17 The Acquisition of Temporal Patterns

MELISSA A. REDFORD

17.1 Introduction[1]

Speech can be defined in terms of the signal, as a time-varying acoustic waveform. We associate speech segments with abrupt spectral or amplitude changes in that waveform. Some of the variation in the temporal aspects of segments can be ascribed to the inherent durations of different articulatory configurations. Other variation is contextually driven. For instance, the order in which segments are sequenced has consequences for coarticulation, which has consequences for segmental duration. The relative speed with which segment sequences are executed also interacts with coarticulation, and so with patterns of segmental duration. Still other variation is linguistically specified. For example, at the segmental level, different languages will use different patterns of inter-articulatory timing to execute the "same" sound (e.g., Spanish /p/ is more similar to English /b/ than English /p/). At the suprasegmental level, there is prosody, which has extensive effects on segmental durations through reduction and lengthening processes. The focus of this chapter is on the acquisition of temporal patterns that are defined by all of these influences. In keeping with the continuity hypothesis, we assume that the motor factors that constrain the shape of pre-speech vocalizations also influence the realization of temporal patterns in children's fluent speech. That said, we recognize that the transition from vocal play and imitation (pre-speech) to concept-driven communication (speech) represents an important discontinuity in development. We propose that this transition marks the development of a speech plan; a representation that guides speech action. Overall, our thesis is that the acquisition of temporal patterns reflects motor skill development; but some patterns are planned, in that they reflect remembered speech action (i.e., stored acoustically-linked articulatory schema), while others emerge during fluent speech from the practiced execution of serially ordered schemas (i.e., multi-word plans). Our thesis structures the content presented in

The Handbook of Speech Production, First Edition. Edited by Melissa A. Redford.

this chapter, which is a review of research on early vocalizations, first words, and developmental changes in children's fluent speech.

17.2 Early sound patterns

The human drive to communicate is overwhelming. Several innate features scaffold the development of speech and, by extension, language acquisition. One of these features is the ability to act and react in an interpersonal context (Trevarthen 1979; Trevarthen and Aitken 2001). Relevant to speech are infants' different facial expressions and vocalizations (acting) combined with adjustments of these based on feedback from another (reacting). Human neonates will also initiate facial movement (e.g., tongue protrusion) with the expectation of a response from their interlocutor (Nagy and Molnar 2004). Another feature that scaffolds the development of communication is the innate ability and drive for infants to imitate facial expressions and movements (Maratos 1973; Meltzoff and Moore 1977). Importantly, the innate capacity for imitation is not restricted to the visual modality. Kuhl and Meltzoff (1996), for example, have shown that infants as young as 12 weeks of age shift their vocal productions in the direction of an auditory, speech-like stimulus. This ability suggests that an auditory-to-articulatory map, critical to the acquisition of target speech sounds, begins to be established early in development. Of course, attempting the production of target speech sounds also requires that they are represented in memory. Thus, a final feature necessary to scaffold speech acquisition is the innate ability to attend to and rapidly learn extended sound patterns (Saffran, Aslin, and Newport 1996). It is presumably this feature that allows the newborn infant to differentiate the rhythms of his mother's language from those of another language (Nazzi, Bertoncini, and Mehler 1998) in addition to enabling the older infant to recognize and extract recurrent sub-sequences (e.g., "words") from extended, fluent speech input (Saffran et al. 1996).

The innate drive to communicate coupled with remarkable imitative and perceptual learning abilities is presumably what pushes the infant to experiment with vocalization. By 2 months of age, infants make a variety of sounds. Most of these are reflexive and either communicate basic needs (i.e., cries) or have no discernible communicative function (i.e., grunts). A few sounds appear, however, to be speech-like. These deliberate vocalizations are the quasi-vowel-like sounds that define the so-called cooing stage of early vocal development (Stark 1980). From the point of view of production, the coo is significant in that it demonstrates the coupling of an articulatory posture with phonation. Specifically, the "coo" or "goo" quality of these sounds suggests a somewhat retracted tongue position. Unlike speech, the tongue is stationary during vocalization and only one dimension (height) of one articulator (tongue dorsum) is engaged to modify the sound across different repetitions (Oller 2000). It is also not clear that the modifications are deliberate in any way. That said, assuming the infant makes a connection between the modifications (based on somatosensory feedback) and the experience of different acoustics, this earliest stage of pre-speech vocalization will provide the foundation for the auditory-to-articulatory mapping that is required for spoken language acquisition.

By 4 months of age, infants' vocal experimentation expands to include a wider variety of sounds (Stark 1980). Some of these new sounds involve the lips as articulators (e.g., raspberries), but most involve changes in pitch or voice quality (e.g., squealing, song-like play, and whispers). Again, there is little evidence of supraglottal articulatory movement or of coordination between articulators in service of achieving a particular sound. There is, however, change over time within the course of a single vocalization. Specifically, infants will vary vocal fold tension and glottal closure during a vocal bout to achieve a variety of melodic patterns (Papoušek and Papoušek 1981; Hsu, Fogel, and Cooper 2000). They can also apparently exercise deliberate control over laryngeal movements, sufficient to imitate melodic contours that a parent produces (Papoušek and Papoušek 1989; Gratier and Devouche 2011). The continuous and deliberate changes in laryngeal posture to meet some acoustic target represents a significant step forward in the acquisition of speech. This stage also likely provides the foundation for the acquisition of language-specific intonation patterns.

At around 7 months of age, infants begin to coordinate phonation with opening and closing movements in the supraglottal vocal tract to produce speech-like sequences of proto-syllables; namely, regular alternations between consonant- and vowel-like sounds. This is the so-called canonical babbling stage (Stark 1980), the final pre-speech stage of vocal play, and a critical stage in the initial acquisition of articulatory timing and coordination (Davis and MacNeilage 1995; MacNeilage, Davis, and Matyear 1997; Oller 2000). Although babbling appears to replicate the basic phonological structure of spoken language, it is substantially different from speech in that the syllabic sequences are highly redundant (reduplicative); for example, *bababa*. An infant might also produce sequences with some variability in vowel height or consonantal manner changes (Davis and MacNeilage 1995), but do not expect to hear an infant produce a speech-like sequence that requires multiple place and manner changes (e.g., *bisogremu*).

The hallmark redundancy of babbling reflects immature control over the individual movements of the supraglottal articulators and over the coordination of their movements. The supraglottal articulator over which infants appear to first gain control is the jaw, as measured by trial-to-trial variability in movement. For example, Green, Moore, and Reilly (2002) showed that jaw movement during infant babbling and adult speech is similarly stable, but upper and lower lip movements are much more variable in infant babbling and in 2-year-old speech than in adult speech. Although there is no equivalent kinematic studies of tongue movement in infants and very young children, the assumption is that control over our most versatile speech articulator takes an especially long time to develop (Green and Nip 2011). What is also clear from kinematic studies is that coordinated supraglottal articulatory movement (i.e., of lip and jaw or of tongue and jaw) remains immature until early adolescence and continues to be refined through late adolescence (e.g., Walsh and Smith 2002; Cheng et al. 2007). This protracted development of speech motor skills has implications for the acquisition of temporal patterns at both the word and phrase level, as we will see below.

Early control over jaw movements coupled with the slow development of coordinated movement is consistent with the Frame/Content account of the sound patterns in babbling and first words (MacNeilage and Davis 1990; Davis and MacNeilage 1995; MacNeilage et al. 1997; see also MacNeilage, this volume, Chapter 16). In this account, the serial alternation of consonants and vowels in babbling and first words is driven mainly by jaw movement: an infant chooses a tongue position and then raises and lowers the jaw to achieve an alternation between vocal tract closure (consonant) and opening (vowel). When the tongue is advanced, the alternation yields the percept of an extended sequence of alternating front consonants and vowels (e.g., *dididi*); when retracted, a sequence of alternating back consonants and vowels (e.g., *gugugu*); and when in a neutral position, a sequence of labial consonants and central vowels (e.g., *bababa*). MacNeilage and Davis's work has shown that these co-occurrence patterns provide a good description of the perceived sound patterns of babbling. These same patterns also characterize early word productions (MacNeilage et al. 1997).

