Book Section:

Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words*

Marilyn May Vihman, Satsuki Nakai and Rory DePaolis

The broad goal of this study was to understand how children in the earliest stages of word use integrate their perceptual knowledge of ambient language prosody and segmental patterns with their production experience of speech motor control to begin to produce words and phrases with adult-like rhythm. Disyllabic babbling and identifiable words or phrases produced by five infants each acquiring American English, French and Welsh, at two developmental points within the single word period, were compared with elicited adult disyllables, both words and nonwords. The elicited adult forms were designed to resemble the segmental patterns produced by the infants in each group, in order to control for the effects of inherent segmental duration and phonotactic structure in making adult/child comparisons. We also considered an uncontrolled sample of prosodically isolated adult disyllabic productions of (child-directed) words and phrases derived from the same recording sessions which provided the infant vocalisation data in all three language groups.

1. Introduction

In the past 10 or 15 years of experimental studies of infant speech perception great strides have been made in our understanding of what children know about their native language by the end of the first year of life (Jusczyk 1997). Broadly speaking, we see that whereas knowledge of native language prosody is already gained in the womb in the last trimester before birth (Querleu et al. 1988), segmental patterns gradually become familiar as well over the course of the first year, with accelerated learning between 9 and 12 months. Recent work demonstrating infants' capacity for 'statistical learning' of arbitrary distributional patterns (Saffran, Aslin and Newport 1996) suggests that implicit learning is the basis for these advances. Implicit learning is

also reflected in vocal production, which shows ambient language effects as early as 10 months (Boysson-Bardies et al. 1989; Boysson-Bardies and Vihman 1991). However, the relationship between the impressive prelinguistic knowledge of the ambient language and the child's deployment of motoric patterns for the production of identifiable first words remains unclear.

The study of rhythmic patterns in production in children acquiring three accentually and rhythmically distinct languages should help us to gain a purchase on this problem. More specifically, the study of segmental duration provides an opportunity to look at rhythmic factors – the infants' ability to match the durational patterning typical of (C)VCV sequences in the adult language – while at the same time taking account of differences in the segments and the targets attempted and actually produced in each language group. To our knowledge, no cross-linguistic studies of infants have previously addressed these issues, although Vihman, DePaolis, and Davis (1998) included duration along with pitch and amplitude in an investigation of the acoustic and perceptual characteristics of infant disyllabic vocalisations in English and French. Similarly, Vihman and Velleman (2000) compared medial consonant duration at two points in the single word period in three language groups, English, French, and Finnish, showing that while there is great individual variability (and relatively long medial consonants) in all three groups early in this period, by the end of the period the medial consonants of children acquiring both English and French become shorter while those of Finnish infants, exposed to contrastive consonant length in the adult language, grow longer. Neither of these studies included analyses of adult data, however.

English, French and Welsh provide a good basis for comparing the effects of exposure to different accentual systems. All three languages show final syllable lengthening (Delattre 1966; Williams 1986) but they are otherwise reported to have complementary durational patterns. The dominant stress pattern of English disyllabic words is trochaic, or strong-weak (75% of English words, according to Delattre 1965). In fact, the proportion of English disyllabic words attempted by children in the single word period can be over 90% trochaic (Vihman and McCune 1994; Vihman, DePaolis and Davis 1998). On the other hand, children acquiring English tend to produce at least as many monosyllables as disyllables in the early word period (for the same five children included in the present study, 54% of all child word forms were monosyllables, 38% disyllables, while content words in running speech in the input were 70% monosyllabic: Vihman et al. 1994). The children's disyllabic productions include attempts at apparent monosyllabic targets preceded by dummy syllables (Vihman, DePaolis and Davis 1998). Those

vocalisations often give the impression of being inspired by adult phrases, which are typically iambic in the input (75%: Delattre 1965). This means that English presents the child with a relatively complex accentual learning problem, since it is necessary to match at least two distinct prosodic patterns, both of them of high incidence in the input.

In contrast, French accents the phrase-final syllable, mainly by lengthening it, so that disyllables are uniformly iambic, whether words or phrases (Fletcher 1991). Vihman, DePaolis and Davis (1998) found that toward the end of the single word period "the [first-to-second vowel] duration ratios for the French infants were relatively stable and adult-like, whereas the American infants showed only slight second syllable lengthening, on average, and a considerably higher level of variability for each syllable (especially the first) than was found in the French data" [p. 944].

Welsh disyllabic words, like English ones, are predominantly trochaic (Williams 1986); disyllabic phrases may be either iambic (*a 'fo* 'that's it!', *na 'ni* 'here we are', *yn 'dwyt* 'aren't you?') or trochaic ('*da, de?* 'good, hunh?', '*tyd ta* 'come-on then'), although no quantitative data as to the typical distribution are currently available. The nature of stress is quite different in the two languages, however. In English, stress is characterized by a combination of perceptually "strengthening" factors affecting the vowel nucleus: greater intensity, higher pitch, longer duration, and a qualitative difference (full as opposed to reduced vowel). In contrast, the stressed vowel in Welsh trochaic words is identifiably *short* rather than long; it is the consonant following the stressed vowel that is marked by lengthening (Williams 1986). Welsh stress is also characterised by greater intensity on the initial syllable. However, pitch prominence, although less reliable than relative duration as a cue to stress, tends to fall on the final syllable (Watkins 1993; Williams 1986). We will see below how these distinct accentual systems result in a unique durational pattern for the VCV portion of disyllables in each of these languages.

In our analyses of disyllabic units we will also distinguish onomatopoeic forms – rare in adult discourse in all three languages but more common in child-directed speech – from other words and phrases. Onomatopoeic words are more variable in terms of accentual pattern than are conventional words or phrases; they are also more likely to be produced with playful emphasis on either syllable. According to several native speakers, onomatopoeic words are characteristically iambic in Welsh, while in both French (e.g., *miam-miam* 'yum-yum') and American English (*tick-tock*) they are most often even-stressed. Thus, these unconventional lexical items tend to depart from the dominant word-accent pattern in all three languages. With regard

to the homogeneity of the input the child hears, if we limit ourselves to di-syllables, English and Welsh stand as mixed systems, in contrast to French, which has only a single accentual pattern for two-syllable utterances, aside from onomatopoeia.

