EFFECTS OF ACQUISITION RATE ON EMERGENT STRUCTURE IN PHONOLOGICAL DEVELOPMENT

Melissa A. Redford[1] and Risto Miikkulainen[2]

[1]*The University of Oregon* and [2]*The University of Texas at Austin*

ADRESSES:

Melissa Redford

The University of Oregon

Department of Linguistics

1290 University of Oregon

Eugene, OR 97403

redford@uoregon.edu

Risto Miikkulainen

The University of Texas

Department of Computer Sciences

1 University Station C0500

Austin, TX 78712

risto@cs.utexas.edu

EFFECTS OF ACQUISTION RATE ON EMERGENT STRUCTURE IN

PHONOLOGICAL DEVELOPMENT

Individual differences in child phonologies are often correlated with the rate at which language is being acquired. The correlation suggests some relationship between acquisition rate and language structure, but the nature of this relationship is not well understood. This paper presents a computational model, the Vocabulary Expansion Model (VEM), designed to explore how acquisition rate might interact with other constraints on phonological development to give rise to rate-dependent differences in the structure of words in a developing vocabulary. In VEM, words from a simulated adult target vocabulary are evaluated according to well-specified articulatory and perceptual costs, and selected into child vocabularies at different rates. Comparisons of the structure in the simulated child vocabularies show that words acquired early during vocabulary development have simpler phonological structure than words acquired later, and that slow word acquisition results in vocabularies with shorter words, simpler segments, greater segment to segment similarity, and simpler syllable structures than more rapid word acquisition. These results are qualitatively similar to the intra- and inter-individual differences observed in children with normal and delayed language, suggesting that at least some of the structural differences may emerge from the rate at which children acquire words.*

1. Introduction. Typically developing children acquire language at different rates, even though they all acquire similar grammars by age 5 or 6. The rate of language acquisition is measured initially in terms of vocabulary acquisition. The speed at which children acquire words is meaningful to parents and clinicians, since deviations from the average can indicate either language talent or language disorder (Stoel-Gammon & Dunn 1985; Bates, Dale, Thal 1995; Leonard 1998). Different rates of vocabulary acquisition are also correlated with differences in linguistic structure. For instance, slow vocabulary acquisition is correlated with smaller phoneme inventories and simpler syllable shapes than average (Stoel-Gammon 1989; Paul & Jennings 1992; Paul 1993; Thal, Oroz, McCaw 1995; Rescorla & Ratner 1996; Mirak & Rescorla 1998; Pharr, Ratner, Rescorla 2000; Edwards, Fox, Rogers 2002; Storkel 2004). The relationship between fast vocabulary acquisition and phonological structure is only just beginning to be studied, but it seems likely that rapid word learners may have larger phoneme inventories and more complex syllable shapes than average (e.g. Smith, McGregor, Demille 2006).

The relationship between acquisition rate and linguistic complexity is intuitively satisfying—slow learners have difficulty grasping complex structure and fast learners are more able than average to do so because the slow are 'slow' and the fast are 'quick'. However, this intuition conflicts with the data. Many late-talking children with simpler phonologies have normal motor, perceptual, and cognitive skills (Paul & Jennings 1992; Mirak & Rescorla 1998; Leonard 1998). It is thus not even clear why children acquire words at different rates. Some have argued that a delay reflects difficulty with the arbitrary mapping of sound to meaning (Storkel 2004), poor working memory (Edward & Lahey 1998), or even a lack of motivation (Paul & Shiffer 1991), but none of these factors explain why phoneme inventories and syllable structure are involved. So why

might the rate of vocabulary acquisition correlate with phonological complexity? One possibility is that rate itself affects the complexity of emerging structure. Rate could interact with the various selection criteria a child uses to acquire words such that different rates of word selection would be characterized by the stronger influence of one set of criteria over another. The resulting lexicons would be differently filtered and so characterized by different phonological patterning. The current paper presents a computational model of phonological development that was built to test this possibility.

Specifically, the model formalizes the idea that a source-filter mechanism[1] shapes phonological development. Children acquire words from the most frequent and semantically transparent subset of words in the adult vocabulary (source), and the source items are selected according to the phonetic principles of articulatory ease, perceptual salience, and perceptual distinctiveness (filter). Phonological structure emerges from the word shapes that are thus selected and represented. The effect of word acquisition rate on emergent phonological structure is tested in the model by varying the rate of word acquisition across different simulations, and by evaluating the resulting vocabularies along paradigmatic and syntagmatic dimensions; namely, in terms of sound and syllable types as well as in terms of how these were combined in different word shapes. Although faster rates might be expected to favor salient types and distinctive combinations over articulatorily simple types and recurrent combinations, the results indicate that the phonetic filter is simply more effective during slow acquisition than during fast acquisition. Since a more effective filter also yields smaller phoneme and syllable inventories with simpler structures than a less effective one, the results are consistent with the observed language data on how rate and complexity correlate. The model therefore provides a novel explanation of the empirical data that; if confirmed with natural data, allows us to understand this aspect of language development in precise, computational terms.

2. <u>A source-filter model of phonological development</u>. The source-filter model of phonological development describes the phonetic influences on developing phonological knowledge. Specifically, the model assumes that phonetic pressures on the input define the mental soundscape from which phonological knowledge is abstracted. This idea is taken from the literature on child language acquisition, as shown below.