The similarities between patterns observed during the canonical babbling stage and first words strongly suggest that the motor skills gained during babbling provide the foundation for early attempts at producing words. These attempts in turn provide the foundation for the acquisition of fluent speech, and so for the acquisition of the complex temporal patterns that arise from multiple articulatory and linguistic influences.

17.3 First words

Typically developing children usually attempt their first words by 12 months of age. We acknowledge that the notion of "word" may strike some as a bit tricky. In particular, linguists will point out that word-like units in an agglutinative language, such as Finnish, are comprised of many individual units, each with a separate and unique meaning. In these languages, words are more akin to the English sentence than to the English word, which very often represents a correspondence between a sound pattern and an indivisible meaning. Yet, when we talk about first words in acquisition, the bias derived from our experience with English words is perfectly acceptable. Children – even Finnish children – do not attempt utterances that combine or layer individual meanings. Instead, they produce units that conventionalize a relationship between a sound pattern and a meaning, one that is recognized by the caregiver or others as communicative. So, for example, "uh-oh" counts as a word when learned and repeated in an appropriate context: a child drops a food item on the floor that he is supposed to eat, mom picks it up and hands it back to the child, child drops it again, and the game continues with "uh-oh"s all around. Vihman and colleagues (1985: 402–408), writing for linguists, carefully elaborate on this lay (and perhaps English-centric) understanding of a word. Some of their discussion is worth quoting because it also makes two central points more relevant to our interest in motor skill development; namely, that the sound patterns of first words are distant approximations of adult targets and that children utter these with intentionality. Vihman et al. (403) explain that:

> Before we credited a child with a spontaneous use of a word, we required that he or she produce a phonetic form that was a recognizable attempt at the adult word, given frequent child-reduction rules (cf. Ingram, 1974, 1979). In addition, the child had to use the word appropriately, with an apparently intentional meaning that was plausible in terms of the adult meaning or use of the word and commonly occurring child-semantic rules, such as over-extension of *doggy* to cats and other animals (cf. Clark, 1973).

With regard to "child-reduction rules," Vihman et al. (1985) are referring to processes like reduplication, fronting, and other simplifications. For example, a 12–14-month-old child is likely to produce "guk" for "duck" or "ditty" for "kitty," patterns that recall the redundancies and simplified syllable structure of babbling. Vihman and colleagues in fact carefully document the sound pattern similarities between babbled utterances and first words. Their study corpus was obtained from a set of 10 children, whom they observed and recorded weekly over a 7-month period starting when the children were just 9 months of age and only babbling. Vihman and colleagues found that individual children had specific sound preferences~~, which were manifest~~ in their babbled utterances. These same sound preferences then appeared in their first words. There were also consistent dependencies between first words and babbled utterances, when these overlapped in developmental time.

A more general way to think of the sound pattern similarities across the prespeech and early speech period is in terms of continuity. Vihman et al. (1985) were among the first to articulate the *continuity hypothesis* for speech sound acquisition: that babbling and first words reflect different developmental stages in the continuous acquisition of the sound patterns of language. Their work also helped to define the current view that the same immature motor skills that shape babbling patterns also shape the patterns observed in first words (Vihman 1996; MacNeilage et al. 1997; Davis, MacNeilage, and Matyear 2005; Nip, Green, and Marx 2009). Reduplicated forms, for example, can be produced by exploiting the open-close cycle of the jaw, requiring minimal intra-utterance movement of the other supraglottal articulators (MacNeilage et al. 1997; Davis et al. 2005). Similarly, the simpler syllable structures of early words reduce the number of supraglottal articulatory movements within a single open-close jaw cycle. As such, these structures stand in contrast to the more complex forms that adults produce, which require repositioning lip, tongue, and/or velum movement and coordinating these with the downward or upward movements of the jaw (see, e.g., Redford 1999).

In spite of the evidence that the sound patterns in babbling and first words are both shaped by the still limited speech motor skills of the young child, it is important to recognize that Vihman et al.'s (1985) definition alludes also to a critical difference between babbling and first words. Whereas babbling is content free and a feature of relaxed play, first words have meaning and are uttered with communicative intent. This difference suggests that an important *discontinuity* in the acquisition of speech and language occurs around a child's first birthday: the transition from speech-like vocal behavior that has no specific goal to behavior that is goal directed.

Goal-directed behavior requires that the to-be-achieved targets are represented in memory with a sufficiently detailed motor plan to be realized again and again in the same way. How are these targets represented in first word production? McCune and Vihman (1987) hypothesized that children might initially rely on the "vocal-motor schemes" established during babbling to produce first words, which we will refer to as schemas. These schemas are the acoustic-articulatory memories of the child's preferred sound patterns; that is, those patterns that the child produced over and over again during babbling. McCune and Vihman hypothesized that a child might select one of their preferred schemas to attempt a proximal word target, thus associating the schema with meaning. Different stored schemas could be associated with different meanings by perceptually matching the stored forms to ambient perceptual targets. The schemas may then serve as initial word representations. We propose that this is what happens, and that the schemas are then continually modified over developmental time in order to more and more closely match the ambient target. The refinement process results in more sophisticated speech motor skills. At the same time, different moments in the development of these skills are reflected in the revised schemas that guide behavior.

Children's extension of acoustically linked articulatory schemas to word production requires the deliberate selection of a stored representation to match a specific perceptual target in order to convey a specific concept in a specific moment in time. In this way, children's first words mark the transition from unplanned speech-like motor behavior to behavior that is guided by a plan. According to McCune and Vihman (1987), the relevant representation is essentially an abstract articulatory specification for the production of a sound pattern associated with meaning. This kind of representation recalls the gestural scores of Articulatory Phonology (Browman and Goldstein 1992), which are also thought of both as speech motor plans and as lexical representations. Assuming continuity in development, we suggest that the acoustically linked articulatory schemas of early childhood are the lexical representations activated during speech planning and execution throughout life. Again, what changes over developmental time is the extent to which these schemas code actions that enable the speaker to produce acoustics that match perceptual representations built up from the ambient language input.

The acoustically linked articulatory schemas that we imagine as the phonological aspect of lexical representation clearly include temporal information because they include information about sequential articulatory action. Consider, for example, the near minimal pairs that a child might produce in attempting the words *powder*, [pʰɑʊdɚ], and *spider*, [spɑɪdɚ]. The child would likely simplify the cluster in *spider* so that the two words would be produced as *powder* and *pider*. Written like this, one might think that the child produces the same "p" in both words; but this is not the case. Consistent with the perceptual input, *pider* will be rendered with an unaspirated /p/, [pɑɪdɚ], and *powder* with an aspirated /p/, [pʰɑʊdɚ] (see, e.g., Redford and Gildersleeve-Neumann 2007). That is, the representation that guides the child's speech action is a plan that the child has built, within the parameters of his or her motor abilities, to approximate a whole-word perceptual target. In order to achieve this best approximation, the plan must include detailed information about articulatory timing. In the present case, that would be information about the relative timing of voice onset after the release of stop closure. Given that laryngeal control is achieved relatively early and coupled with supraglottal articulatory action during the pre-speech phase of development, this is the kind of temporal information that the child could incorporate into an early lexical representation. In the next two sections, we consider the acquisition of temporal patterns that emerge above the level of a single word.

17.4. Multi-word utterances

In the previous section, we introduced the concept of a plan that guides speech action. Although this plan may initially be just a single schema, guiding the production of a single word, we hypothesize that the plan expands with mean length of utterance in development. Specifically, as children's linguistic and cognitive capacities grow, their default speech plan becomes a sequence of schemas, each associated with a single word. Multi-word speech plans are hypothesized to account for the fact that fluent speech is defined by a smooth temporal flow from one word to another, with strong junctures (e.g., pauses) occurring only to define or delimit conceptually coherent sequences of words. Although the structure and extent of the plan that governs continuous speech is an area of active research (see, e.g., Shattuck-Hufnagel, this volume, Chapter 19), we concur with the view in the adult literature that the plan extends over at least the length of one intonational phrase (Keating and Shattuck-Hufnagel 2002; Krivokapić 2007; Choe and Redford 2012). An intonational phrase is defined by a continuous intonational contour and is delimited by strong junctures, often pauses, at the beginning and end of its realization.