Another potential source of variability in the median lengths of V-C-V elements in the word production of children exposed to different languages has been proposed, however. This is the difference in the inherent rhythmic variability of the adult languages, based on purely phonetic acoustic analyses rather than on phonological classification (Grabe and Low 2002; Ramus, Nespor and Mehler 1999). The durational variability captured by these rhythm class models derives from a calculation of the duration of vocalic and intervocalic intervals (exclusive of pauses), based on acoustic analysis of controlled sentences (Ramus, Nespor and Mehler 1999) or of longer passages of read speech (Grabe and Low 2002). Ramus and colleagues found the best acoustic correlate of rhythm classes to be a combination of the proportion of time allocated to vocalic intervals and the standard deviation of the duration of consonantal intervals. Grabe and Low suggested that a better indicator of rhythmicity could be obtained by separately calculating a "Pairwise Variability Index" across successive vocalic and intervocalic intervals, with normalisation for speech rate for the vocalic index. The difference between the two methods for assigning rhythm class is in the index derived for vocalic intervals: proportion of vocalic relative to intervocalic intervals in the Ramus et al. model, variability in the duration of vocalic intervals in the Grabe and Low model.

Languages are then placed within the matrix defined by the intersection of the two indices to define a (graded) rhythm class space. The incidence of vowel reduction, diphthongs and tense vowels plays a role in defining rhythm class under either model, as does the incidence of consonant clusters, including consonant length or geminates, and final consonants, all of which contribute to the variability of syllable types in a language (one of the factors thought to enter into the impressionistic classification of languages by rhythm types: Dauer 1983). Using this approach English, the classic example of the "stress-timed" language type, again stands in contrast with French, long considered to be a prototypical "syllable-timed language." Welsh falls in between. According to Grabe and Low's calculations (see their Fig. 2, p. 530), Welsh is closer to French than to British English, while the rhythm class characterization used by Ramus et al. (1999) would place Welsh about equidistant from British English and French but would classify Welsh as stress-timed, like English (see Fig. 3, Grabe and Low 2002, p. 534).[1]

The empirical goal of this study was to test the effects on child rhythmic learning of differences in input speech as regards relative homogeneity at (1) the level of word or phrase accentual pattern and (2) the level of C-V alternation, or segmental sequencing. Although prosodic patterns elicit the earliest learned (i.e., language-specific) perceptual responses in infancy, segmental patterns begin to be known perceptually and to influence infants' vocal production patterns within the first year as well, as noted above. On the other hand, the segmental sequences found in early child words and contemporaneous babble are highly similar cross-linguistically. Within the limits of children's motoric planning skills, the specific rhythmic patterns of the adult language should have some influence; the limits of those planning skills remain largely unknown for children at this stage, however.

At the level of both larger (word, phrase) and smaller units (sublexical segmental sequences) French is less variable than either English or Welsh, leading to the expectation that children acquiring French will advance more rapidly in approximating the adult pattern. At the larger unit level, there is little to choose between the variability of model structures for disyllables in English and Welsh, however. Both have mainly trochaic words in input speech and a mix of trochaic and iambic phrases, although disyllabic words are far more common in the input in Welsh than in English, as we will see. At the segmental sequence level English is more variable than either French or Welsh (based on Grabe and Low 2002) and could thus be expected to provide children with the greatest challenge; Welsh falls between English and French in this regard.

2. Method

Data from five children each were drawn from longitudinal studies conducted in three language communities: American English (California; 10 participants in the original study, five of them boys: Vihman et al. 1985), French (Paris, France: Boysson-Bardies and Vihman 1991), and Welsh (North Wales). All of the infants were normally developing. All five of the American children whose data are included here, three of the French, and one of the Welsh were first-born. Two American, three French and one Welsh child were male. Children were recorded at home on a weekly (English) or biweekly basis (French and Welsh) on audio and video, in natural interaction with their mothers and sometimes with the observer, who was always a native speaker.

Two word points were identified for sampling the data in a comparable way cross-linguistically: the 4-word point, the first month in which the child used four or more identifiable adult-based words spontaneously in a half-hour session (4wp: two sessions sampled) and the 25-word point (25wp: one session sampled). The latter corresponds to approximately a 50-word cumulative vocabulary (for word identification procedures and other methodological details used in all three studies see Vihman and McCune 1994). For English we used data only from the five children who had reached the 25wp by the age of 17 months. In both the French and the Welsh groups a sixth child dropped out of the study when word production proved to be slow in getting started. There is thus a small bias toward precocity in word production in all three groups. All analyzable disyllables, including both words and nonwords or babble vocalisations, were extracted from each child's 4- and 25-word points. In some instances supplementary disyllables were selected from the week immediately preceding the 25-word point for the English group, since these children produced less disyllables overall. Mean group ages and numbers of disyllables analysed are indicated in Table 1.

Table 1.   Ages and tokens analyzed.

Language	Mean age in months		Tokens analyzed	
Word point	*4*	*25*	*4*	*25*
English	11.5	16	111	117
French	11.5	16.5	114	169
Welsh	13	17	168	167

2.1. Selection Criteria

The study was limited to disyllabic vocalisations only, including both identifiable words and babble, for two reasons: (1) the disyllable is the minimal unit needed for the investigation of intervocalic consonantal length, and (2) only monosyllables and disyllables are of high incidence in infant production at this stage, cross-linguistically (Vihman et al. 1994). Utterances selected for inclusion minimally contained two open (vocalic) phases separated by a closed (consonantal) phase. We included every disyllable which lent itself to objective analysis by the methods available. Items whose medial consonant was a glide, which poses particular problems for segmentation, were

excluded. Disyllables with interfering talking or other noise were not used. Utterances which showed excessive shifts from modal register, excessive vocal effort, whisper, or creaky voice were also excluded. No more than three successive repetitions of a single word type were included in the analysis, on the grounds that a "prosodic set" could be inferred and such mechanical repetition might bias the results.

2.2. Spontaneous adult data extracted from recordings

In order to obtain a representative sample of spontaneous adult speech whose durational properties could provide an idea of the range of models the infants are exposed to we searched through the 4wp and 25wp recordings of the children included in this study, in all three languages, for all prosodically isolated adult disyllabic productions, whether words or phrases. We defined "isolated disyllables" as those that were separated from adjacent utterances by at least 300ms (following Fernald et al. 1989). In the case of Welsh, mothers as well as children wore a wireless microphone connected to a transmitter. In the case of English and French, only the infants wore a microphone in most sessions, however. Since the mothers were located at varying distances from the infant in the course of the sessions, only a relatively small subset of the disyllables extracted lent themselves to analysis for those languages. Disyllables that were whispered or overlaid with noise, or that lacked a medial consonant, were excluded.