2.1. <u>Vocabulary acquisition and phonological development</u>. The consensus in the field of child language is that phonological development and vocabulary acquisition are linked (Stoel-Gammon & Dunn 1985; Smith & Locke 1988; Shore 1995; Beckman & Edwards 2000a; McCune & Vihman 2001; Storkel 2004). The link can be immediately and intuitively grasped by considering the phonological simplicity of children's very first words (e.g. <u>mama</u> / <u>dada</u>) compared with the more complex forms they produce as vocabulary size increases (e.g. <u>bottle</u> / <u>diaper</u>, then <u>snail</u> / <u>spider</u>). There is less agreement, however, on the mechanism that accounts for the link. At least three possibilities exist: a developing phonology drives vocabulary acquisition (e.g. Locke 1983; Schwartz 1988; Vihman 1993; MacNeilage, Davis, Matyear 1997); vocabulary acquisition drives phonological development (e.g. Ferguson & Farwell 1975; Lindblom 1992; Beckman & Edwards 2000a); vocabulary acquisition and phonological development interact so that increases in one drive increasing complexity in the other and vice versa (e.g. McCune & Vihman 2001; Storkel & Moriesette 2002). Most researchers would probably agree with some version of the third alternative, since it integrates the major insights of the other two hypotheses. In order to understand these insights and the hypothesized interaction between vocabulary acquisition and phonological development, we need to understand the arguments for how each might influence the other.

To some extent, the emphasis on a particular direction of influence—from phonology to vocabulary or from vocabulary to phonology—depends on how phonology is defined. Most researchers in child language define phonology operationally in terms of

the word shapes that the child produces at any given point in time.[2] Phonological development is therefore characterized as changes in the linguistic output capabilities of the child. If one were to take the strong stance that phonology is perfectly characterized by its operational definition, then one would elevate the importance of output capabilities in phonological development. Since infants also demonstrate output capabilities prior to word acquisition (i.e. babbling), one might expect that these prior abilities help to set the pace and define the character of phonological development.

There is substantial evidence to support such a view. For instance, MacNeilage, Davis, and Matyear (1997) found basic similarities in the structure of babbling and early words that they attributed to a strong motor constraint in child speech; namely, a near-complete dependence on the jaw cycle to organize serial action in speech. Further, MacNeilage and Davis (2000) argued that this constraint has a deep evolutionary history and that its tell-tale segment co-occurrence patterns are evident in the core vocabulary items of adult languages today. From the point of view of phonological development, the constrained sound patterns of adult core vocabularies mean that in many instances children are acquiring words that match their preferred mode of production (MacNeilage, Davis, Kinney, Matyear 2000).

Many others would interpret the match between acquired words and a preferred motor pattern as an instance of phonological selectivity—a widely accepted idea that 'children choose to produce certain words, but not others, specifically because of the phonological characteristics of words relative to their existing phonologies' (Schwartz 1988:187). Although many researchers seem to agree that phonological selectivity is grounded in motor constraints, Schwartz cautions that universal articulatory considerations cannot be solely responsible for selectivity since large individual differences are observed. For example, the developing vocabulary of one child may reflect that child's strong preference for words with bilabials, while another's may reflect a preference for words with alveolars (Vihman, Ferguson, Elbert 1986).

Vihman (1993) argued, however, that even individual differences suggest the influence of articulatory constraints on selectivity. She proposed that phonological development proceeds in large part from vocal practice, and expounded a model in which 'an "articulatory filter" acts to "screen in" … words which resemble the child's existing motor production patterns; only those words which roughly match the child's pre-existing production patterns would be represented in sufficient detail to allow early identifiable word production' (Vihman 1996:126). In addition to explaining phonological selectivity, Vihman has argued that production factors account for other phenomena of early phonological development. For instance, she suggested that homonymy is relatively frequent in early child language because young children try to maximize vocabulary items while minimizing production patterns (Vihman 1981).

At some point, however, the articulatory pressures on a developing system must be counterbalanced by requirements for perceptual salience and distinctiveness, just as they are in adult languages (Martinet 1965; Lindblom, MacNeilage, Studdert-Kennedy 1984; Ohala 1993; Demolin & Soquet 2001). If children acquire a vocabulary in order to communicate their wishes and desires to others, then they need to code different concepts with different sound patterns. Put another way, absolute homonymy may be an optimal strategy for minimizing production patterns, but it is not a viable strategy for a child motivated to communicate. The listener prefers that the child produce salient and distinct sound shapes to code distinct concepts, a preference that historical linguists have assumed to apply in adult language as well (e.g. Vennemann 1978; Anttila 1989). This emphasis on the communicative function of language brings us to another definition of phonology, one that supports the view that vocabulary acquisition drives phonological development.

For most linguists, phonology is characterized not by vocal output, but by those structural properties that are used to code and distinguish meanings. According to this view, phonological development might be properly defined as the development of an

increasingly rich, abstract knowledge of the correspondences between sounds and sound structures in a particular language. Since the fundamental emphasis is on the coding and distinguishing of meaning, this view would naturally predict that phonological development follows from vocabulary acquisition (i.e. the acquisition of sound patterns that are linked to specific meanings). For example, the sound categories /p/ and /b/ might emerge from the understanding that the sound shapes pig and big denote separate concepts. On this view, phonological knowledge is abstracted from the language that the child has acquired, namely, the vocabulary (Ferguson & Farwell 1975; Lindblom 1992; Beckman & Edwards 2000a). Clinical evidence supports this idea. For example, Girolametto, Pearce, and Weitzman (1997) showed that phonological complexity increases with lexical training in children who have expressive language delay.[3]

If one adopts the view that phonological knowledge is abstract and emerges from the lexicon, as we do here, one is left with a problem of explanation. In contrast to the hypothesis that vocabulary acquisition is driven by phonological development (i.e. output capabilities), the hypothesis that phonology emerges from the vocabulary does not explain why some word shapes are selected before others. Although the emergentists do not address this question directly, the suggestion seems to be that social, cognitive, and perceptual-motor constraints still define vocabulary acquisition. The problem then becomes how to acknowledge the influence of all of these factors when proposing a coherent and explicit mechanism of vocabulary acquisition to account for phonological development. Our solution is to characterize the process in terms of a source-filter mechanism.