So, our proposal is that an intonational phrase defines the temporal extent of the speech plan in development, just as in adult speech. This proposal is supported by the characterization of early language production as holophrastic (see, e.g., Tomasello 2003: 36–40): it has long been observed that children at the one-word stage of language acquisition utter each word with a particular intonational contour and, often, the same word with different intonational contours. The intonationally inflected single word productions of early child language has led to the notion that, with every word, the child tries to convey – in some sense – the information of a sentence (= a holophrase). Whatever its relation to sentences, holophrastic speech is consistent with the view that speech action is guided by a plan defined by a coherent intonational contour. Clearly the number of words that are incorporated under this contour increases with developmental time. To wit, the child moves from the one-word stage to producing an average of two words per utterance, then three, then four, and so on (Brown 1973). The developmental increases in phrase length reflect a child's linguistic knowledge and increasing cognitive capacity, but also an ability to execute increasingly long and complex action sequences.

The remainder of this chapter addresses the acquisition of temporal patterns associated with fluent speech, that is, the execution of multi-word speech plan. More specifically, we review developmental changes in speech rate, and the acquisition of language rhythm. These two phenomena, rate and rhythm, together subsume all articulatory and linguistic influences on the complex temporal patterns of speech. The review of rate is meant to underscore the protracted development of speech motor skills, with increasing rate tied to advances in skill. The review of rhythm acquisition will be used to argue that some global temporal patterns emerge fortuitously with the development of speech motor skills while others are themselves the targets of acquisition.

17.4.1 Speech rate: Speaking versus articulation

Speech rate is measured as the number of syllables per second or words per minute that an individual produces. The clinical literature makes a distinction between speaking rate and articulation rate following Miller, Grosjean, and Lomanto (1984; see, e.g., Tsao and Weismer 1997; Hall, Amir, and Yairi 1999; Flipsen 2002; Tsao, Weismer, and Iqbal 2006): speaking rate includes pauses; articulation rate does not. Speaking rate is thought to reflect language planning processes; articulation rate, purer motor processes. However, the distinction is not totally clear cut. Speaking rate incorporates articulation rate, and articulation rate can be modified by the speaker to achieve different communicative goals.

Both speaking and articulation rate continue to increase throughout childhood and into early adolescence (Kowal, O'Connell, and Sabin 1975; Sabin et al. 1979; Smith, Sugarman, and Long 1983; Haselager, Slis, and Rietveld 1991; Walker et al. 1992; Hall et al. 1999; Flipsen 2002). Kowal and colleagues, who investigated speaking rate in 168 American-English-speaking children, present data from narrative tasks that indicate an increase in rate until 13 or 14 years old (8th grade), with the most substantial increases occurring between 5 and 8 years old (kindergarten through 2nd grade). For example, Kowal et al. (1975) report speaking rates of 2.15 syllables per second (SD = .75 syll/sec) in 5- and 6-year-old speech, 2.86 syllables per second (SD = .53 syll/sec) in 7- and 8-year-old speech, and 3.84 syllables per second (SD = .52 syll/sec) in 17- and 18-year-old speech. We have found similarly small increases in articulation rate over the same developmental period in a corpus of spontaneous narratives produced by 68 children and their parents. Our data indicate mean rates of 3.16 syllables per second (SD = 0.43 syll/sec) in 5-year-old speech, 3.49 syllables per second (SD = 0.51 syll/sec) in 7-year-old speech, and 4.13 syllables per second (SD = 0.67 syll/sec) in adult speech. Mean articulation rates in our sample of 6-year-old speech are 3.38 syllables per second (SD = 0.57 syll/sec), not significantly different from either the 5-year-old or 7-year-old rates. The articulation rates we find in our corpus are consistent with the developmental changes in articulation rates reported for Dutch-speaking school age children (Haselager et al. 1991) and for American-English-speaking pre-school age children (Walker et al. 1992; Hall et al. 1999; Flipsen 2002).

The parallelism between developmental increases in speaking rate and articulation rate suggests a correlation between the two. Robb and colleagues (2003) have confirmed this correlation in pre-school Australian-English-speaking children, age 2–4 years. Sabin et al. (1979) also found that speaking and articulation rates were significantly correlated in their data. In spite of this, they advocated for treating the two quite differently. Citing Lenneberg (1967), Sabin et al. argued that "speech rate is limited by the 'cognitive aspects of language' rather than simply (by) the physical ability to articulate speech" (46), and pointed to task-dependent differences in pause frequency and duration. Like Goldman-Eisler (1968) before them, Sabin et al. found higher rates of pausing when speakers retold a previously read narrative than when they produced spontaneous narratives based on cartoon sequences. As in Goldman-Eisler (1968), the task-dependent differences were attributed to different demands on cognitive processing. The idea is that certain kinds of language (e.g., spontaneous narratives) require more time to formulate than others (e.g., retold narratives), and that this processing time is reflected in the amount of pausing that occurs. Sabin and colleagues suggested that developmentally related changes in pausing could be understood in the same way. In particular, their idea was that children spend more time formulating an utterance than adults.

Sabin et al.'s (1979) measure of pause frequency and duration was calculated as a function of language produced (i.e., per 100 syllables). Such a measure leaves open the possibility that pauses are more frequent and longer in younger children's speech compared to older children's speech, not because of differences in formulation time, but because younger children produce shorter utterances on average than older children. When pause frequency is assessed as the number of pauses per unit time and pause duration in terms of absolute duration, the developmental effect on pause frequency and duration disappears (Redford 2013): children pause for as much time as adults when speaking, even though they produce less speech over time than adults. That said, the effect of spontaneous telling versus retelling on pause frequency and duration is as robust in children as in adults (ibid.). Children and adults also pause for more time before longer utterances than before shorter utterances, and before discourse markers and other conceptual junctures than elsewhere (ibid.). Overall, these findings support the idea of an association between pausing and language formulation, but not the idea that children require more time to formulate language. Instead, the findings suggest that pause frequency is negatively correlated with developmental changes in mean utterance length.

Changes in mean utterance length are usually attributed to the acquisition of syntax, with longer utterances indicative of more complex syntactic structures (see, e.g., Brown 1973; Tomasello 2003). But utterance length is also likely to be conditioned by speech motor skills and non-language cognitive factors. With respect to speech motor skills, it is clear that multi-word utterances require the ability to effectively instantiate planned actions, one after another and without interruption. The difficulty of this task increases with the size of the plan (see, e.g., Maassen and Terband, this volume, Chapter 15), which means that immature motor skills could limit the size of the plan. With respect to non-language cognitive factors, it seems likely that working memory is involved in speech planning, which means that working memory capacity could also limit the size of the plan. One could even imagine an interaction between the "physical ability to articulate speech" and working memory capacity: slower articulation rates increase the amount of time that a plan must be held in working memory, which could lead either to the premature decay of elements not yet executed in the plan or to an accommodation in the planning process, such that fewer elements are prepared for execution at any one time. Although even a partial review of the working memory literature is outside the scope of this chapter, this novel hypothesis for developmentally related changes in mean length of utterance is consistent with theories that emphasize temporal constraints (e.g., processing speed and memorial decay) over task switching (i.e., executive control) to explain developmental changes in working memory capacity (cf. Towse, Hitch, and Hutton 1998; Gathercole et al. 2004). It also aligns well with the robust finding of a close relationship between speech rate and number of words recalled in studies of child and adult working memory (see Hitch and Towse 1995).

17.4.2 Articulation rate and speech motor skills

Rapid, stable, and efficient execution of complex movement sequences requires extensive practice, and that performance can continue to improve over many years (Schmidt and Lee 2005). It is likely for this reason that children's articulation rates are slower than adults' until early adolescence.

Every speech posture is achieved by the precise spatial-temporal coordination of articulators to achieve an acoustic target – a specific phone or segment – that is embedded in a sequence of such targets, whose order defines the sound/movement pattern of a particular word. Although motor constraints on coordination are particularly apparent in the production of early sound patterns and first words, kinematic evidence suggests that adult-like performance is not achieved until adolescence (e.g., Walsh and Smith 2002; Cheng et al. 2007). In the acoustic domain, the protracted acquisition of postural control is reflected in the slow decrease of segmental durations over developmental time (Smith 1978; Kent and Forner 1980; Lee, Potamianos, and Narayanan 1999). For example, Lee et al. (1999), who report data from 56 adults and 436 children between the ages of 5 and 17 years, find that segmental durations are longer in younger children's speech compared with older children's speech, and that adult norms are not attained until age 12. Recall that Kowal et al.'s (1975) data indicate that adult speaking rates (pause inclusive) are not attained until age 13 or 14. Thus, the Lee et al. results are in strikingly parallel to the Kowal et al. findings, underscoring the important contribution of articulation to speaking rate.

