vihman et al., 2006). The consistent placement of intensity and $f0$ may require more guidance than is available to the babbling infant. This skill may require the increased attention to segmental aspects of new words that comes online in the second year for the infant to begin to disentangle language-specific uses of $f0$ and intensity.

Studies by Snow (1994, 2006) and Behrens and Gut (2005) support the idea that prosodic competence requires linguistic knowledge. Snow (1994) found that infants' rate of $f0$ change is related to the duration of the final syllable. The rate of $f0$ change becomes more independent of duration, as in adult speech, at the onset of word combinations. Similarly, Snow (2006) found that intonation development followed a U-shaped pattern, with the 6–8-months-olds producing a wide range of intonation patterns, followed by a restricted $f0$ range emerging at 9–11 months of age and then a return to an expansion of the $f0$ range at 18–20 months. Again, the later adult-like $f0$ ranges in this study at 18–20 months of age coincide with the increase in word use preceding word combinations. Behrens and Gut (2005) examined a German child's early combinations and found that intensity, $f0$, and duration patterns differed across different types of two-word utterances. For example, they found that noun + noun combinations showed no sign of increasing prosodic integration (i.e. movement towards the adult system), while noun + particle combinations did show such signs, suggesting again that it is the linguistic advances that trigger the integration of adult-like prosodic cues into segmental patterning.

Finally, this study raises questions about the infants' ability to perceive and produce elements of prosodic structure in their pre-linguistic utterances. The participants in this study are just beginning to exhibit language-specific tendencies. Interestingly, all infants are producing acoustically trochaic, iambic, and spondaic patterns (consistent with findings by Davis et al., 2000 and Vihman et al., 1998), with only the Finnish and Welsh infants showing signs of a preferred pattern. The nearly identical proportion of spondees in each language suggests that this may be a default pattern that is altered as infants begin to integrate prosodic cues into their production. At the time that the data were collected the infants were just beginning to produce words, so it is doubtful that these vocalizations represent explicit attempts to model the prosody of the adult language. Instead, it is more likely that the infants are producing patterns that have been learned implicitly (see the studies demonstrating distributional learning for 8-month-olds, such as Maye, Werker, & Gerken, 2002; Saffran, Aslin, & Newport, 1996). As word production begins to overtake babble in the infants' vocalizations, stress or accent may become a more integral part of how a word or phrase is represented in the infant's developing lexicon. What is surprising is that infants with very little linguistic knowledge have the ability to extract some acoustic patterns from what amounts to a very noisy input. However, the influence of prosody on infant output as early as 10 months is consonant with the well-documented sensitivity of pre-linguistic infants to prosodic patterns (see Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003 for a brief review).

What then is the state of infant production of prosodic cues at the onset of word use? Early attempts at producing prosodic patterning may reflect biological predispositions, such as longer duration of the final syllable and a decrease in pitch and intensity over the course of the breath group (Kent & Murray, 1982; Lieberman, 1986), coupled with infant attempts to increase vocal effort in a stress-like manner (see Vihman et al., 1998, for an outline of biological predispositions and ambient language effects for English and French). In this case, intensity and $f0$ would tend to co-occur as the infant increases vocal effort to highlight a portion of an utterance. An interaction between the infant's own productions and adult patterns that are maximally distinctive, such as duration in Finnish, may initially make the infant aware of the potential match between certain aspects of stress in adult production and acoustic effects that he or she is able to create. DePaolis (2006) found such a match between well-practiced consonants (vocal motor schemes or VMS) and increased interest in passages with an individual infant's VMS over matched passages lacking that infant's VMS. This evidence for an articulatory filter working at the segmental level (Vihman, 1996, p. 142) makes it plausible that the same mechanism could be at work at the suprasegmental level. That a suprasegmental articulatory filter would work first at the level of rhythm is supported by Ramus (2002), who found that newborns could discriminate languages based upon rhythmic properties even when most of the intonational cues were removed. He suggested that rhythm might be a 'perceptual cornerstone' for the infant (cf. also Cutler & Mehler, 1993), supporting the finding here that duration was the cue that differed the most across the four languages. Settling upon the stable production of specific prosodic cues may facilitate awareness of word stress or accent as comprised of controllable elements that can vary independently, as seems to be the case with Finnish infants in relation to $f0$ and intensity. However, more finely tuned use of prosody may require a level of attention to linguistic detail that begins to be possible only as word production becomes well established.

## Acknowledgments

The authors would like to thank Satsuki Nakai for help with the Welsh data collection and analysis. This work was partially supported by Grant # R000237087 from the Economic and Social Research Council (UK), a European Commission Marie Curie Incoming International Fellowship, an educational leave from the College of Integrated Science and Technology at James Madison University, and from The Academy of Finland.