2.3. Elicited adult data

In order to form a clear idea of the "end state" toward which the children might be supposed to be heading while abstracting away from the differences in phonotactic structure between child and adult productions, we elicited child-like disyllabic patterns from five female native speakers of each language. Since target words, or the adult models for words attempted by the children, occurred only rarely as isolated disyllables in the recorded sessions, we included in our stimuli the most common child word targets for each language group alongside nonwords modeled on frequently occurring children's disyllabic nonword patterns (Table 2). Note that one phonological pattern – /babi/ – was elicited in all three groups for direct comparison across languages (Eng. *bobby,* Fr. *babie,* Wel. *babi* 'baby').

Table 2.   Elicited adult word and nonword disyllables (word spelling is in italics).

English		French			Welsh		
spelling	*phonetic*	*spelling*	*phonetic*	*gloss*	*spelling*	*phonetic*	*gloss*
apple	æpəl	*bébé*	bebe	baby	*babi*	babi	baby
baby	beɪbi	*chapeau*	ʃapo	hat	*choo choo*	tʃutʃu	c
Big Bird	bɪgbəd	*maman*	mamã	mama	*eto*	ɛtɔ	again
bottle	bɑɾəl	*papa*	pɑpɑ	papa	*si-so*	sisɔ	see-saw
button	bʌʔən	*poupée*	pupe	dolly	*ta ta*	tɑtɑ�	bye-bye
					tedi	tɛdi	teddy
bebba	bɛbə	aba	aba		ada	ada	
bebby	bɛbi	aideux	ɛdø		dede	dɛdɛ	
bobby	bɑbi	babie	babi		dwdw	dudu	
doodoo	dudu	bobeau	bobo		gaga	gɑgɑ	
edda	ɛdə	doudou	dudu		gagak	gɑgak	
gogga	gɑgə						

2.4. Acoustic Analysis

In the case of the spontaneous English and French adult and child data, disyl-
lables were extracted from the audio tapes and digitized to 16 bits using an
Audiomedia sound board in a PowerPC (sampling rate 22.2 kHz). The Welsh
data had been recorded onto a DAT deck and so were transferred digitally
for further analysis. All measurements were made using Soundscope speech
analysis software. Duration measurements used concurrent information
from the amplitude trace, narrow and wide band spectrograms, and intensity
curve. Additional screens were used to expand the beginning and endpoints
of the segments to be measured in order to obtain more detailed signal infor-
mation related to each manually placed marker (see Vihman, DePaolis and
Davis 1998, Fig. 2, for an illustration). Rules for segmentation of the first
vowel (V_1), medial consonant (midC), and second vowel (V_2) were based on
relevant transition cues, depending on the surrounding segments. Utterance-
initial consonants were excluded from the measurements. Glides occurring
between vowel and consonant or consonant and vowel were included in the
vocalic measurement.

3. Results

3.1. Durations of V-C-V in the three languages: Adults

Figure 1 shows the proportional durations of the individual V-C-V elements of the form /babi/ as produced in isolation by five adult speakers of each of the three languages. The patterns in the three languages are clearly distinct. As expected, based on prior reports, English is marked by long V_1, French by long V_2, and Welsh by long midC. A repeated measures ANOVA (Language: Between-subject factor; Element: Within-subject factor) performed on the proportional durations of adult /babi/ indicated that the Language x Element interaction is highly significant [$F(4, 24) = 44.7$, $p < .001$]. Multiple comparisons (Bonferroni corrections applied) revealed that the following differences are significant: %V_1: English > French and Welsh, %midC: Welsh > English and French, %V_2: French > Welsh.
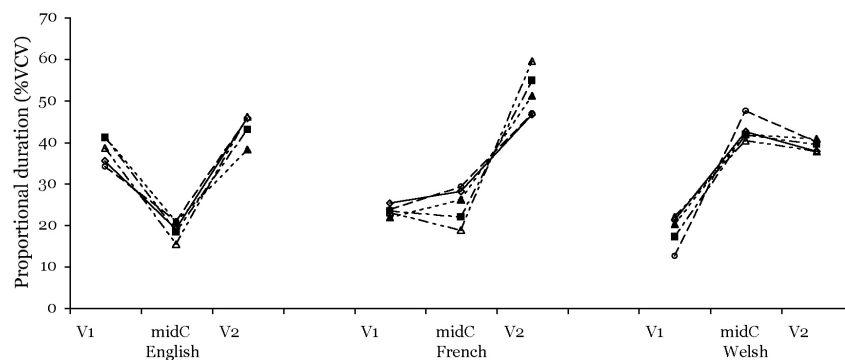

Figure 1. Proportional durations of elements of adult productions of /babi/. (Individuals are plotted in different lines.).

3.2. Durations of V-C-V in the three languages: Children

A comparison of the particular segments used in the child vocalisations measured in the three language groups revealed a number of significant differences that could be expected to affect the overall durations of VCV sequences. At both word points in all three language groups stops accounted for the highest proportion of medial consonants (close to 50% or more). The

second most frequent consonant type differed across groups and ages: nasals for the American children at both age levels but fricatives for both French and Welsh at the 4wp, with a shift to more nasals by the 25wp in both those languages.[2] The duration of children's consonant types differed in the same way across all language groups, with the ranking stops > fricatives > nasals (median 173, 148 and 133ms, respectively).[3]

Since the particular consonants produced in the disyllables used for analysis could affect the results of cross-linguistic comparisons we limited our durational analyses to vocalisations with medial stops. We also limited the vocalic portions to monophthongs, excluding tokens consisting of either diphthongs or syllabic consonants, which amounted to less than 12% of all vocalic nuclei for any language at either word point but which tended to be longer than monophthongs. For example, the median duration of monoph-thongal V2 was ca. 210 ms while the median duration of diphthongs and syllabic consonants in that position was ca. 300ms.

Figure 2 shows the vowel-stop-vowel (V-S-V) profile for the children's vocalisations in each language group, at the 4wp and the 25wp, with the segmental types restricted to stops and monophthongs. Median durational values for adult production of all the elicited nonword disyllables included in Table 2 are plotted on the right for comparison (recall that only stops and monophthongs were included in the elicited nonwords). Since the pro-portional durations of V-S-V elements differ for different consonants and vowels, even within the limits we imposed, error bars represent the range of median values of the various child-form disyllables.[4]

At the 4wp in all three groups the majority of the children, like the adults, show proportionately longer V_2 than V_1, giving the effect of final syllable lengthening. The proportional durations of the three elements taken together reveal relatively little evidence of specific ambient adult language shaping, however. Only one American child produced elements whose relative dura-tions resembled those of the American adults' elicited forms (broadly, midC $< V_1, V_2$), while two Welsh children produced a pattern resembling that of the Welsh adults ($V_1 <$ midC, V_2) and three French children produced a pattern re-sembling that of the French adults (V_1, midC $< V_2$). Furthermore, considering the individual children in each group, only one American child, one French child and two Welsh children produced all three elements within the range of the adult values ($\pm$10%) at the 4wp. The French children's patterns give a more homogeneous impression than do those of the other two groups, but Co-chrans C tests indicate that the cross-linguistic differences are not statistical-ly significant (C[4, 3]= .46, p = .74 for V_1; C[4, 3]= .61, p = .24 for midC).