The longer segmental durations of child speech compared to adult speech are frequently discussed in conjunction with temporal variability, which is also greater in child speech compared to adult speech (Tingley and Allen 1975; Kent and Forner 1980; Smith et al. 1983; Lee et al. 1999). Temporal variability refers to variation in the duration of a linguistic unit (segment, syllable, or word) across repetitions of the unit when other linguistic, cognitive, and social factors are held constant. Of course, variation in duration will be proportional to the mean, especially when measured in standard deviations. For this reason, Kent and Forner (1980) proposed that temporal variability may be larger in child speech than in adult speech only because mean durations are larger. Smith et al. (1983) and Smith (1992) referred to this proposal as the statistical artifact hypothesis, and argued against it.

In a first study, Smith et al. (1983) asked 5-, 7-, and 9-year-old children as well as a group of adults to repeat a sentence multiple times at a normal speech rate (i.e., their default rate) and at slow and fast rates. Phrase and syllable durations were found to vary systematically as a function of age and speech rate, but there was no interaction between these factors. Next, Smith et al. investigated temporal variability as a function of speech rate using a mean normalized measure of variation, namely, the coefficient of variation. The results were that temporal variability was higher in 5- and 7-year-old children's speech than in 9-year-old children's speech. Older children's speech was in turn more variable than adults' speech. Temporal variability was also higher at slower and faster speaking rates than at default speaking rates across speakers regardless of age, suggesting independence between absolute duration and variability. In a later study, Smith (1992) directly examined the relationship between duration and temporal variability in children's speech, aged 2–9 years. He found that in spite of overall decreases in both duration and variability with age, intra-subject correlations between the measures were relatively low. The combined results suggest that developmental decreases in temporal variability, though correlated with decreases in segmental duration, reflect the development of different motor skills than those required to achieve a specific linguistic target. The distinction may be one of coordinating independent articulators through time (i.e., articulatory timing) versus independently controlling individual articulators to achieve a target (i.e., articulatory coordination).

The explanation that different motor control processes underlie changes in variability and duration also makes sense of the finding that variability is higher at non-default speech rates (Smith et al. 1983). Whereas individual targets do not change with speech rate, voluntary manipulations of rate entail changes in the sequential timing of articulatory action, which also has consequences for articulatory coordination. For example, we know from studies of intra-speaker rate control in adults that fast speech is achieved primarily through vowel reduction, and especially through the temporal compression of stressed vowels (Gay 1978, 1981). These changes may or may not result from direct control. Direct control over rate implies that a clock-like mechanism, extrinsic to the representations underlying speech production, drives the rate with which each sound target embedded in a sequence is executed. Indirect control over rate implies that changes follow from the manipulation of parameters within the speech production system (see also Fowler 1980).

Whether control is direct or indirect, the child must acquire a strategy for executing speech movements more rapidly in time. The adult literature makes clear that at least two strategies are available (e.g., Ostry and Munhall 1985; Adams, Weismer, and Kent 1993; Matthies et al. 2001). Some speakers decrease articulatory displacement (e.g., increase damping) at faster rates of speech, resulting in incomplete target attainment. Others increase movement velocity (e.g., increase stiffness), thereby maintaining targets even at fast rates of speech. Children presumably acquire the ability to manipulate both displacement and velocity, even if they choose to manipulate just one to effect changes in speech rate. Children also presumably gain more and more fine-grain temporal control over changes in articulatory displacement and velocity during the protracted development of speech motor skills, and use this control to meet communicative demands. It is this modulatory control, along with practice of basic articulatory timing patterns, that allows for the emergence of prosodically related temporal patterns.

17.4.3 Prosodically related temporal patterns

Insofar as the speech plan guides the production of multi-word utterances, it encodes both the order of words in a phrase as well as the articulatory timing information that is part of the phonological specification of the word. Whereas just this information may be sufficient to realize language-specific phonetic and phonemic patterns at different rates, it is not sufficient to account for prosodically related temporal patterns, which introduce variable lengthening and shortening (reduction) unequally across the phrase. To make this point explicitly, let us consider a few seconds of speech produced by a 5-year-old American-English-speaking boy. The speech was collected in a narrative task that children completed with their caregiver. This task used wordless picture books (Mercer Mayer's frog stories) to elicit fluent, structured spontaneous speech (see Redford 2013 for task details). The transcribed extracted speech sample shown below has been rendered in normal orthography and syllabified (txt tier). Slashes indicate pauses. The numbers below each syllable show approximate syllable and pause durations in milliseconds (dur tier). The various notations above the text show a trained analyst's judgment of weak ("x") and strong ("X") prominences at the level of the phrase and of weak (")") and strong ("))") prosodic phrase boundaries (pros tier; see Breen et al. 2012 for details regarding labeling).

pros:		x))			))?	X					X?))	)?
txt:	/	and	/	he	was	try-	ing	to	get	the	frog	but
dur:	1530	565	789	218	154	495	161	91	211	80	471	299
pros:		X?	)?				X?	)			x))	
txt:	he	could-	n't	be's	it	just	jumped	up	to	the	branch	/
dur:	279	175	178	271	131	349	366	198	196	83	323	1944
pros:	x))			)?	X		x	))				x
txt:	and	/	he	was	ver-	ry	mad	at	/	<teh>	/	so
dur:	566	1069	73	176	272	117	258	151	837	270	621	327

The snippet illustrates well the temporal variability of naturally produced speech in relation to the phrase-level coding of rhythm. Note, for example, that the stressed syllable of the main verb "try" is over three times as long as the auxiliary that precedes it, even though both syllables are comprised of just three phonemes each. Note also that some lexical items, such as "he" and "to," are realized with substantially different absolute durations in the same grammatical context, while another, "was," is realized with substantially similar durations in two different grammatical contexts. Finally, note the relatively long syllables in the vicinity of a boundary judgment, and that only these relatively longer syllables are heard as most prominent.

How are these prosodically related temporal patterns acquired? To answer this question, let us first introduce the notion of rhythm more completely.

17.4.4 What is speech rhythm?

Rhythm is best defined with respect to alternations in prominence (Liberman and Prince 1977; Hayes 1984), a perceptually-based linguistic construct (Terken and Hermes 2000; Arvaniti 2009). The acoustic-phonetic study of rhythm, which had been more or less abandoned after extensive arguments about the psychological reality of isochrony (equally timed stresses), was reinvigorated when Ramus, Nespor, and Mehler (1999) and Grabe and Low (Low, Grabe, and Nolan 2000; Grabe and Low 2002) introduced a few simple, interval-based measures that seemingly differentiated languages on the basis of their perceived rhythm pattern (see Cummins, this volume, Chapter 8 for a discussion of these measures and their historical context). The researchers who introduced these measures interpreted their results to support the so-called rhythm class hypothesis, which provided the original impetus for the isochrony debate. The rhythm class hypothesis states that languages belong to one of three rhythm classes: stress-timed, syllable-timed, and mora-timed. Stress-timed languages, such as English and Russian, have lexical stress (e.g., the verb, to *reCORD* versus the noun, *REcord*), syllable- and mora-timed languages do not. Mora-timed languages make extensive use of gemination (phonologically long consonants and vowels), stress- and syllable-timed languages do not.