## Appendix I. Segmentation guidelines

Numbers indicate importance of each cue to segmentation. Two cues with equal numbers were weighted equally.

### Stops

*CV Boundary*

*Voiceless*

- Identify burst in the time plot
- Find first glottal pulse and place marker on first positive peak

*Voiced*

- Identify burst in the wide-band (WB) spectrograph
- Place marker after burst at the onset of the following:

1. First identifiable glottal pulse (positive peak)
2. Formation of F1 (if pre-voicing present, then the change in amplitude of F1)
3. Onset of harmonic structure in the narrow-band (NB) spectrograph
4. Rapid change in slope of the intensity curve
5. Onset of periodic striations in the WB spectrograph

- For pre-voiced stops, place marker immediately after burst; if the burst is unidentifiable use formants as a guide.

*VC Boundary*

*Voiceless and voiced*

- Identify the point of closure by the following:

1. Abrupt change in the amplitude of F1
2. Sharp reduction in the upper harmonics in the NB spectrograph
3. Loss of F2 and F3 structure
4. Abrupt change in slope of the intensity curve

### Nasals

The appearance/disappearance of harmonics often occurs before/after the abrupt change in F1/nasal murmur amplitude. This is likely due to the difference in time between the velar and oral closure, or the incomplete closure of the velum during a nasal utterance.

*CV Boundary*

- Identify boundary by noting the onset of the following:

1. Abrupt change in amplitude from nasal murmur to F1
2. Formation of higher harmonics in NB spectrograph
3. Onset of rising slope in intensity curve

*VC Boundary*

- Identify boundary by noting the onset of the following:

1. Abrupt change in amplitude from F1 to nasal murmur
2. Reduction of higher harmonics in NB spectrograph
3. Onset of falling slope in intensity curve

### Semi-vowels

Segmentation of semi-vowel boundaries requires the ability to identify F2 or F3 transitions. If the transition was not identifiable (for example /ji/ or poor resolution of the spectrograph) the utterance was not segmented.

*CV Boundary*

- Identify beginning of transition into vowel and beginning of steady state portion of vowel (identification was facilitated by noting the 'knee' or change in slope of the formant. The beginning or end of the transition was assumed to be the 'bend' in the knee)
- The marker was placed between these two points ('splitting the transition')

*VC Boundary*

- Identify the beginning of the transition out of the vowel and the beginning of the steady state portion of the glide (bend in the knee as above)
- The marker was placed between these two points

*Special case:* initial semi-vowels with no clear steady portion were simply halved from the beginning of the consonant to the onset of the steady state portion of the vowel.

*Special case:* medial semi-vowels with no clear static portion were quartered, with the initial quarter counted as part of the initial vowel and the final quarter counted towards the final vowel.

**Fricatives**

*CV Boundary*

*Voiceless and voiced*

- The onset of the vowel was identified by the following:

1. Onset of F1; if identifiable, the first positive glottal pulse was used for voiceless fricatives
2. Formation of harmonics in NB spectrograph
3. Onset of higher formant structure
4. Marked increase in slope of the intensity curve

*VC Boundary*

*Voiceless and voiced*

- The time plot was used to identify the onset of noise in the signal
- When the onset of noise was difficult to identify the onset was determined by the following (listed in order of importance):

1. Reduction in F1 structure
2. Reduction in higher formant structure
3. Reduction in upper harmonics in the narrow-band spectrograph
4. Dramatic reduction in intensity curve

**Final vowel**

- The last identifiable glottal pulse was used as the end of the vowel. If the very end of the vowel was harsh or creaky, the last glottal pulse was considered to be the last pulse before the pitch shift.

## Appendix II. Acoustic classification of stress/accent patterns by language and child

Acoustic classification of stress/accent patterns by language and child is given in Table A1