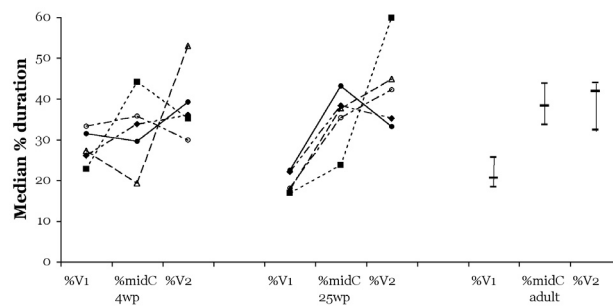American children
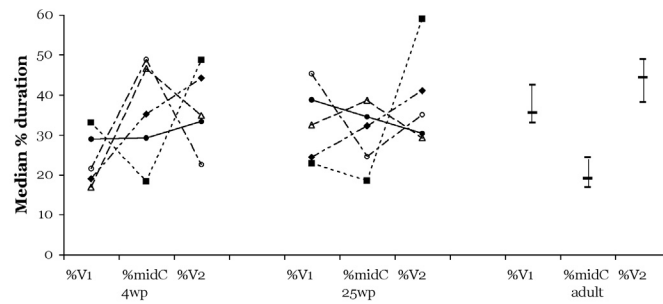
French children
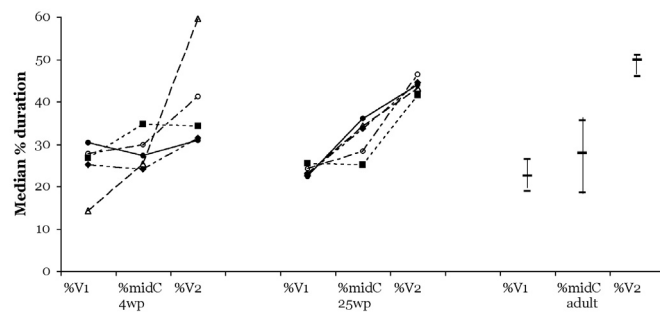
Welsh children

Figure 2. Proportional durations of elements of child disyllables at 4 and 25wp (leftmost panels); right panels show median and range of durations of each element in all elicited adult nonword disyllables for each language.

We see from Figure 2 that in all three language groups the children's vocalisations do show a change between the two sampling periods such that at the 25wp they more closely approximate the adult pattern of the target language. This necessarily required different patterns of change within each language group, and for the different children within each group. Overall, in proportional terms, American children's V_1 became longer and midC shorter; French children's V_1 became shorter and V_2 longer; Welsh children's V_1 became shorter while their midC and V_2 became longer. Considering individual children, however, French and Welsh children had achieved far more adult-like V-S-V proportions than had American children by the later developmental point. Of the five children in each group, five French and four Welsh children produced V-S-V proportions that approximately matched the adult shape while only one of the American children matched the ratio for all three elements. A better match between child and adult productions for French and Welsh in comparison with English at the 25wp results in greater homogeneity for the groups of French and Welsh children as a whole. A Cochrans C test indicated that at the 25wp the difference in the homogeneity of variance is significant for the proportional duration of V_1, where American children differed from the remaining two groups: C(4, 3) = .92 p = .001. The differences for the other elements are not statistically significant, however.[5]

Table 3.   Number (proportion of total) and structures of mothers' isolated disyllables.

		Total (proportion)	long /tense V, diphthong, glide + V		med. cluster	final consonant
			V1	V2		
English	*words*	44 (.37)	24	31	8	16
	phrases	65 (.54)	27	37	35	59
	onomat.	11 (.09)	10	10	2	10
	Total	120	61 (.51)	78 (.65)	45 (.38)	85 (.71)
French	*words*	101 (.61)	3	5	21	40
	phrases	46 (.28)	3	7	2	7
	onomat.	19 (.11)	0	0	4	0
	Total	166	6 (.04)	12 (.07)	27 (.16)	47 (.28)
Welsh	*words*	154 (.62)	13	9	54	54
	phrases	54 (.22)	8	6	15	12
	onomat.	39 (.16)	27	9	8	17
	Total	247	48 (.19)	24 (.10)	77 (.31)	83 (. 37)

3.3. Crosslinguistic comparison of mothers' isolated disyllables

Table 3 characterises the isolated disyllables extracted from the mothers' child-directed speech and analysed for the duration of the V-C-V elements. The disyllables are categorised into the three units that might be expected to differ in accentual pattern: words, phrases, and onomatopoeic forms. As expected based on earlier studies, words make up a far smaller proportion of the disyllables produced in child-directed speech in English (37%) in comparison with French (61%) and Welsh (62%). The proportion of onomatopoeic forms is also somewhat smaller in English and French than in Welsh. Phrases make up over half of the mothers' isolated disyllables in English, as compared to well under a third in the other two languages. In particular, certain phrases are of very high incidence in the English data (over half are wh-questions, such as *who's that? what's this? where else?*). Furthermore, although trochaic words are the dominant English pattern, only 23 out of the 44 tokens analyzed were standard trochaic words (e.g., *bunny, cupcake vs. allgone, whoopee,* which have variable accent).

To better show the differences between English and the other two languages with respect to the complexity of the nucleus and other aspects of phonotactic structure that affect the rhythmic profile, we also indicate the occurrence of diphthongs and contrastively long or tense V_1 and V_2, medial clusters, and final consonants in the mothers' disyllables. From this it is clear that the phonotactic complexity of our spontaneous English data exceeds that of the other two languages on all three measures.
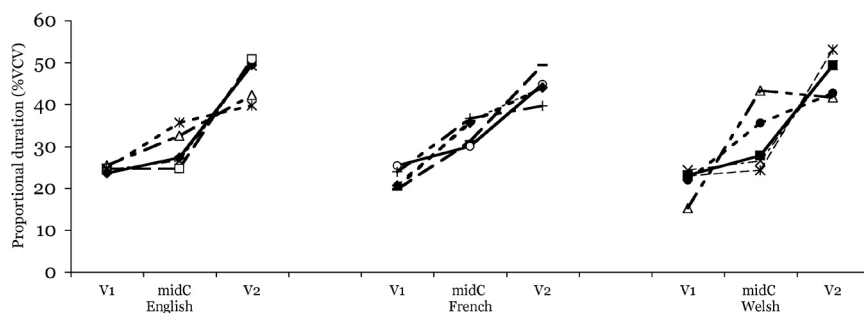

Figure 3. Proportional durations of V-C-V elements of mothers' disyllables.