2.2. The source, the filter, and representation. As noted in the preceding section, children appear to select words from the target language according to what they can produce (i.e. phonological selectivity). If phonology is defined as abstract knowledge emerging from the lexicon, then phonological selectivity suggests that young children

unconsciously shape the structure of their developing linguistic knowledge. In this section, we elaborate on this suggestion and address the questions of how word selection occurs and what is represented.

Source. Vocabulary acquisition is influenced by factors other than the phonological structure of target words. Two major additional factors include word frequency and meaning (e.g. Getner 1982; Markman 1987; Hart 1991; Beckman & Edwards 2000b). These factors can be defined as characteristics of the source in the following way.

The source for word selection and so for phonological development is a subset of the words in the target vocabulary. These words will be available for selection from language use with and around the child. High-frequency source words will be more available to a child than low-frequency words. A number of studies support this observation (for a review see Rescorla, Alley, Christine 2001); early vocabularies typically include words that children have heard thousands of times (e.g. mama, diaper, bottle). Nonetheless, some of the most frequent words of all—function words, for example—enter a child's vocabulary remarkably late (Bates et al. 1995). Put another way, early vocabularies only include certain high-frequency words, such as concrete nouns and common verbs. This strong preference for a subset of content words might be explained as a selection effect, but it is better understood as a source characteristic. Only words with transparent meanings will be available for the child to select and include in her vocabulary (see e.g. Getner 1982; Markman 1987; Choi & Gopnik 1995). Thus, the source is defined as the subset of those words in the adult language that are available to a child; namely, the ones that are frequently heard and are (more) transparent in meaning.

Filter. As noted above, the child language data suggest that acquisition is highly constrained by immature motor processes (Menn 1983; Locke 1983), which has led researchers to propose that word selection is guided by articulatory constraints (Schwartz 1988; Vihman 1996). However, articulatory pressures must be counterbalanced

somewhat by perceptual pressures. The child is a listener as well as a speaker, and so the need for perceptual saliency and contrast will also influence the type of words selected for inclusion in a growing vocabulary. For instance, Ferguson and Farwell (1975) suggested that children favor new words with distinctive sound patterns that are readily distinguishable from those already present in their inventories (see also Storkel 2004). Other work indicates that young children may have fewer similar sounding words in their vocabularies than older children or adults (Charles-Luce & Luce 1990; 1995).

In light of the evidence, we propose that children select words from the target vocabulary that maximize articulatory ease, perceptual salience and distinctiveness. This proposal thus extends to phonological development the idea from historical linguistics that phonetic pressures influence sound change.[4] Whereas the phonetic pressures are assumed to be balanced in adult language (e.g. Lindblom 1990), in acquisition the pressure for ease is initially stronger than the perceptual pressures. The greater importance of articulatory over perceptual pressures in early acquisition explains why language production lags behind language perception, and is also consistent with the fact that earlier vocabularies are composed of words with simpler sounds and syllable structures than later vocabularies (Locke 1983; Vihman 1996). The pressure for distinctiveness becomes more important as vocabulary size increases and the developing sound space becomes more crowded. As the pressure for distinctiveness increases, sound structures diversify.

To summarize the ideas developed so far, vocabulary acquisition can be understood in terms of a source-filter mechanism, namely, as the phonetically-motivated selection of the most frequent and conceptually accessible adult words that are used with and around children. Insofar as phonology emerges from the lexicon, the source-filter conception of word acquisition also characterizes developing phonological knowledge. This knowledge will be dependent on the representation of stored vocabulary items.

Representation. Currently, there is no sure way to know how a child represents the sound shape of the words that she acquires. On the one hand, the child understands the adult target word and so must register an acoustic pattern associated with its meaning. On the other hand, she systematically produces a simpler version of the target. One might argue that when these two observations are taken together they more strongly suggest an adult-like, acoustic-perceptual representation of the word rather than a child-like articulatory representation. A faithful acoustic-perceptual representation of adult speech would allow the child to comprehend adult speech and would also provide her with an imagined target for production. The fact that she does not achieve this target might be attributed to an immature motor system, and need not entail a degraded acoustic-perceptual or child-like articulatory representation.[5]

Positing an adult-like acoustic-perceptual representation is, of course, the simplest assumption one can make about lexical representation in the child language learner: we do not need to imagine how the signal is transformed by the child's perceptual system or how it may be mediated by the child's motor abilities. In spite of its simplicity, the idea is consistent with the common observation that perception precedes production as well as with the idea of acoustic targets in speech (Kingston & Diehl 1994; Lindblom 1996; Coleman 1998). Further, the idea is supported by evidence from infant speech perception studies, which suggests that infants encode speech stimuli in fine acoustic detail (Werker & Tees 1999; Jusczyk & Luce 2002; Swingley & Aslin 2002). If anything, these latter studies suggest that infants are more sensitive to acoustic detail in the adult speech waveform than older infants or adults are. This assertion is best exemplified by the well-supported finding that infants initially discriminate between speech sound categories that vary along any of the dimensions exploited by languages. Near the onset of first word production, infants reportedly lose this ability and start behaving like adults; that is, they only discriminate between sound categories that encode language-specific contrasts (Werker & Tees 1984; Werker & Lalonde 1988; Best 1993; Polka & Werker 1994).[6]

The abstract knowledge about the sound patterns of a language that infants begin to demonstrate just prior to first word production may help to focus production effort on the most salient and linguistically significant aspects of a word. However, we propose that this knowledge will not directly drive further word selection. Rather, word selection will remain a phonetic filtering process; a process that is sensitive to the articulatory costs and the inherent perceptual saliency of words as well as to the number and shapes of previously represented words. The give and take between articulatory ease and perceptual salience and distinctiveness will ensure that the richness of stored sound patterns increases systematically through time, which will ensure that the developing phonology does as well. Next, we consider whether the rate at which word selection and storage occurs could affect the structure of a developing system.