| | Infant | Trochee | Spondee | Iamb | Total |
|---|---|---|---|---|---|
| American | 1 | 2 | 15 | 7 | 24 |
| | 2 | 6 | 9 | 1 | 16 |
| | 3 | 6 | 12 | 11 | 29 |
| | 4 | 1 | 7 | 7 | 15 |
| | 5 | 4 | 18 | 13 | 35 |
| | 6 | 15 | 14 | 2 | 31 |
| | 7 | 3 | 3 | 1 | 7 |
| | 8 | 6 | 5 | 4 | 15 |
| | 9 | 0 | 7 | 1 | 8 |
| | 10 | 0 | 5 | 5 | 10 |
| | Total | 43 | 95 | 52 | 190 |
| French | 1 | 6 | 25 | 9 | 40 |
| | 2 | 2 | 7 | 7 | 16 |
| | 3 | 5 | 8 | 6 | 19 |
| | 4 | 15 | 9 | 3 | 27 |
| | 5 | 4 | 15 | 6 | 25 |
| | Total | 32 | 64 | 31 | 127 |
| Finnish | 1 | 6 | 7 | 2 | 15 |
| | 2 | 4 | 7 | 0 | 11 |
| | 3 | 4 | 4 | 1 | 9 |
| | 4 | 2 | 6 | 0 | 8 |
| | 5 | 3 | 8 | 4 | 15 |
| | 6 | 7 | 9 | 1 | 17 |
| | 7 | 8 | 8 | 3 | 19 |
| | 8 | 19 | 10 | 1 | 30 |
| | 9 | 0 | 5 | 7 | 12 |
| | 10 | 2 | 3 | 1 | 6 |
| | Total | 55 | 67 | 20 | 142 |
| Welsh | 1 | 0 | 16 | 2 | 18 |
| | 2 | 6 | 23 | 7 | 36 |
| | 3 | 3 | 24 | 40 | 67 |
| | 4 | 13 | 16 | 12 | 41 |
| | 5 | 3 | 15 | 0 | 18 |
| | Total | 25 | 94 | 61 | 180 |

## References

Behrens, H., & Gut, U. (2005). The relationship between prosodic and syntactic organization in early multiword speech. *Journal of Child Language*, *32*, 1–34.

Boysson-Bardies, B. de., Bacri, N., Sagart, L., & Poizat, M. (1981). Timing in late babbling. *Journal of Child Language*, *8*, 525–539.

Boysson-Bardies, B. de., & Vihman, MM. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language*, *67*, 297–319.

Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language*, *6*, 411–421.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system. *Journal of Speech, Language, and Hearing Research*, *43*, 721–736.

Cutler, A., & Mehler, J. (1993). The periodicity bias. *Journal of Phonetics*, *21*, 103–108.

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J. K. (2000). Prosodic correlates of stress in babbling: An acoustical study. *Child Development*, *71*, 1258–1270.

DePaolis, Rory A. (2006). The influence of production on the perception of speech. In D. Bamman, T. Magnitskaia, & C. Zaller (Eds.), *Proceedings of the 30th Boston University conference on language development* (pp. 142–153). Somerville, MA: Cascadilla Press.

Di Cristo, A. (1998). Intonation in French. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems* (pp. 195–218). Cambridge: Cambridge University Press.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language*, *36*, 202–225.

Engstrand, O., Williams, K., & Lacerda, F. (2003). Does babbling sound native? Listener responses to vocalizations produced by Swedish and American 12- and 18-month-olds. *Phonetica*, *60*, 17–44.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, *27*, 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *105*, 126–152.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, *105*, 611–633.

Hallé, P. A., Boysson-Bardies, B., & Vihman, M. M. (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, *34*, 299–318.

Iivonen, A. (1998). Intonation in Finnish. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems* (pp. 311–327). Cambridge: Cambridge University Press.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, *14*, 147–172.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress pattern of English words. *Child Development*, *64*, 675–687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, *24*, 252–293.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*, 159–207.

Kehoe, M., Stoel-Gammon, C., & Buder, E. H. (1995). Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research*, *38*, 338–350.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America*, *72*, 353–363.

Kessen, W., Levine, J., & Wendrich, K. A. (1979). The imitation of pitch in infants. *Infant Behavior and Development*, *2*, 93–100.

Klein, H. B. (1984). Learning to stress: A case study. *Journal of Child Language*, *11*, 375–390.

Konefal, J. A., Fokes, J., & Bond, Z. S. (1982). Children's syntactic use of vowel duration. *Journal of Phonetics*, *10*, 361–366.

Konopczysnki, G. (1993). The phonological rhythm of emergent language: A comparison between French and English babbling. *Kansas Working Papers in Linguistics*, *18*, 1–30.

Kunnari, S. (2000). *Characteristics of early lexical phonological development in children acquiring Finnish*. Acta Universitatis Oulensis B34, University of Oulu.

Kunnari, S., Nakai, S., & Vihman, M. M. (2001). Cross-linguistic evidence for acquisition of geminates. *Psychology of Language and Communication*, *5*, 13–24.

Levitt, A. G. (1993). The acquisition of prosody: Evidence from French- and English-learning infants. In B. de. Boysson-Bardies, S. de. Schonen, P. Jusczyk, P. MacNeillage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 385–398). Dordrecht: Kluwer Academic.

Levitt, A. G., & Wang, Q. (1991). Evidence for language specific rhythmic influences in the reduplicative babbling of French- and English-learning infants. *Language and Speech*, *34*, 235–249.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustical Society of America*, *32*, 451–454.