Figure 3 summarizes the distribution of proportional durations of V_1, midC, and V_2 for the analyzable disyllables extracted from each mother's

child-directed speech. The differences between language groups are far less evident here than in the elicited disyllabic nonwords (adult panels, Fig. 2), in which accentual pattern, segmental type, and phonotactic structure were all controlled. Furthermore, the cross-linguistic differences in variability that we expected to see are not apparent either. We see, instead, the same differences in variability across the elements of the V-C-V sequence in all three languages: V_1 covers a smaller proportion of the total duration and is the least variable, midC takes up a larger proportion of the duration and varies somewhat more widely, and V_2 is proportionately the longest and the most widely varying. Phonotactic complexity appears to interact with other factors to obscure any durational differences due to accentual patterning alone in this uncontrolled data sample.

If these findings provide an accurate picture of the V-C-V duration profiles in input speech, however, we are left with a puzzle: How can we account for the children's progress toward distinct rhythmic patterns by the 25wp if isolated disyllables taken from mothers' child-directed speech are so similar across the languages? What then are the children's models?

3.4. Children's disyllabic words in relation to targets

The children differed across languages in the types of adult disyllabic targets they attempted and produced; to some extent this reflects differences in adult language structures and in the nature of the input to children in the different groups (see Table 4). As expected, the French children produced the most words following the dominant adult word pattern (iambic words: 61%), with the remainder divided between a few onomatopoeic forms (8%) and a larger number of "other" forms including interjections (*allo, bravo)*, other words or phrases whose accentual pattern is variable (*non-non*), and monosyllabic words preceded by a likely "filler syllable" ([*à*] *voir* [aβa]*, [de] l'eau* [dəlo]), giving the impression of an iambic phrase, although the actual target can only be guessed at (on filler syllables in French, see Veneziano and Sinclair 2000).

The Welsh children produced twice as many onomatopoeic forms as either of the other two groups (21% of all target types), with less than half of all target word types conforming to the dominant word pattern (42% trochaic words). No putative phrases consisting of filler + monosyllabic content word were identified as targets for Welsh child productions with the exception of *oh God! (*produced by a child with several older siblings).[6]

Table 4.   Adult targets for children's disyllables (proportion of all target types).

	Dominant pattern		*Onomatopoeia*		*Other*		*Total*	
	4wp	*25wp*	4wp	*25wp*	4wp	*25wp*	4wp	*25wp*
English	5	23	2	4	6	19	13	46
Total	28 (.47)		6 (.10)		25 (.42)		59	
French	7	29	3	2	3	15	13	46
Total	36 (.61)		5 (.08)		18 (.31)		59	
Welsh	10	20	8	7	9	18	27	45
Total	30 (.42)		15 (.21)		27 (.37)		72	

Just under half of the targets attempted by American children were tro-
chaic words (47%), while a sizable minority fell in the category "other", in-
cluding at least five monosyllabic words preceded by a filler syllable with
unverifiable target (*[a/the] bead[s]*) in addition to one iambic word (*balloon*)
and a certain number of longer words or monosyllables treated in some other
way.

To better understand the difference in variability of rhythmic patterns at
the 25wp across the different groups of children we analyzed the elicited
adult forms of five target disyllables in each language (Table 2). We selected
for analysis the most frequently attempted target words in each group (tak-
ing into account the number of tokens attempted by each child as well as the
number of children attempting each word type). Figure 4 presents propor-
tional V-C-V profiles for these elicited adult productions.
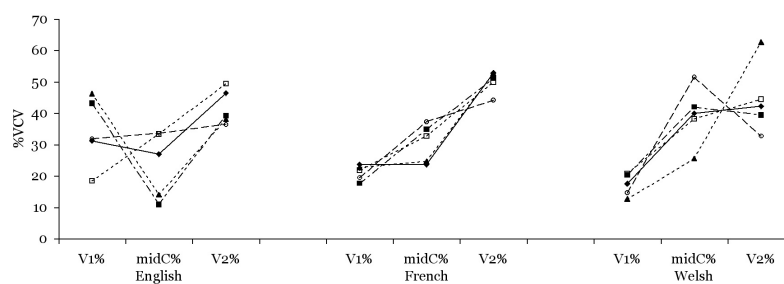

Figure 4.   Proportional durations of elements of elicited adult productions of target
        disyllabic words.

Here we begin to gain a better idea of the source of the differential variabil-
ity of children's productions in the three languages. That is, we can see that

although all of the English words measured here are trochaic, the durational variability of all three elements remains considerable, reflecting the variability in the intrinsic duration of different types of segments (e.g., tense vs. lax stressed vowel, full vs. reduced unstressed vowel) as well as in phonotactic complexity (consonant singletons vs. clusters).

The similarity between the median proportional durations plotted for the different children's disyllables produced at the 25wp (Fig. 2) and those of the common target words (Fig. 4) is striking, especially in the case of French and Welsh. In order to investigate this finding more closely we undertook further analysis of these individual word targets. Figures 5–7 provide a direct comparison of the proportionate V-C-V durations of selected adult (median) and child productions.

In Figure 5 we see the proportionate durations of four French words (*chapeau* 'hat', *maman* 'mother', *papa, poupée* 'doll') as produced by two to four children as well as the median production pattern for the fifteen tokens elicited from adults for each of these words. The four adult V-C-V patterns are generally similar. On the whole the children are producing moderately good matches to the particular word targeted; the children's patterns are relatively similar across the four words as well.
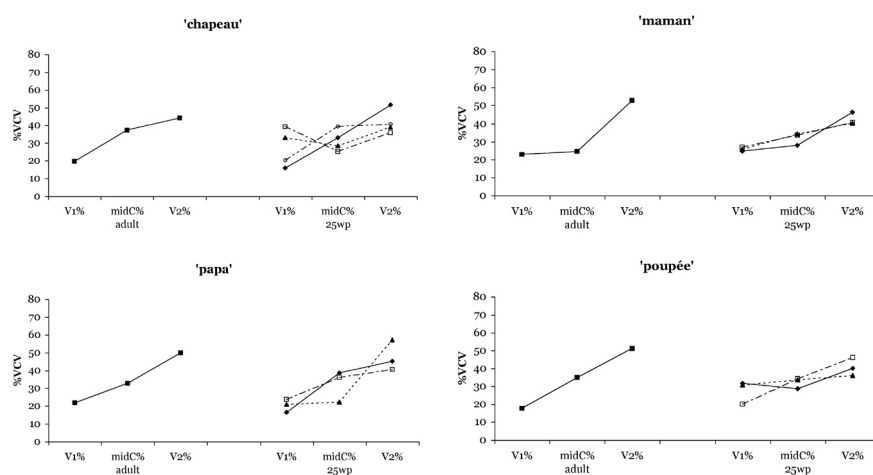

Figure 5.   Individual French words; elicited adult targets and child productions.