3. <u>Acquisition rate and emergent structure</u>. Typically developing children usually acquire (produce) their first words around 12 months of age.[7] By 18 months, they have usually acquired 50 words. After this point, words are added more quickly so that by the time they start combining words into phrases at around 24 months they will have, on average, around 300 words (Fenson et al. 1994). The nonlinear vocabulary growth that results from the acquisition of ever increasing numbers of words per month has been called the <u>vocabulary spurt</u> or the <u>vocabulary burst</u> (e.g. Dromi 1987; Goldfield & Reznick 1990). Explanations for the spurt are varied, but usually invoke cognitive factors. For instance, Baldwin and Markman (1989) suggested that the vocabulary spurt is due to a 'naming insight,' that is, the sudden realization that everything can be labeled. Other explanations assume gradual cognitive development, such as slowly improving categorization abilities (Gopnik & Meltzoff 1987). The gradual explanations may be more in line with the cumulative data, which suggest a curvilinear increase in vocabulary rather than a true burst (Ganger & Brent 2004). Importantly, though, the explanations for

rate of vocabulary acquisition typically focus on cognitive development rather than on the development of motor and perceptual capabilities (but see Plunkett 1993).

If cognitive factors account for word acquisition rate, then the rate of selection is explained independently from the what of selection. The what of selection are the word shapes selected into a developing vocabulary, and these can be adequately characterized by the input and by the pressures for articulatory ease, perceptual salience and distinctiveness (i.e. motoric and perceptual factors). The rate of selection may nevertheless interact with the what of selection. A cognitive factor that increases the number of words selected into a vocabulary at any given point in time may effectively reweight the perceptual-motor factors that define which word shapes are selected at that point in time. In particular, the pressure for distinctiveness might be more important when many words are being added to a vocabulary over a short period of time than when these words are added slowly. The idea is that when words are added quickly they form an assembly, which is more likely to highlight system-level attributes like distinctiveness over word-level attributes like ease and salience.

3.1. Rate and structure in natural language acquisition. Research on the relationship between vocabulary acquisition and phonological structure has focused primarily on correlations between size and structure, rather than on correlations between rate and structure. Nonetheless, there is some indication that the major increase in word acquisition rate occurring at about 2 ½ years of age precipitates a refinement in output capabilities (Dromi 1987; Roberts 1998). The refinement of production might be attributable to increasing phonological knowledge, which would guide the child to refine production of the linguistically significant. The knowledge, in turn, would have been abstracted from the sound patterns represented in the lexicon. These sound patterns would have diversified during the spurt, since the motivation to acquire new words

quickly would have increased the importance of the perceptual pressures that focus on contrasts between words.

The relationship between rate and structure is even more evident in the clinical literature. In particular, it is well-established that children who acquire words slowly have simpler phonologies than children who acquire words at the typical rate. Children who have not achieved 50 words by 24 months of age have smaller phonemic repertoires, simpler syllable structures, and other hallmarks of a simpler phonology than their age-matched peers (Paul & Jennings 1992; Thal et al. 1995; Rescorla & Ratner 1996; Pharr et al. 2000). These children are characterized as late talkers because the slow rate at which they acquire words cannot be explained in terms of motoric, perceptual, or cognitive deficits, and because these children often catch up to their typically developing peers later on (Leonard 1998). The otherwise unremarkable profile of late talkers has led some to suggest that phonological development in these children is normal, just slow (Paul & Jennings 1992; Mirak & Rescorla 1998; Leonard 1998).[8] However, there are a few studies that compare the early phonologies (i.e. output capabilities) of late talkers with those of younger, typically developing children with similar vocabulary sizes. The studies that have made these comparisons show that late talkers use simpler phonological structures more often than the younger typically developing children (Schwartz, Leonard, Folger, Wilcox 1980; Aguilar-Mediavilla, Sanz-Torrent, Serra-Raventos 2002). These data are intriguing because they suggest that phonological development in late talkers is different, even if it is not deviant (see also Leonard 1998:31-36). Since there is no perceptual-motoric or cognitive explanation for the phonological difference, perhaps it is simply attributable to the slow rate at which these children select and represent words. In particular, it could be that the phonetic pressures that filter word selection are not reweighted to favor distinctiveness over ease and salience. Slow word acquisition could mean that each word is evaluated independently of other words, which would favor the articulatory and perceptual merits of a word over its contrastiveness in the system. This

type of evaluation could result in simpler phonological patterns, assuming that diversity and complexity of structure emerge from pressures for distinctiveness.

3.2. <u>Testing the effects of acquisition rate on structure</u>. How could one test the hypothesis that the rate of word selection affects emergent phonological structure by changing the shape of a phonetic filter? Such a hypothesis is difficult to test directly in child language studies since neither the phonetic filter nor the represented sound shapes are accessible to the researcher. Also, linguistic abilities are highly complex and the factors that are hypothesized to interact with one another to influence a particular ability are difficult to isolate with sufficient precision to test their individual or combined effects. An artificial system would provide a better environment for testing the hypothesis. In such a system, the various factors influencing acquisition could be evaluated directly because they would be well specified and isolated from others. Simulations could confirm or disconfirm whether a hypothesized interaction between factors is possible in principle. The results would then suggest avenues for empirical research that would provide the necessary natural language evidence for or against the hypothesis.