Although dissatisfaction with the interval-based measures of rhythm abounds (see, e.g., Arvaniti 2009), the measures clearly capture something about speech rhythm, which is evident even in our snippet above. Duration is also the key correlate of lexical stress in English (Terken and Hermes 2000; Kochanski et al. 2005). For example, the unstressed vowel /ə/ in the verb "record" is likely to be shorter than the stressed vowel /ɛ/ in the noun "record" when produced by the same speaker under the same conditions. The stressed vowel in both the verb and noun is also likely to be significantly longer than the unstressed vowel in the same word. The duration contrast is due to the fullness of the stressed vowel and to the short, centralized quality of the vowel in the unstressed syllable. This quality of unstressed vowels in stress-timed languages like English and Russian has been referred to as "reduced." The phonological notion of reduction assumes that an underlying full vowel is transformed at some point during the production process, which is why it surfaces as a reduced version of itself. Whereas such a transformation may have occurred over historical time, we do not think that it occurs during production. In our view, the central quality and length of the unstressed vowel is available in the acoustic waveform and is thus the target of acquisition. As a target of acquisition, it becomes represented directly as part of the articulatory timing pattern of the word when children have gained the motor skills necessary to produce consecutive vowels of different lengths.

In a stress-timed language, lexically-based differences in vowel length co-occur with complex syllables structures. This means that stress-timed languages also allow for long and short sequences of consonants intervocalically. By contrast, syllable structures are simpler in syllable- and mora-timed languages. Thus, stress-timed languages are differentiated from syllable- and mora-timed languages by higher variability in vocalic and consonantal interval durations and relatively lower proportions of total vocalic to total consonantal interval durations.

17.4.5 Rhythm acquisition

A few studies have applied interval-based rhythm measures to child speech and language (Grabe, Post, and Watson 1999; Bunta and Ingram 2007; Sirsa and Redford 2011; Payne et al. 2012). These studies generally confirm an observation that Allen and Hawkins (1978) originally made; namely, that children's speech is more syllable-timed than adult speech. For example, Grabe et al. (1999) and Bunta and Ingram (2007) find that children acquire the distinctive rhythm pattern of a syllable-timed language earlier than they do a stress-timed language. Even so, Payne and colleagues showed that language-specific rhythm differences can be detected in early child speech. They also showed that, though more vocalic than adult speech, children's speech is also not exactly syllable-timed.

Payne et al. (2012) used several interval-based rhythm measures to compare the rhythm patterns produced by 27 children acquiring either English, Catalan, or Spanish in a cross-sectional study of 2-, 4-, and 6-year-old speech. For comparative purposes, they also analyzed the speech produced by the children's parents. The findings for adult speech were that English and Spanish were well distinguished by all measures: English vocalic and consonantal interval durations were more variable than in Spanish, and the proportion of vowel to total duration was lower in English than in Spanish. Catalan tended to cluster with Spanish, but was more similar to English on certain measures. The findings for child speech mirrored the findings for adult speech and the cross-linguistic differences were evident by age 2. Nonetheless, children's speech did differ from adult speech along a number of dimensions across all three languages. In keeping with the idea that child speech is more syllable-timed than adult speech, the proportion of vowel to total duration was higher and variability in vowel durations was lower in child compared to adult speech, though this difference disappeared by age 6. Children's speech was also characterized by more variability in consonant durations than adult speech, and this difference persisted even at age 6.

In discussing their findings, Payne et al. (2012) suggested that persistent differences in the rhythms of child and adult speech might reflect an interaction between developing phonological and phonetic abilities. In particular, they suggested that the finding of higher variability in consonant durations might emerge from simplifications of syllable structure as well as from immature postural control and sequencing abilities. Insofar as cluster simplification and other processes that change syllable shape also have their origins in immature speech skills, we might conclude that speech rhythm is influenced by many of the same factors that account for slower speech rates in children compared to adults.

Whereas Payne et al. (2012) concentrate on the finding of cross-linguistic differences in the speech rhythms of very young children, Grabe et al. (1999) concentrate on the differential rate at which different language rhythms are acquired. Grabe et al. investigated speech of 6 mother–child dyads. Half of the dyads were English speakers and half were French speakers. Children in both groups were 4 years of age. Using a variety of temporal measures, including the interval-based measure of sequential vowel durations (PVI-V), Grabe et al. found that the French-speaking children had acquired the syllable-timed pattern of their language, but the English-speaking children had not acquired the stress-timed pattern of their language. A study by Bunta and Ingram (2007), though focused on the acquisition of rhythm by Spanish-English bilinguals, also indicated that Spanish monolinguals acquire the syllable-timed rhythm pattern of their language prior to English monolinguals. In particular, the rhythm patterns of Spanish-speaking children did not differ from Spanish-speaking adults at 4½ years of age, but those of English-speaking children did differ from English-speaking adults at this age. We found similar differences in the rhythm patterns of 5- and 8-year-old monolingual English speakers (Sirsa and Redford 2011). Younger children were found to have lower normalized vocalic PVI scores on average than older children.

Overall, studies on the acquisition of rhythm suggest that syllable-timed patterns are acquired before stress-timed patterns, and that the English stress-timed pattern may not be fully acquired until age 6 or 7. Whereas it is likely that the vocalic nature of early child speech can be accounted for largely in terms of simplified syllable structures, these simplifications originate from immature motor skills, codified by the stored acoustically-linked articulatory schemas that guide speech action. Similarly, the higher variability in consonant durations and lower nPVI-V scores in 5-year-old speech are also likely due to immature motor skills, albeit at the level of on-line articulatory timing control.

17.4.6 Rhythmic groupings and speech motor skills

One might expect that the protracted acquisition of stress-timing can be explained by the delayed acquisition of lexical stress. However, several acoustic-phonetic studies on the implementation of lexical stress in English suggest instead that stress is acquired fairly early (Pollock, Brammer and Hageman 1993; Kehoe, Stoel-Gamon, and Buder 1995; Schwartz et al. 1996). For example, Kehoe et al. (1993) found no measurable differences in the acoustic marking of stress in familiar words produced by 2-year-old children (+/– 6 months) and adults. Pollock et al. (1993) also found that 2-year-old children use duration to distinguish stressed and unstressed syllables, but the other correlates of lexical stress, fundamental frequency and intensity, are not used appropriately until age 3. Still, if the duration patterns associated with lexical stress are acquired so early, then the protracted acquisition of a stress-timed rhythm, as measured by changes in the sequential variation in vowel durations, is unlikely to be explained in terms of the acquisition of lexical stress.

It turns out that in spite of the early marking of lexical stress, 2-year-olds and adults do not produce unstressed syllables in the same way. Kehoe and colleagues found that unstressed syllables were 55% longer in children's speech than in adults' speech, even though the relative difference between stressed and unstressed syllable durations was not different. Schwartz et al. (1996) found that even the relative difference of stressed and unstressed syllables within a word was smaller in 2-year-old and adult speech, and attributed this smaller difference to relatively longer unstressed syllables in child speech compared to adult speech. Pollock et al. (1993) found that unstressed syllable durations decreased over developmental time, but stressed syllable durations stayed the same. Relatedly, kinematic data indicates that even older children (4+ years of age) are more variable in their production of unstressed syllables compared to stressed syllables (Goffman 1999), suggesting that reducing unstressed syllables may require more advanced motor skills than the producing stressed syllables. Ballard et al. (2012) have recently advanced a similar argument based on acoustic evidence, showing that 7-year-olds continue to deviate from adults in their realization of weak-strong syllable sequences. The suggestion that shorter durations are especially hard for children to realize is consistent with the notion that efficient and reduced movements are characteristic of expert motor control (e.g., Green et al. 2000, 2002), and with data showing that young children make relatively larger amplitude speech movements than adults (Riely and Smith 2003).

Unstressed syllable (vowel) reduction is also a feature of phrasal stress patterns. Function words like "a" and "the" in English, when cliticized (attached) to a following noun, are particularly reduced. This is evident even from the short snippet of speech presented in section 17.4.3 above; there, "the" is less than one-third the length of the following monosyllabic nouns it determines. The perceptual effect of cliticization is the creation of a so-called clitic group or prosodic word, such that a phrase like "The boy walked the dog" with five orthographic words is produced with just three prosodic words (i.e., [the boy] [walked] [the dog]). Allen and Hawkins (1978) may have been the first to observe that young English-speaking children do not reduce function words to the same extent as adults. Further, they suggested that it was the less reduced function words of child speech that contributed to the percept that children's spoken English is more syllable-timed than adult English. Allen and Hawkins's evidence was largely transcription-based, but Goffman (2004) provided some acoustic and kinematic evidence for the idea. Goffman's interest was in whether or not the morphosyntactic status of the unstressed syllable affected its production. Accordingly, she compared lexically unstressed syllables to function word syllables in child and adult speech. She found that adults reduce unstressed syllables in a function+content word phrase (e.g., "a *bab*") more than they do in a disyllabic word (e.g., "a*bab*"). Children, between 4 and 7 years of age, did not differentiate between unstressed syllables as a function of morphosyntactic status.