In Figure 6, where the proportionate durations for four Welsh words are plotted (*choochoo, eto* 'again', *tata* 'bye-bye' and *tedi* 'teddy'), we can see

a more striking match of child productions to individual adult word models. Each of the target words has a distinct V-C-V pattern, and the children's productions constitute good matches to the individual pattern of at least three of them. This seems to reflect item learning. The difference in the between-word similarity for adult French as compared with Welsh further suggests that the source of the child variability that we have observed may lie primarily in the variability of different adult models.
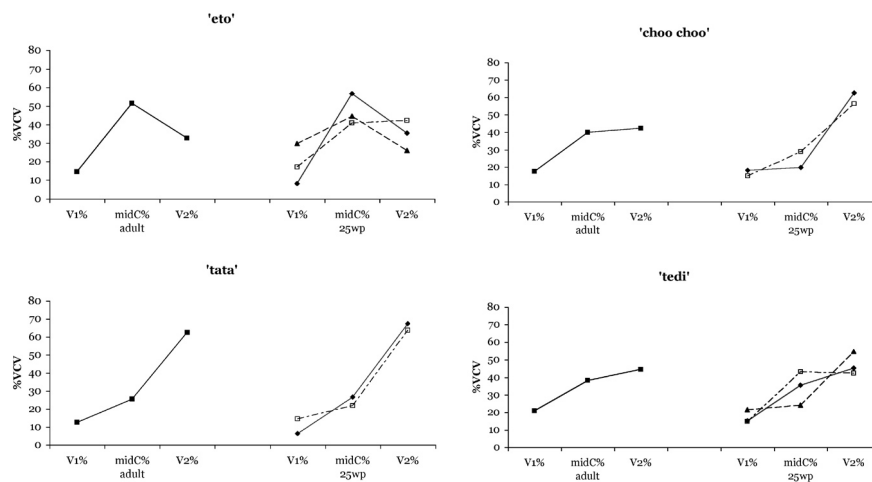

Figure 6. Individual Welsh words; elicited adult targets and child productions.

In Figure 7 we see a comparable set of comparisons for American English. Notice that although all four words are trochaic, their adult rhythmic profiles are not the same. In particular, contrast *baby*, which elicits three relatively good matches, with the other three words. In the case of *button,* in particular, the tokens produced by the two children look quite different: One of these tokens is transcribed as [bætːi], the other as [paːŋa], with the final nasal moved to medial position as in many of the child's other words (for an account of the development of this word template, see Vihman and Velleman 1989).

More generally, the differences between the fidelity of the individual children's productions as regards rhythmic match to the adult targets in the three languages can plausibly be attributed to the difference in the structural complexity of the target words. Of the five words frequently attempted in each group, all of the Welsh target words and all but one of the French target

words have a simple $(C_1)VC_1V$ structure (the exception is *chapeau,* with child forms [pœpœ, bɔbo, habœ, (h)apo]: Note the variability of these tokens in Figure 5). For English, three of the target word forms include stops differing in place of articulation *(Big Bird, bottle, button)*, three have a syllabic consonant as the second nucleus, *Big Bird* has a medial consonant cluster and even *baby*, which inspires the best English matches, has a diphthong for its first nucleus. None of these relatively more complex segmental types occur in any of the French or Welsh target words that were frequently attempted by the children. The relative difficulty posed by the English models has its equivalence in the phrases found in the input: Contrast the frequent English wh-questions noted above, with their two- to three-consonant clusters, with the French equivalents found in our data: *et ça?* 'and that?', *et là?* 'and there?', *qui c'est?* 'who is it?', *c'est quoi?* 'what is it?'.
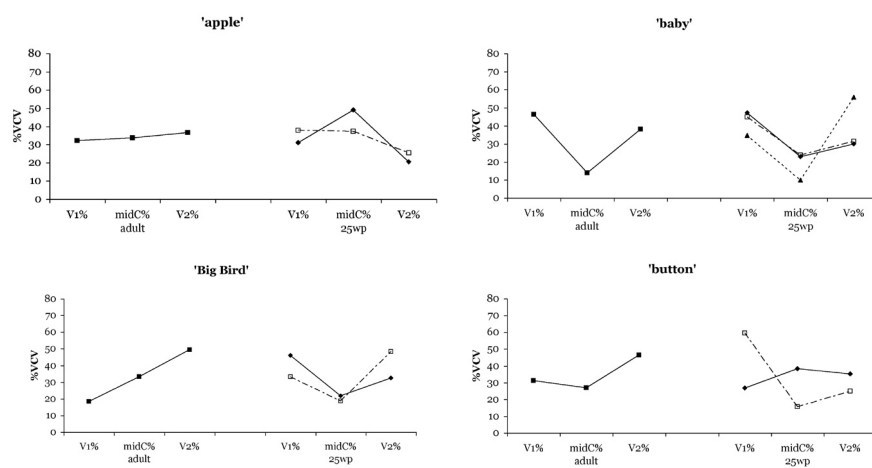

Figure 7. Individual English words; elicited adult targets and child productions.

In view of these differences it is reasonable to suppose that, in comparison with English, it is more straightforward for the child acquiring French or Welsh to succeed in making a rhythmic mapping from adult target to a form in his or her existing articulatory repertoire, at least where disyllables are concerned. (Recall that monosyllables are the dominant production pattern for most children acquiring English, but not for the other two languages.) Thus, the greater rhythmic variability that has been noted for English in comparison with French and also Welsh (Grabe and Low 2002) seems to give a

reasonably good account of the children's relative success in matching the adult rhythmic standard at the 25wp.

4. Discussion and conclusion

Analysis of the elicited child-like disyllabic patterns produced by adults in the three language groups showed little within-language variability and clear between-language differences in the relative length of V_1, midC and V_2. Comparison of these results with the V-S-V productions of children at two developmental points yielded three primary findings:

(1) At the 4wp children could not be readily assigned to the appropriate language group on the basis of the way their rhythmic pattern apportioned length to V-S-V. All three groups seemed to show final vowel lengthening, also seen in the adult languages. As a group the French children appeared relatively more homogeneous and adult-like than the American and Welsh children, but the difference was not significant.