Artificial systems have often been employed in linguistic and language acquisition research to rigorously specify theory driven hypotheses and test their in-principle validity. An early example of this is Liljencrants and Lindblom's (1972) model exploring the extent to which vowel inventories in languages are explained in terms of a simple metric of perceptual distance. Subsequent modeling studies have refined the phonetic pressures explaining the structure of vowel inventories (Schwartz, Boë, Vallée, Abry 1997; de Boer 2000) and explored how similar pressures might define the emergence of syllable structure (Lindblom et al. 1984; Joanisse 2000; Redford, Chen, Miikkulainen 2001). Modeling efforts in language acquisition have been even more intense, and have often been used to engage in strong theoretical argument (e.g. Marcus,

Pinker, Ullman 1992; Elman, Bates, Johnson, Karmiloff-Smith, Parisi, Plunkett 1996). Other systems have been built to explore the development of abilities underlying language acquisition. For instance, Guenther's DIVA model (1995; Guenther, Hampson, Johnson 1998) examined the development of speech motor control processes underlying speech sound acquisition.

All of these models are abstract, yet able to provide very specific explanations for complex phenomena. In so doing, they clarify our understanding of the phenomena and identify particular problem areas for further research. Given these virtues, a computational model provides the best method for exploring the difficult question of how the rate of word selection might affect developing phonological structure. The model developed here for this purpose is called the Vocabulary Expansion Model (VEM), and it is described next.

4. The Vocabulary Expansion Model (VEM). VEM models vocabulary acquisition by first developing a source vocabulary that represents those adult words that are available to the child. Candidate words are created by concatenating existing syllables. Articulatory ease, perceptual salience and distinctiveness are then measured. Words that minimize ease and maximize salience and distinctiveness are chosen for inclusion in the source vocabulary. In this way, the source vocabulary is optimized according to the phonetic pressures, ensuring that sound structure emerges.

Once the source vocabulary is complete, child vocabularies are developed. Subsets of words are selected from the source, according to the same phonetic pressures, until the vocabulary is roughly equal in size to children's vocabularies at the end of the first word stage. Different child vocabularies are acquired at different rates.

The overall structure of the model is outlined in Figure 1, and described in more detail in the following subsections.

Insert Figure 1 about here.

4.1. <u>Words</u>. In VEM, words are word shapes (i.e. phonological forms) without associated meaning. Each individual word shape nonetheless represents a unique sound-meaning correspondence. So, multiple occurrences of a single word shape represent homonymy in VEM.

Word shapes are generated for the source vocabulary by randomly selecting and concatenating syllables (a maximum of four per word in the present simulations) from a fixed syllable repertoire. The syllable repertoire represents a space of possible sound patterns in the artificial language. To ensure that a sufficiently large sound space is modeled, significant variability is introduced via a detailed production algorithm that generates syllables. The algorithm also provides the basis for rigorously defining articulatory ease as well as the low-level information necessary for evaluating salience and distinctiveness.

The syllable production algorithm generates articulatory sequences by temporally overlapping the movements of several different speech articulators: the jaw, the upper and lower lips, the tongue tip, the tongue body, and the glottis. With the exception of the jaw, the movements of all the articulators are represented categorically in a neutral (= rest) or end state (= target) position. The opening and closing movements of the jaw are discretized into a series of time steps. Although there are some co-occurrence constraints on movement (e.g. the lower-lip cannot be used to form a labiodental fricative while the upper-lip is used for rounding), variability in movement/sound production is introduced by randomly selecting end states for the different articulators and randomly selecting the amplitude of the jaw cycle.[9, 10] The resulting well-defined movement sequences are in many ways analogous to the time varying representations of vocal tract variables in

Articulatory Phonology (i.e. gestural scores; Browman & Goldstein, 1986; 1992). A relevant conceptual difference between the models is that, in VEM, the movements of the lips, tongue, and glottis are nested within the cyclic movements of the jaw, which is conceived of as the articulatory manifestation of the rhythm pattern in speech.

After movement sequences are generated they are segmented to provide information that is used to define different syllable shapes in VEM. Segmentation is accomplished by sliding a temporal window across a sequence. Transitions between the different postures of the lips, tongue, and glottis within the movement sequence are used to indicate segment boundaries and assign segment type. Thus, phones emerge in VEM as they do in Articulatory Phonology: from stable articulatory configurations.

The segment types that are recognized in VEM are determined largely by jaw height (representing degree of constriction). These are obstruent consonant (C), sonorant consonant (S), and vowel (V). The unique combinations of articulatory postures associated with C, S, or V represent unique phones (e.g. a voiced labial obstruent versus a voiced alveolar obstruent).[11] Syllable boundaries are defined by the boundaries of a single open-close jaw cycle (Guenther 1995; Guenther et al. 1998; Redford et al. 2001), and syllable shape refers to the number and types of segments that occur within a single cycle (e.g. CV and CSV represent two different syllable shapes). Once segmented, a syllable shape label is associated with the movement sequence to allow for future operations that refer either to syllable structure or to subsyllabic structure or to both.