Lieberman, P. (1985). The physiology of cry and speech in relation to linguistic behavior. In B. Lester (Ed.), *Infant crying: Theoretical and research perspectives* (pp. 29–57). New York: Plenum Press.

Lieberman, P. (1986). The acquisition of intonation by infants: Physiology and neural control. In C. Johns-Lewis (Ed.), *Intonation and discourse* (pp. 239–257). London: Croom Helm.

Marcos, H. (1987). Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language*, *14*, 255–268.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.

Menn, L., & Matthei, E. (1992). The "two-lexicon" account of child phonology: Looking back, looking ahead. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research and implications* (pp. 211–247). Timonium, MD: York Press.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development*, *16*, 495–500.

Oller, D. K., & Smith, B. L. (1977). Effect of final-syllable position on vowel duration in infant babbling. *Journal of the Acoustical Society of America*, *62*, 994–997.

Papoušek, M., & Papoušek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language*, *9*, 137–158.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllabic nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693–703.

Pollack, K. E., Brammer, D. M., & Hageman, C. F. (1993). An acoustic analysis of young children's productions of word stress. *Journal of Phonetics*, *21*, 183–203.

Ramus, F. (2002). Language discrimination by newborns. *Annual Review of Language Acquisition*, *2*, 85–115.

Robb, M. P., & Saxman, J. H. (1990). Syllable durations of preword and early word vocalizations. *Journal of Speech and Hearing Research*, *33*, 583–593.

Robb, M. P., & Tyler, A. A. (1995). Durations of young children's word and nonword vocalizations. *Journal of the Acoustical Society of America*, *98*, 1348–1354.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Snow, D. (1994). Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research*, *37*, 831–840.

Snow, D. (2006). Regression and reorganization of intonation between 6 and 23 months. *Child Development*, *77*, 281–296.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*, 249–267.

SPSS for Windows (2005). *Release 14.0.0 [computer software]*. Chicago: SPSS Inc.

Suomi, K., Toivanen, J., & Ylitalo, R. (2003). Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*, 113–138.

Suomi, K., & Ylitalo, R. (2004). On durational correlates of word stress in Finnish. *Journal of Phonetics*, *32*, 35–63.

Umeda, N. (1975). Vowel duration in American English. *Journal of the Acoustical Society of America*, *58*, 434–445.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child.* Oxford: Blackwell.

Vihman, M. M. (2000). Prosodic development: A cross-linguistic analysis of the first word period. End of award report, Economic and Social Research Council, R000237087.

Vihman, M. M. (2001). A cross-linguistic study of the acquisition of geminate consonants. End of award report, Economic and Social Research Council Award, R000223134.

Vihman, M. M., & DePaolis, R. A. (1998). Perception and production in early vocal development: Evidence from the acquisition of accent. In M. C. Gruber, D. Higgins, K. S. Olson, & T. Wysocki (Eds.), *Chicago Linguistic Society 34, Part 2: Papers from the panels* (pp. 373–386). Chicago, IL: CLS.

Vihman, M. M., DePaolis, R. A., & Davis, B. L. (1998). Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development*, *69*, 935–949.

Vihman, M. M., & McCune, L. (1994). When is a word a word? *Journal of Child Language*, *21*, 517–542.

Vihman, M. M., Nakai, S., & DePaolis, R. A. (2006). Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. In L. Goldstein, C. Best, & D. Whalen (Eds.), *Laboratory phonology 8* (pp. 341–366). Cambridge: Cambridge University Press.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (1985). From babbling to speech: A re-assessment of the continuity issue. *Language*, *61*, 397–445.

Vihman, M. M., Thierry, G., Lum, J., Keren-Portnoy, T., & Martin, P. (2007). Onset of word form recognition in English, Welsh and English–Welsh bilingual infants. *Applied Psycholinguistics*, *28*, 475–493.

Vihman, M. M., & Velleman, S. L. (2000). Phonetics and the origins of phonology. In N. P. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge: Its nature and status* (pp. 305–339). Oxford: Oxford University Press.

Westermann, G., & Miranda, E. R. (2004). A new model of sensorimotor coupling in the development of speech. *Brain and Language*, *89*, 393–400.

Whalen, D. H., Levitt, A. G., & Wang, Q. (1991). Intonational differences between the reduplicative babbling of French- and English-learning infants. *Journal of Child Language*, *18*, 501–516.

Williams, B. (1985). Pitch and duration in Welsh stress perception: The implications for intonation. *Journal of Phonetics*, *13*, 381–406.

Williams, B. (1986). An acoustic study of some features of Welsh prosody. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 35–51). London: Croom-Helm.