(2) By the later developmental point there was progress toward the adult model in all groups. This agrees with earlier cross-linguistic studies of segmental production over this same period (Boysson-Bardies and Vihman 1991; Vihman et al. 1994).

(3) Inter-group differences were found in the extent to which the children succeeded in matching adult rhythmic patterning. Both the French and the Welsh children generally conformed relatively closely to the adult pattern by the 25wp. In contrast, the American children remained more variable and less closely matched to the adult models.

4.1. Sources of variability

We explored two potential sources for the lower variability and greater conformity to the adult pattern seen in the French children's production, namely, greater homogeneity at the level of both larger (word, phrase, onomatopoeia) and smaller rhythmic units (sublexical segmental sequences). On the one hand, French seems to provide a more homogeneous set of target forms to model than either English or Welsh, due to the fact that French disyllabic speech forms are largely iambic, whether words or phrases. Exceptionally, forms may be evenly accented on both syllables or produced with initial-syllable stress for idiosyncratic reasons (affect, playful varia-

tion, etc.). In contrast, both English and Welsh present a mix of trochaic, iambic, and even-stressed models. However, our analyses of isolated disyllables extracted from the mothers' speech failed to provide support for the hypothesis that differences in the rhythmic patterns of different lexical types is the primary source of differences in the extent of child variability across groups.

An alternative proposal holds that the inherent rhythmic variability of a "stress-timed" language like English provides a relatively more difficult model for children to match than does French, one of the classic "syllable-timed" languages. Grabe, Post and Watson (1999), in a study of speech production in 4-year-olds acquiring (British) English vs. French, argue that the greater success shown by French children in matching the rhythmic variability level of adults is due to the fact that French is rhythmically simpler than English because French has a less variable rhythm overall, as detected in fluent adult speech.

The present study generally supports the rhythm class models (Grabe and Low 2002; Ramus, Nespor and Mehler 1999) as predictors of children's relative ease of acquisition of rhythmic patterning. On the one hand, in our elicited adult nonword productions we controlled the rhythmic complexity of the adult models by eliciting the kinds of simple patterns typically produced by children and found that French and Welsh children's vocalisations were more adult-like than those of the American children at the 25wp. On the other hand, when we compared elicited adult productions of words attempted by the children in the three languages, the difference in variability in the adult word targets was evident (Fig. 4). Furthermore, the relative variability of the five elicited target words in each language – ordered as English > Welsh > French, in agreement with the predictions of the rhythm class models – bore a close resemblance to the relative variability across the five individual children within the three language groups at the 25wp. In other words, the target words for the French children's early productions were themselves closely similar to one another, rhythmically speaking (Fig. 4), and the different children's tokens were similar as well (Fig. 5). In Welsh, in accordance with the "larger unit" variability discussed above, the words targeted by the children were more disparate in their rhythmic pattern than the French target words (Figs. 4, 6) but the children generally achieved good matches to them. In English, in contrast to both French and Welsh, the elicited target words, although all trochaic, differed considerably, apparently reflecting differences at the sub-lexical level (Figs. 4, 7). Specifically, the word targets attempted by the American children included

more phonotactically complex sequences and more segments of a kind not typically found in children's earliest production repertoires. One could infer that it is a consequence of the greater phonetic challenge posed by the adult models that some of the English child word productions depart more radically from the adult forms than do any of the French or Welsh child word productions.

It may be worth noting, finally, that although the word targets attempted most often by the children learning English were more complex than those of the children learning the other languages, those targets were nevertheless less complex than the American mothers' isolated disyllabic words. Only one out of five of the children's frequent target words (20%) includes either a diphthong or a long or tense vowel (vs. 81% of the mothers' words) or a final consonant (vs. 35% of mothers' words). Without laboring the point, the children's frequent choices of early words to produce seem to reflect the kind of selectivity often reported in the literature (see Schwartz 1988).

4.2. The representation of rhythm and item learning

Returning to the broader issue of implicit prelinguistic knowledge in relation to early word production with which we began, we are now in a position to reflect again on what the child needs to know in order to "get the rhythm right". To achieve the native language pattern in production the child must eventually be able to match both the overall melody and the rhythmic pattern of individual words. In the single-word production period this might seem not to be overly challenging, since the complication of fitting content and function words, or even a succession of content words, into a single intonational contour does not yet arise. The problem is rather one of representation: In the first year, as indicated above, there is good evidence that the child is able to gain a sense of the rhythmic patterning of the language sufficient to give greater attention to words fitting the dominant pattern than to other words. Our results suggest that it is not this global representation of the rhythmic patterning in the language that underlies children's early word production, however.

We saw, first of all, that at the outset of word production children's disyllables do not yet match adult rhythmic patterning (Fig. 2). In fact, despite the extensive perceptual learning of the first year, we see here that different children acquiring the same language show distinct patterns, only a few achieving a rough approximation of the dominant rhythmic pattern of the ambient

language. Thus, the construction of representations at a new level seems to be needed, involving considerable additional learning before ambient language rhythms can be successfully matched in production. Specifically, as the evidence from Welsh demonstrates the most clearly, early word learners seem to be developing adult-like rhythmic patterns for production on a word-by-word basis. A more abstract knowledge of the rhythmic patterning appropriate to the native language can be expected to emerge only gradually from the combined perceptual and production knowledge or representation of increasing numbers of individual words, with further reorganization once larger and more varied units begin to be combined and integrated into word combinations (see Snow 1994).

It has been well established that children's early words build on articulatory patterns already available in babbling. Assuming that early vocal production patterns are grounded in biomechanical constraints and are thus common to children learning different languages, the opportunities for easy matching will differ according to the particular ambient language. In addition, the findings of this study agree with earlier work in suggesting that the challenges posed by the adult language will be met differently by different children. Cross-linguistic phonological analyses of first words (4wp) in comparison with later words (25wp or beyond) suggest that each child must begin by developing a stock of individual representations of particular lexical items (Vihman 2002). These initial representations, or word production patterns, may derive from matches between the child's existing vocal patterns and (implicitly "pre-selected") adult words (Vihman and Nakai 2003).