VEM uses the mechanism of variation and selection to generate an initial syllable repertoire that carves out a portion of the possible sound space for the source vocabulary. Many populations of syllables are generated on each run of the model (1500 populations of 300 syllables each in the present simulations). These populations are then evaluated for overall articulatory ease, which is calculated on the basis of the articulatory costs for each syllable in the population.

The articulatory cost $C_{A,xs}$ for an individual syllable $\underline{s}$ appearing in word $\underline{x}$ is the sum of the number of movements made by the articulators during a single jaw cycle that produces $\underline{s}$:

(1) $C_{A,xs} = \sum_{a}\sum_{t} M_{at},$

where $\underline{a}$ is an articulator, and $\underline{t}$ is a time step, and

(2) $M_{at} = \begin{cases} 1 \text{ if articulator } \underline{a} \text{ moved over time } \underline{t}, \\ 0 \text{ otherwise.} \end{cases}$

This way, more complex segmental articulations and syllable structures are more costly than simpler articulations and syllable structures, which leads, for instance, to a preference for singly articulated obstruents over doubly articulated obstruents and for CV syllables over CSV or CVC syllables.

The syllable population with the lowest cost is selected as the repertoire for word generation. The process of evaluation and selection thus restricts the possible sound space from which words are developed, which is important since languages may also be assumed to develop under such restrictions. Nevertheless, substantial diversity remains in VEM's initial syllable repertoire.[12]

4.2. Vocabularies. Once a syllable population is selected, VEM develops a source vocabulary and several child vocabularies (30 in the present simulations). The source vocabulary is developed by first establishing a small core vocabulary and then by systematically expanding it into the full source. One thousand 5-word vocabularies are generated randomly, and the one with the lowest articulatory and perceptual costs (described in detail in section 4.3) is selected as the core. This core is then expanded by adding selected words to it. At first, one word at a time is added; later on, several words

are added at once, as determined by a growth rate parameter (described in section 4.4). At an expansion step where n-words are to be added, a new population of 1000 sets of n-words is generated randomly, and the set that results in the lowest overall cost is added to the vocabulary. In this way, the source vocabulary expands nonlinearly over time, modeling the type of growth that occurs during language acquisition. Vocabulary expansion ends when a final vocabulary size of 1500 words is achieved. This size is not equal to an adult vocabulary size, but is large enough to model a vocabulary that provides between 75% and 80% coverage of written texts (Francis & Kucera 1982) and so is presumably large enough to model the subset of the vocabulary that is available to children for selection (see section 2.2).

VEM develops child vocabularies in the same way that it develops the adult source vocabulary, except that words are selected from the source rather than from randomly generated populations of new words at each time step. A small core vocabulary is first selected from a population of 200 small vocabularies randomly drawn from the source. The core is then systematically expanded by adding selected words to it (these are selected according to the phonetic costs from among 200 randomly drawn n-word sets at each time step). Again, the number of words selected for inclusion in the vocabulary increases with the size of the vocabulary, thus modeling nonlinear growth during language acquisition. Expansion ends when the vocabulary reaches 300 words, which is roughly the number of words a child has at the end of the first word stage (Fenson et al. 1994). This stage is an appropriate stopping point because word selection is likely filtered by additional pragmatic and grammatical factors once children begin to combine words into phrases.

Although many details of vocabulary growth are abstracted in the model,[13] the algorithm captures two important assumptions about the structure of existing and developing phonological systems: structure emerges (1) over time and (2) in response to

selection pressures. In VEM, phonological structure emerges from phonetic factors that have long been suggested to influence such structure (e.g. Martinet 1965; Liljencrants & Lindblom 1972; Lindblom et al. 1984; Ohala 1993; Demolin & Soquet 2001). These pressures are defined next.

4.3. Selection pressures. The articulatory and perceptual pressures are implemented as specific evaluation functions in VEM. These functions evaluate the cost of selecting a particular word or word set into the vocabulary. One articulatory pressure and three specific perceptual pressures are included in VEM. The articulatory pressure and one of the perceptual pressures affect word-level structure only, that is, the cost of selecting a word is considered only on the basis of the structure of that word. The other two perceptual pressures affect system-level structure, that is, the cost of selecting a word is considered based on the word's relationship to the sound structure of other words in the growing vocabulary.

Word-level costs. The articulatory cost $C_{A,x}$ evaluates the ease with which word x is produced. It is simply the sum of the number of movements needed to produce that word, which is calculated by summing the articulatory costs of the word's component syllables:

$$(3) \qquad C_{A,x} = \sum_s C_{A,xs} \,,$$

where s is a syllable in word x.

Similarly, the word-level perceptual cost $C_{I,x}$ evaluates the perceptual saliency of word x based on the structure of its component syllables.[14] It is calculated by summing the costs of each component syllable s and dividing by the number of syllables N in a

word. Whereas articulatory costs increase with word length, perceptual costs should not. It is for this reason that the summed cost of the syllables in $\underline{x}$ is divided by $\underline{N}$:

$$(4) \qquad C_{I,x} = \frac{\sum_s C_{I,\,xs}}{N}\;.$$

The inherent perceptual cost of a syllable $\underline{s}$ equals the sum of the costs of its onset, offset, and nucleus:

$$(5) \qquad C_{I,xs} = \sum (n_s + f_s + v_s),$$

where $\underline{n}$ is a syllabic onset cost, $\underline{f}$ is a syllabic offset cost, and $\underline{v}$ is a syllabic nucleus cost.