Further evidence for the idea that the slow acquisition of function word vowel reduction provides the key to explaining the late acquisition of English rhythm comes from our study of rhythm acquisition in 5- and 8-year-olds (Sirsa and Redford 2011). As previously noted, we found that, on average, 5-year-old speech had lower nPVI-V scores than 8-year-old speech. In order to understand why this might be, we took a number of more specific temporal measures on the children's speech. These included a measure of the relative duration of lexically stressed and unstressed vowels, a measure of the relative duration of function and content word vowels in determiner noun phrases with monosyllabic nouns, and a measure of phrase-final lengthening. We then used these measures to predict nPVI-V scores and found that only the measure of function-to-content word vowel duration accounted for a significant amount of the variance: higher function-to-content word vowel duration ratios (i.e., less reduced function word vowels) predicted lower nPVI-V scores. The relationship between function word vowel reduction in determiner noun phrases and adult-like rhythm production was very strong; it accounted for 46% of the variance in nPVI-V scores.

In sum, unstressed vowel reduction is acquired slowly, perhaps because smaller, faster speech movements are a feature of motor expertise, which children only develop over a long period of time. The unstressed vowels of function words are even more reduced than the unstressed vowels of di- or multisyllabic content words in adult speech. This could be why adult-like function word vowel reduction is acquired especially late and contributes the most to explaining the immature English rhythm patterns of school-age children, as measured by the variability in sequential vowel durations.

17.4.7 Temporal modulation in service of meaning

If unstressed vowel reduction is tied to the development of speech motor skills, as we argue above, then the prosodic words that result from the reduction of function words adjacent to content words are not themselves the target of acquisition, but instead emerge fortuitously with faster, more fluent speech. That said, the semantically light nature of the specific items that are reduced suggests that motor skills alone cannot provide a complete picture of prosodic word acquisition. Before concluding this chapter, we must underscore that meaning also matters to the acquisition of temporal patterns. For example, children must learn that only certain syllables within a word and certain lexical items within a phrase can be reduced without sacrificing speech intelligibility. The role of meaning in the acquisition of temporal patterns is even more evident when we consider lengthening processes.

In English and many other languages, lengthening is one of several acoustic cues used to mark focus (a grammatical category related to information structure) and utterance-final boundaries. These types of lengthening are known as accentual lengthening and phrase-final lengthening, respectively. The domain of accentual lengthening in English is primarily the stressed syllable (Turk and White 1999), but lengthening effects are also observed at the prosodic word level

(Turk and Sawusch 1997; Turk and White 1999). In the snippet presented in section 17.4.3, accentual lengthening occurred on the words “trying,” “jumped,” and “very.” Note that the length of the stressed syllables in these words was especially long relative to the surrounding syllables. The domain of phrase-final lengthening may be somewhat less extensive: lengthening effects are especially notable on ultimate syllables of the phrase, whether these are lexically stressed or not (Turk and Shattuck-Hufnagel 2007). Lengthening effects can also extend to non-final syllables in phrase-final multisyllabic words, but only if these are lexically stressed (ibid.).

Even though there is some evidence that pre-speech vocalizations replicate accentual and phrase-final lengthening patterns of the language (de Boysson-Bardies, Sagart, and Durand 1984; Robb and Saxman 1990), it is equally clear that their semantic-pragmatic functions must also be acquired. Acquiring these functions necessitates coordinating the modulation of F0 with patterns of lengthening. Although there is very little acoustic-phonetic study on the acquisition of focus marking, some careful acoustically-based transcription work by Ajou Chen (Chen and Fikkert 2007; Chen 2011a) suggests that, at a minimum, the intonational marking of focus is not fully developed in Dutch-speaking children until after age 8. Chen (2011b) argues that younger children, aged 2 and 3, have trouble just with the phonetic realization of the complex tonal patterns associated with focus marking; and that older children struggle with the form-to-function mapping, confusing topic and focus marking. As for phrase-final lengthening, there is some indication that it may originate in physiological constraints (Robb and Saxman 1990), but Snow (1994) has shown that its linguistic aspect is not fully acquired until three months after the onset of combinatorial speech; a finding he then uses to highlight the challenges inherent to coordinating supraglottal articulatory timing and F0 patterns.

In addition to marking focus and utterance boundaries, lengthening is also used to indicate utterance-internal junctures. Speakers’ implicit knowledge of this function is demonstrated when they are asked to disambiguate a potentially ambiguous sequence. For example, the most natural way to disambiguate the scope of “old” in the sentence “The old men and women stayed home” is to lengthen “men” relative to “and women,” assuming the speaker wants to indicate that the scope of “old” is restricted to “men” (Lehiste 1973). Lengthening is also used to mark ambiguous word boundaries. For example, Christie (1977) showed that listeners use consonantal duration to distinguish between minimal pair sentences such as “help u*s n*ail” versus “help a *sn*ail.” Redford and Gildersleeve-Neumann (2007) found that pre-school children have trouble instantiating this specific cue to word boundaries. We compared the production of /s/+sonorant and /s/+stop offset-onset and onset cluster sequences (*this nail* versus *bitty snail*; *nice top* versus *I stop*) in 20 pre-school children, aged 3 and 4 years, with their parents’ production. Whereas all children produced the primary cue to juncture in the /s/+stop sequences – the presence versus absence of stop aspiration – even 4-year-olds could not reliably distinguish singleton sonorant onsets from /s/+sonorant clusters in two-word phrases. This differential acquisition of the juncture cues is in line with the distinction in this chapter between articulatory timing patterns at the level of the word versus those at the level of the phrase: the allophonic cue to juncture (+/− aspiration) is a pattern that is acquired and represented with the lexical items; the lengthening pattern, by contrast, requires focal control over lengthening at a boundary.

In sum, certain aspects of rhythm may emerge with the extensive practice required to efficiently execute extended movement sequences (i.e., multi-word plans) and from the development of control over the parameters that affect movement amplitude and velocity. However, when phrase-level temporal patterns are tied to meaning, they become – like words – the targets of acquisition. The challenges inherent to the acquisition of meaningful, phrase-level temporal patterns lies first in their coordination with separately generated intonational patterns, and second in their context-dependency. Unlike the patterns that guide word production, phrase-level temporal patterns cannot be learned and stored directly. Instead, they must be abstracted based on an understanding of the semantic-pragmatic context and generated based on a plan that marks out specific stretches as requiring a change in the parameters affecting movement amplitude and velocity, and thus time.

17.5 Summary

The present chapter examined the role that motoric factors play in the acquisition of the complex temporal patterns that characterize spoken language. We began by reviewing the motoric constraints that so clearly influence pre-speech vocalizations and early word production. In keeping with the continuity hypothesis, we proposed that the motor constraints that are evident in pre-speech also influence the later acquisition of temporal patterns at every level of linguistic analysis. However, an important discontinuity between pre-speech and speech was also noted; namely, the transition from vocal play and imitation to concept-driven communication. This transition marks the development of a speech plan, which guides speech action and represents what the child has learned about the structure of language. The temporal patterns that are produced from the first word stage onward thus reflect an interaction between motor skill development and the representation of language. This interaction presents a significant challenge, frequently overlooked by those who study speech and language acquisition: the challenge of distinguishing variance in performance due to immature motor skills from variance due to immature linguistic representations. Our review of rate and rhythm was motivated by this challenge and by the hypothesis that the acquisition of linguistic representation interacts with motor skill development. For example, we suggested that increases in the size of the speech plan may depend in part on developmental changes in the ability to rapidly attain sequences of different articulatory postures. We also argued that rhythm, though a linguistic construct, is not phonology-driven. Instead, we proposed that practice with different articulatory timing routines and the development of control over increasing movement efficiency is manifested as age-related differences in the realizations of stressed and unstressed syllables at the word and phrase levels. Increasing movement efficiencies over developmental time results ultimately in the emergence of prosodic words. Finally, we acknowledged a critical role for meaning in the acquisition of temporal patterns. In addition to encoding inter- and intra-articulatory timing at the level of the (stored) word, the speech plan must specially mark stretches in the plan for focal changes in timing so that context-dependent temporal patterns are appropriately realized.