That is, the highly selected first words reflect existing vocal patterns developed in babbling; these differ in complexity from one child to the next. Through production practice with a growing stock of adult-based words the child is then very gradually able to induce the more abstract structure of the adult language. This implicates explicit as well as implicit learning, as the child moves from the use of situationally primed first words to more intentionally targeted and more flexibly deployed later "referential" words (McCune and Vihman 2001). Our study has captured two early points in that process. At the 4wp, when babbling vocalizations make up much of the child's production, the influence of the overall rhythmic patterns of the adult language is weak, if detectible at all. By the 25wp, when the children are using 50 words or more, individual words are relatively successfully reproduced, resulting in greater homogeneity for the French group, with its less variable rhythmic patterning, than for the other groups. The achievement of adult-like rhythmic patterns will require mastery of considerably greater phonotactic

complexity for all the children, but particularly for those learning English; accurate deployment of the range of accentual patterns available for words, phrases, and onomatopoeic forms can be expected to emerge in parallel with that developing mastery, as part of the process of lexical learning. Thus the implicit learning of native-language rhythms that occurs in the prelinguistic period is only a first step in the long apprenticeship that will culminate in adult-like rhythmic production.

Notes

*   The authors thank the Economic and Social Research Council of the UK for its financial support. We also thank the participating families from California, Paris and North Wales and the adults who produced the elicited data. Lucy Evans collected and transcribed the Welsh infant data, Pam Martin helped with Welsh word identification, Dr. Llinos Spencer and Dr. Enlli Thomas kindly answered questions about Welsh words and phrases.
1. The classification of Welsh by the Ramus et al. method was carried out by Grabe and Low, based on their own data.
2. Note that these differences were identified only in the sample of disyllables that lent themselves to analysis in this study. For analyses based on a larger sample of English and French child vocalisations see Boysson-Bardies and Vihman (1991).
3. The greater duration of stops than of fricatives in all three groups of children is surprising. We speculate that this may be due to greater intentionality on the child's part in producing a stop than a fricative, as suggested by the greater incidence of fricatives in babbling than in words in this period (Boysson-Bardies and Vihman 1991)
4. Note that for quantitative analyses of the children's data, given skewed distribution and some extreme values, we have entered the median measurement as a summary figure for each child and, for the sake of comparability, for the adult data as well.
5. Cochrans' C tests performed on proportional durations of elements produced at the two developmental points indicate that French children's proportional durations of V1 and V2 are significantly more homogeneous at the 25wp than at the 4wp ($C[4, 2] = .96$, $p = .009$ for V1; $C[4, 2] = .98$, $p = .003$ for V2). No other groups differed in terms of homogeneity of variance.
6. The primary language of all of the Welsh children's homes was Welsh. These children rarely produced English words, but since beyond early childhood virtually all speakers of Welsh are bilingual today, it would have been impossible to avoid including in the study any Welsh children who are also exposed to English.

References

Boysson-Bardies, Bénédicte de, Pierre Hallé, Laurent Sagart and Catherine Durand
    1989   A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language* 6: 1–17.

Boysson-Bardies, Bénédicte de and Marilyn M. Vihman
    1991   Adaptation to language: Evidence from babbling and first words in four languages. *Language* 67: 297–319.

Dauer, Rebecca M.
    1983   Stress-timing and syllable timing reanalyzed. *Journal of Phonetics* 11: 51–62.

Delattre, Pierre C.
    1965   *Comparing the Phonetic Features of English, French, German and Spanish*. Heidelberg: Julius Groos Verlag.
    1966   A comparison of syllable length conditioning among languages. *International Journal of Applied Linguistics* 4: 182–198.

Fernald, Anne, Traute Taeschner, Judy Dunn, Mechtilde Papousek, Bénédicte de Boysson-Bardies and Ikuko Fukui
    1989   A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16: 477–501.

Fletcher, Janet
    1991   Rhythm and final lengthening in French. *Journal of Phonetics* 19: 193–212.

Grabe, Esther and Ee Ling Low
    2002   Durational variability in speech and the rhythm class hypothesis. In: Natasha Warner and Carlos Gussenhoven (eds.), *Papers in Laboratory Phonology VII*, 515–546. Cambridge: Cambridge University Press.

Grabe, Esther, Brechtje Post and Ian Watson
    1999   The acquisition of rhythmic patterns in English and French. In: *Proceedings of the XIVth International Congress of Phonetic Sciences*, 1201–1204.

Jusczyk, Peter W.
    1997   *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

McCune, Lorraine and Marilyn M. Vihman
    2001   Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research* 44: 670–684.

Querleu, D., X. Renard, F. Versyp, L. Paris-Delrue and G. Crépin
    1988   Fetal hearing. *European Journal of Obstetrics and Reproductive Biology* 29: 191–212.

Ramus, Franck, Marina Nespor and Jacques Mehler
  1999   Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265–292.
Saffran, Jenny R., Richard N. Aslin and Elyssa L. Newport
  1996   Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.
Schwartz, Richard
  1988   Phonological factors in early lexical acquisition. In: Michael D. Smith and John L. Locke (eds.), *The Emergent Lexicon*, 185–222. New York: Academic Press.
Snow, David
  1994   Phrase-final syllable lengthening and intonation in early child speech. *Journal of Speech and Hearing Research* 37: 831–840.
Veneziano, Edy and Hermine Sinclair
  2000   The changing status of 'filler syllables' on the way to grammatical morphemes. *Journal of Child Language* 27: 461–500.
Vihman, Marilyn M.
  2002   Getting started without a system. *International Journal of Bilingualism* 6: 239–254.
Vihman, Marilyn M., Rory DePaolis and Barbara L. Davis
  1998   Is there a "trochaic bias" in early word learning? *Child Development* 69: 933–947.
Vihman, Marilyn M., Edwin Kay, Bénédicte de Boysson-Bardies, Catherine Durand and Ulla Sundberg
  1994   External sources of individual differences? *Developmental Psychology* 30: 651–662.
Vihman, Marilyn M., Marlys A. Macken, Ruth Miller, H. Simmons and Jim Miller
  1985   From babbling to speech: A reassessment of the continuity issue. *Language* 61: 395–443.
Vihman, Marilyn M. and Lorraine McCune
  1994   When is a word a word? *Journal of Child Language* 21: 517–542.
Vihman, Marilyn M. and Satsuki Nakai
  2003   Experimental evidence for an effect of vocal experience on infant speech perception. In: M. J. Solé, Daniel Recasens and J. Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona.
Vihman, Marilyn M. and Shelley L. Velleman
  1989   Phonological reorganization. *Language and Speech* 32: 149–170.
  2000   The construction of a first phonology. *Phonetica* 5: 255–266.
Watkins, T. Arwyn
  1993   Welsh. In: Martin J. Ball (ed.), *The Celtic Languages*. London: Routledge.

Williams, Briony
    1986   An acoustic study of some features of Welsh prosody. In: Catherine
              Johns-Lewis (ed.), *Intonation in Discourse*. London: Croom Helm.