Different syllable structures have different inherent perceptual costs due to a variety of factors. The cost $C_{I,xs}$ reflects some known and some assumed costs associated with particular structures. For example, pre-vocalic consonants are more perceptible than post-vocalic consonants (Redford & Diehl 1999; Benkí 2003), single consonants are more perceptible than consonant clusters as evinced by the perceptual assimilation of one consonant to another in a sequence (Hura, Lindblom, Diehl 1992; Ohala 1990), and obstruent-sonorant combinations (CS or SC) are likely to be more perceptible than obstruent-obstruent combinations (CC) because sonorant transitions are more likely than most obstruent transitions to preserve place information of the preceding or following obstruent. The specific values for onsets and offsets in the present simulations were: C/S onset = 10, C/S offset = 20, CS onset = 30, SC offset and CC onset = 40, CC offset = 50, CCS onset = 60, SCC offset and CCC onset = 70, and CCC offset = 80. The value for syllable nuclei equaled the number of vowels in the nucleus (i.e. a single V nucleus = 1, a diphthong = 2, a triphthong = 3). If the nucleus was devoiced, the cost was 50 times its value.[15]

In summary, the word-level costs that operationalize the concepts of articulatory ease and perceptual salience favor word shapes with simple sounds and syllable structure. This means that words like <u>mama</u> and <u>papa</u>, with one constriction gesture and reduplicated sounds as well as CV structure, will be selected over words like <u>sleepy</u> and <u>cracker</u>, that include two or more constriction gestures (e.g. the [l] in <u>sleepy</u> and [ɹ] in <u>cracker</u>), more differentiated movement across the word, and more complex syllable shapes (e.g. the CSV syllable in <u>sleepy</u> and <u>cracker</u>).

<u>System-level costs</u>: The first of the two system-level perceptual costs, $C_{P,x}$, evaluates how distinctive the syllables of a word are with respect to the other syllables in the system. $C_{P,x}$ is calculated as the sum of the cost for each of its syllables <u>s</u> divided by the number of syllables <u>N</u> in the word:

$$(6) \qquad C_{P,x} = \frac{\sum\limits_{s} C_{P,xs}}{N} \ .$$

To calculate $C_{P,xs}$, the sounds and structure of syllable <u>s</u> in word <u>x</u> are compared with the sounds and structure of each syllable <u>r</u> in every other word <u>y</u> in the vocabulary. Identical structures with identical sounds are penalized, thus estimating the extent to which syllable <u>s</u> adds to the diversity/homogeneity of paradigmatic structure:

$$(7) \qquad C_{P,xs} = \sum_{x \neq y} \sum (J_{xsyr} + K_{xsyr} + L_{xsyr})$$

where

(8) $\qquad J_{xsyr} = \begin{cases} 1 \text{ if the onset of syllable } \underline{s} \text{ in word } \underline{x} \text{ has the same structure and} \\ \quad \text{sounds as the onset of syllable } \underline{r} \text{ in word } \underline{y}, \\ 0 \text{ otherwise,} \end{cases}$

and K_{xsyr} is calculated similarly for offsets, L_{xsyr} similarly for nuclei.

The second system-level perceptual cost $C_{S,x}$ evaluates how distinctive the syntagmatic structure of word $\underline{x}$ is compared to other words $\underline{y}$ in the vocabulary:

(9) $C_{S,x} = H_x \sum\limits_{y} C_{S,xy}$

where H_x is the number of times word $\underline{x}$ occurs in the vocabulary. The similarity $C_{S,xy}$ is calculated by comparing the sounds of $\underline{x}$ with the sounds of $\underline{y}$ from left to right (i.e. from the onset to the end of the shortest of the two words participating in the comparison) and counting how often they match divided by the number of points of comparison $\underline{R}$:

(10) $C_{S\,xs} = \dfrac{\sum\limits_{a} \sum\limits_{t} M_{ywa}}{R}$

where $\underline{a}$ is an articulator at time step $\underline{t}$, and

(11) $M_{xyat} = \begin{cases} 1 \text{ if articulator } \underline{a} \text{ made the same movement over time } \underline{t} \text{ in words} \\ \quad \underline{x} \text{ and } \underline{y}, \\ 0 \text{ otherwise.} \end{cases}$

In other words, $C_{S,x}$ estimates how similar a word is to other words in the vocabulary and provides a multiplicative penalty for homophones (i.e. identical word shapes). Such a penalty is consistent with the idea that optimal vocabularies have one word shape per meaning (Venneman 1978; Anttila 1989).

In summary, the system level costs that operationalize the concept of perceptual distinctiveness favor word shapes that are maximally dissimilar from one another. This in turn creates a selection pressure for sound pattern and syllable structure diversity in the expanding vocabulary.

Overall costs: The overall cost C_x for a word participating in a source or child vocabulary is a weighted sum of the individual costs:

(12) $C_x = W_1 C_{A,x} + W_2 C_{I,x} + W_3 C_{P,x} + W_4 C_{S,x}$

The weights are adjusted so that each of the costs contribute an appropriate amount to the overall cost of the word over the course of vocabulary expansion. The appropriate values were found experimentally to be $W_1 = 1$, $W_2 = 5$, $W_3 = 3$, $W_4 = 1$. Importantly, the different costs do not contribute equally in determining selection at different stages of expansion. In particular, the system-level costs increase significantly with vocabulary size. Accordingly, articulatory costs are more important in determining selection during the initial stages of expansion (i.e. in smaller vocabularies) than during later stages of expansion (i.e. in larger vocabularies). For this reason, VEM is able to model the stronger articulatory constraints on child vocabularies than on adult vocabularies without reweighting the individual costs for acquisition.