NOTES

1 This work was supported by Award Number R01HD061458 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely my responsibility and does not necessarily reflect the views of NICHD. I am grateful to Zahra Foroughifar, Jeffery Kallay, and Paul Olejarczuk for reading and commenting on an early version of the chapter.

REFERENCES

Adams, Scott G., Gary Weismer, and Raymond D. Kent. 1993. Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research* 36: 41–54.

Allen, George D. and Sarah Hawkins. 1978. The development of phonological rhythm. In Alan Bell and Joan Bybee Hooper (eds.), *Syllables and Segments*, 173–185. Amsterdam: North-Holland Publishing.

Arvaniti, Amalia. 2009. Rhythm, timing, and the timing of rhythm. *Phonetica* 66: 46–63.

Ballard, Kirrie J., Danica Djaja, Joanne Arciuli, Deborah G.H. James, and Jan van Doorn. 2012. Developmental trajectory for production of prosody: Lexical stress contrastivity in children ages 3 to 7 years and in adults. *Journal of Speech, Language, and Hearing Research* 55: 1822–1835.

Breen, Mara, Laura C. Dilley, John Kraemer, and Edward Gibson. 2012. Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory* 8: 277–312.

Browman, Catherine P. and Louis M. Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica* 49: 155–180.

Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Bunta, Ferenc and David Ingram. 2007. The acquisition of speech rhythm by bilingual Spanish- and English-speaking 4- and 5-year-old children. *Journal of Speech, Language, and Hearing Research* 50: 999–1014.

Chen, Ajou. 2011a. Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language* 38: 1055–1083.

Chen, Ajou. 2011b. The developmental path to phonological focus-marking in Dutch. In S. Frota, E. Gorka, and P. Prieto (eds.), *Prosodic Categories: Production, Perception and Comprehension*, 93–109. Dordrecht: Springer.

Chen, Ajou and Fikkert, Paula. 2007. Intonation of early two-word utterances in Dutch. In J. Trouvain and W.J. Barry (eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)*, 315–320.

Cheng, Hei Yan, Bruce E. Murdoch, Justine V. Goozée, and Dion Scott. 2007. Electropalatographic assessment of tongue-to-palate contact patterns and variability in children, adolescents, and adults. *Journal of Speech, Language, and Hearing Research* 50: 375–392.

Choe, Wook Kyung and Melissa A. Redford. 2012. The distribution of speech errors in multi-word prosodic units. *Laboratory Phonology* 3: 5–26.

Christie, William M. 1977. Some multiple cues for juncture in English. *General Linguistics* 17: 212–222.

Clark, Eve. 1973. What's in a word? On the child's acquisition of semantics in his first language. In T.E. Moore (ed.), *Cognitive Development and the Acquisition of Language*, 65–110. New York: Academic Press.

Davis, Barbara L. and Peter F. MacNeilage. 1995. The articulatory basis of babbling. *Journal of Speech, Language, and Hearing Research* 38: 1199–1211.

Davis, Barbara L., Peter F. MacNeilage, and Christine L. Matyear. 2005. Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. *Phonetica* 59: 75–107.

de Boysson-Bardies, Bénédicte, Laurent Sagart, and Catherine Durand. 1984. Discernible differences in the babbling of infants according to target language. *Journal of Child Language* 11: 1–15.

Flipsen, Peter. 2002. Articulation rate and speech-sound normalization failure. *Journal of Speech, Language, and Hearing Research* 46: 724–737.

Fowler, Carol. 1980. Coarticulation and theories of extrinsic timing. *Journal of Phonetics* 8: 113–133.

Gathercole, Susan E., Susan J. Pickering, Benjamin Ambridge, and Hannah Wearing. 2004. The structure of working memory from 4 to 15 years of age. *Developmental Psychology* 40: 177–190.

Gay, Thomas. 1978. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63: 223–230.

Gay, Thomas.1981. Mechanisms in the control of speech rate. *Phonetica* 38: 148–158.

Goffman, Lisa. 1999. Prosodic influences on speech production in children with specific language impairment and speech deficits. *Journal of Speech, Language, and Hearing Research* 42: 1499–1517.

Goffman, Lisa. 2004. Kinematic differentiation of prosodic categories in normal and disordered language development. *Journal of Speech, Language, and Hearing Research* 47: 1088–1102.

Goldman-Eisler, Freida. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. London: Academic Press.

Grabe, Esther and Ee Ling Low. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7: 515–546.

Grabe, Esther and Paul Warren. 1995. Stress shift: Do speakers do it or do listeners hear it? *Papers in Laboratory Phonology* 4: 95–110.

Grabe, Esther, Brechtje Post, and Ian Watson. 1999. The acquisition of rhythmic patterns in English and French. In *Proceedings of the 14th International Congress of Phonetic Sciences, (ICPhS-99)*, 1201–1204.

Gratier, Maya and Emmanuel Devouche. 2011. Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology* 47: 67–76.

Green, Jordan R. and Ignatius S.B. Nip. 2010. Some organization principles in early speech development. In B. Maassen and P. van Lieshout (eds.), *Speech Motor Control: New Developments in Basic and Applied Research*, 171–188. Oxford: Oxford University Press.

Green, Jordan R., Christopher A. Moore, and Kevin J. Reilly. 2002. The sequential development of lip and jaw control for speech. *Journal of Speech, Language, and Hearing Research* 45: 66–79.

Green, Jordan R., Christopher A. Moore, Masahiko Higashikawa, and Roger W. Steeve. 2000. The physiological development of speech motor control: Lip and jaw coordination. *Journal of Speech, Language, and Hearing Research* 43: 239–255.

Hall, Kelly D., Ofer Amir, and Ehud Yairi. 1999. A longitudinal investigation of speaking rate in preschool children who stutter. *Journal of Speech, Language, and Hearing Research* 42: 1367–1377.

Haselager, G.J.T., I.H. Slis, and A.C.M. Rietveld. 1991. An alternative method of studying development of speech rate. *Clinical Linguistics and Phonetics* 5: 53–63.

Hayes, Bruce. 1984. The phonology of rhythm in English. *Linguistic Inquiry* 15: 33–74.

Hitch, Graham J. and John N. Towse. 1995. Working memory: What develops. In F.E. Weinert and W. Schneider (eds.), *Memory Performance and Competencies: Issues in Growth and Development*, 3–21. Mahwah, NJ: Lawrence Erlbaum Associates.

Hsu, Hui-Chin, Alan Fogel, and Rebecca B. Cooper. 2000. Infant vocal development during the first 6 months: Speech quality and melodic complexity. *Infant and Child Development* 9: 1–16.

Ingram, David. 1974. Phonological rules in young children. *Journal of Child Language* 1: 49–64.

Ingram, David. 1979. Phonological patterns in the speech of young children. In P. Fletcher and M. Garman (eds.), *Language Acquisition*, 133–148. Cambridge: Cambridge University Press.

Keating, Patricia and Stefanie Shattuck-Hufnagel. 2002. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics* 101: 112–156.

Kehoe, M., Carol Stoel-Gammon, and Eugene H. Buder. 1995. Acoustic correlates of stress in young children's speech. *Journal of Speech and Hearing Research* 38: 338–350.

Kent, Raymond D. and L.L. Forner. 1980. Speech segment durations in sentence recitations by children and adults. *Journal of Phonetics* 8: 157–168.

Kochanski, Greg, Esther Grabe, John Coleman, and Burton Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America* 118: 1038–1054.

Kowal, Sabine, Daniel C. O'Connell, and Edward J. Sabin. 1975. Development of temporal patterning and vocal hesitations in spontaneous narratives. *Journal of Psycholinguistic Research* 4: 195–207.

Krivokapić, Jelena. 2007. Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics* 35: 162–179.

Kuhl, Patricia K. and Andrew N. Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America* 100: 2425–2438.

Lee, Sungbok, Alexandros Potamianos, and Shrikanth Narayanan. 1999. Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105: 1455–1468.

Lehiste, Ilse. 1973. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America* 54: 1228–1234.

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.

Liberman, Mark and Allen Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249–336.

Low, Ee Ling, Esther Grabe, and Francis Nolan. 2000. Quantitative characterisations of speech rhythm: "Syllable-timing" in Singapore English. *Language and Speech* 43: 377–401.

MacNeilage, Peter F. and Barbara L. Davis. 1990. Acquisition of speech production: Frames, then content. In Marc Jeannerod (ed.), *Attention and Performance XIII: Motor Representation and Control*, 453–476. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacNeilage, Peter F., Barbara L. Davis, and Christine L. Matyear. 1997. Babbling and first words: Phonetic similarities and differences. *Speech Communication* 22: 269–277.

Maratos, Olga. 1973. The origin and development of imitation in the first six month of life. PhD thesis, University of Geneva.

Matthies, Melanie, Pascal Perrier, Joseph S. Perkell, and Majid Zandipour. 2001. Variation in anticipatory coarticulation with changes in clarity and rate. *Journal of Speech, Language, and Hearing Research* 44: 340–353.

McCune, Lorraine and Marilyn Vihman. 1987. Vocal motor schemes. *Papers and Reports on Child Language Development* 26: 72–79.

Meltzoff, Andrew N. and M. Keith Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198(4312): 75–78.

Miller, Joanne L., François Grosjean, and Concetta Lomanto. 1984. Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica* 41: 215–225.

Nagy, Emese and Peter Molnar. 2004. Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior and Development* 27: 54–63.

Nazzi, Thierry, Josiane Bertoncini, and Jacques Mehler. 1998. Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 24: 756–766.

Nip, Ignatius S.B., Jordan R. Green, and David B. Marx. 2009. Early speech motor development: Cognitive and linguistic considerations. *Journal of Communication Disorders* 42: 286–298.

Oller, D. Kimbrough. 2000. *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ostry, David J. and Kevin G. Munhall. 1985. Control of rate and duration of speech movements. *Journal of the Acoustical Society of America* 77: 640–648.

Papoušek, Mechthild and Hanuš Papoušek. 1981. Musical elements in the infant's vocalization: Their significance for communication, cognition, and creativity. In L.P. Lipsitt and C.K. Rovee-Collier (eds.), *Advances in Infancy Research*, vol. 1, 163–224. Norwood, NJ: Ablex.

Papoušek, Mechthild and Hanuš Papoušek. 1989. Forms and functions of vocal matching in interactions between mothers and their pre-canonical infants. *First Language* 9: 137–158.

Payne, Elinor, Brechtje Post, Luïsa Astruc, Pilar Prieto, and Maria del Mar Vanrell. 2012. Measuring child rhythm. *Language and Speech* 55: 203–229.

Pollock, Karen E., Diane M. Brammer, and Carlin F. Hageman. 1993. An acoustic analysis of young children's productions of word stress. *Journal of Phonetics* 21: 183–203.

Ramus, Franck, Marina Nespor, and Jacque Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73: 265–292.

Redford, Melissa A. 1999. An articulatory basis for the syllable. PhD dissertation, University of Texas.

Redford, Melissa A. 2013. A comparative analysis of pausing in child and adult storytelling. *Applied Psycholinguistics* 34: 569–589.

Redford, Melissa A. and Christina E. Gildersleeve-Neumann. 2007. The acquisition of two phonetic cues to word boundaries. *Journal of Child Language* 34: 815–843.

Riely, Rachel R. and Anne Smith. 2003. Speech movements do not scale by orofacial structure size. *Journal of Applied Physiology* 94: 2119–2126.

Robb, Michael P. and John H. Saxman. 1990. Syllable durations of pre-word and early word vocalizations. *Journal of Speech and Hearing Research* 33: 583–593.

Robb, Michael, Harvey Gilbert, Viki Reed, and Amanda Bisson. 2003. A preliminary study of speech rates in young Australian English-speaking children. *Contemporary Issues in Communication Science and Disorders* 30: 84–91.

Sabin, Edward J., Edward J. Clemmer, Daniel O'Connell, and Sabine Kowal. 1979. A pausological approach to speech development. In A. Siegman, and S. Feldstein (eds.), *Of Speech and Time: Temporal Speech Patterns in Interpersonal Contexts*, 35–55. Hillsdale, NJ: Lawrence Erlbaum Associates.

Saffran, Jenny R., Richard N. Aslin, and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274: 1926–1928.

Schmidt, Richard A. and Timothy D. Lee. 2005. *Motor Control and Learning: A Behavioral Emphasis*, 5th edn. Champaign, IL: Human Kinetics.

Schwartz, Richard G., Kakia Petinou, Lisa Goffman, Gerri Lazowski, and Christine Cartusciello. 1996. Young children's production of syllable stress: An acoustic analysis. *Journal of the Acoustical Society of America* 99: 3192–3200.

Sirsa, Hema and Melissa A. Redford. 2011. Towards understanding the protracted acquisition of English rhythm. In W.-S. Lee and E. Zee (eds.), *Proceedings from the 17th International Congress of Phonetic Sciences, (ICPhS-11)*, 1862–1865.

Smith, Bruce. 1978. Temporal aspects of English speech production: A developmental perspective. *Journal of Phonetics* 6: 37–67.

Smith, Bruce L. 1992. Relationships between duration and temporal variability in children's speech. *Journal of the Acoustical Society of America* 91: 2165–2174.

Smith, Bruce L., Michael D. Sugarman, and Steven H. Long. 1983. Experimental manipulation of speaking rate for studying temporal variability in children's speech. *Journal of the Acoustical Society of America* 74: 744–749.

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

Stark, Rachel E. 1980. Stages of speech development in the first year of life. In G. Yeni-Komshian, J. Kavanagh, and C. Ferguson (eds.), *Child Phonology*, vol. 1: *Production*, 73–90. New York: Academic Press.

Terken, Jacques and Dik Hermes. 2000. The perception of prosodic prominence. In G. Bruce and M. Horne (eds.), *Prosody, Theory and Experiment: Studies Presented to Gösta Bruce*, 89–127. New York: Springer.

Tingley, Beth M. and George D. Allen. 1975. Development of speech timing control in children. *Child Development* 46: 186–194.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Towse, John N., Graham J. Hitch, and Una Hutton. 1998. A reevaluation of working memory capacity in children. *Journal of Memory and Language* 39: 195–217.

Trevarthen, Colwyn. 1979. Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (ed.), *Before Speech: The Beginning of Interpersonal Communication*, 321–347. Cambridge: Cambridge University Press.

Trevarthen, Colwyn and Kenneth J. Aitken. 2001. Infant intersubjectivity: Research, theory, and clinical applications. *Journal of Child Psychology and Psychiatry* 42: 3–48.

Tsao, Ying-Chiao and Gary Weismer. 1997. Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language, and Hearing Research* 40: 858–866.

Tsao, Ying-Chiao, Gary Weismer, and Kamran Iqbal. 2006. Interspeaker variation in habitual speaking rate: Additional evidence. *Journal of Speech, Language, and Hearing Research* 49: 1156–1164.

Turk, Alice E. and James R. Sawusch. 1997. The domain of accentual lengthening in American English. *Journal of Phonetics* 25: 25–41.

Turk, Alice E. and Stefanie Shattuck-Hufnagel. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35: 445–472.

Turk, Alice E. and Laurence White. 1999. Structural influences on accentual lengthening in English. *Journal of Phonetics* 27: 171–206.

Vihman, Marilyn. 1996. *Phonological Development: The Origins of Language in the Child*. Oxford: Blackwell.

Vihman, Marilyn May, Marlys A. Macken, Ruth Miller, Hazel Simmons, and Jim Miller. 1985. From babbling to speech: A re-assessment of the continuity issue. *Language* 61(2): 397–445.

Walker, Jean F., Lisa M.D. Archibald, Sharon R. Cherniak, and Valerie G. Fish. 1992. Articulation rate in 3- and 5-year-old children. *Journal of Speech and Hearing Research* 35: 4–13.

Walsh, Bridget and Anne Smith. 2002. Articulatory movements in adolescents: Evidence for protracted development of speech motor control processes. *Journal of Speech, Language, and Hearing Research* 45: 1119–1133.