4.4. Expansion rate. All vocabularies develop over time in VEM, but the rate at which they do so is controlled by a growth rate parameter. At every developmental time step t, a continuous vocabulary size variable V(t) is multiplied by the growth rate parameter g:

(13) $V(t+1) = gV(t)$

The product is rounded to its nearest integer value, which then determines the number of words added to the vocabulary at time step t+1. In this way, the growth rate parameter ensures that the number of new words added to a vocabulary is proportional to the size of the vocabulary, that is, the vocabulary grows nonlinearly. Nonlinear growth is more

realistic than linear growth in natural systems, and is a known characteristic of vocabulary development (Fenson et al. 1994).

5. Simulations. To test whether or not the rate of word acquisition affects emerging phonological structure, child vocabulary acquisition was simulated at three different rates as shown in Figure 2.

Insert Figure 2 about here.

The growth rate parameter was set to model slow, medium, and fast acquisition with the experimentally-derived values of 1.02, 1.026, and 1.035. During a single run of the model, a single source vocabulary gave rise to 30 child vocabularies—10 vocabularies at each rate of word acquisition. The simulations were repeated four times, so there were a total of 120 child vocabularies (40 slow, 40 medium, and 40 fast growth vocabularies) developed from four different source vocabularies. The different source vocabularies provided different initial conditions (e.g. ambient language conditions) for vocabulary acquisition.

The phonological structure of the acquired vocabularies was analyzed as a function of growth rate (Rate), vocabulary size (Size), and source vocabulary (Source). Rate and Size were treated as within variables (i.e. drawn from a single source vocabulary) and Source was treated as a between variable. Since phonological structure emerges from the lexicon (section 2.2), the sound patterns of the vocabulary will determine the complexity of emergent phonological structure. The sound patterns of interest for evaluating complexity were the word-level costs of word length, articulatory difficulty, and syllable structure complexity, and the system-level costs of diversity of syllable onsets, offsets, and nuclei, and the distinctiveness of sound sequences. Vocabularies with longer words, higher word-level costs, and lower system-level costs

indicate more complex phonologies than vocabularies with shorter words, lower word-level costs, and higher system-level costs. It was predicted that phonological complexity would be more likely to emerge under fast word acquisition than under slow word acquisition, since the system-level costs were expected to outweigh word-level costs earlier in acquisition. The simulation results are presented next.

6. <u>Results</u>. No matter what the rate of word acquisition, child vocabularies became more complex as they grew in size. Words became longer, composed of more elaborate sounds (i.e. multiple articulators were involved), more complex syllable structures, and more distinctive sound sequences. Rate of acquisition also significantly affected complexity. Comparisons between size-matched vocabularies attained at different rates showed that slow-growth vocabularies had shorter words with simpler sounds and simpler syllable structures than either medium- or fast-growth vocabularies. Thus, the simulation results are qualitatively similar to the patterns observed in natural language acquisition where simpler structures are acquired earlier than more complex ones, and children who acquire words slowly have simpler phonologies than children who acquire words quickly. The simulation results are given in more detail below.

6.1. <u>Word length</u>. There was a marked preference for short words over long words in the simulated child vocabularies—all had many more monosyllabic words than disyllabic or polysyllabic words. The early, smaller vocabularies were composed almost exclusively of monosyllabic words. Disyllabic words were not added in significant numbers until vocabulary size exceeded 100 words. These results are shown in Figure 3.

Insert Figure 3 about here.

Figure 3 also shows the statistically significant effect of word acquisition rate on word length [$F(2,72) = 40.50$, $\underline{p} < .01$] and the significant interaction with vocabulary size [Size × Rate, $F(12,432) = 8.98$, $\underline{p} < .01$; Size, $F(6,216) = 397.21$, $\underline{p} < .01$]. The pattern of interaction varied somewhat depending on the source vocabulary in question [Rate × Size × Source, $F(36,432) = 2.354$, $\underline{p} < .01$], but word length always differed in larger slow- and fast-growth vocabularies (< 100 words). From the Figure, it appears that the interaction between acquisition rate and vocabulary size was due to the larger numbers of monosyllabic words in slow-growth vocabularies than in medium- and fast-growth vocabularies, and a corresponding smaller number of disyllabic words.

In addition, one can infer from the large numbers of mono- and disyllabic words shown in Figure 3 that very few polysyllabic words were ever acquired in the child vocabularies, even when words were acquired quickly. For instance, the mean number of trisyllabic words in the 300 word vocabularies of the fast-growth condition was 4.20 (SD = 1.99). Only two of the fast-growth vocabularies, out of the 40 developed, had four-syllable words, and in each case there was only 1 instance of such a word.

The small number of disyllabic and polysyllabic words in the child vocabularies could indicate that monosyllabic word shapes predominated in the adult source vocabularies, but this was not the case. On average, 35.24% of the words in the source vocabularies were monosyllabic, 51.78% were disyllabic and 12.98% were polysyllabic (standard deviations were 0.47%, 0.40%, and 0.37%, respectively). So, the small number of disyllabic and polysyllabic words in the child vocabularies is not due to the source.[16] Rather, it indicates that articulatory costs were minimized in child vocabularies by selecting from among the monosyllabic words of the source. Given that articulatory costs accrue with the number of movements across a word, as long as monosyllables contain only simply-articulated consonants, these are likely to be less costly than equally simple disyllables. The preference for monosyllabic words also indicates that the pressures for salience and distinctiveness could be satisfied in vocabularies by selecting from among

different monosyllabic word shapes, and that this solution was less costly than increasing word length.

6.2. Word-level structure. The relative phonological complexity of the sounds and syllables of a word is reflected in its articulatory and inherent perceptual costs. As just noted, articulatory costs are calculated by summing the number of movements required to execute the sounds of a word. This cost therefore penalizes multiply articulated segments and dissimilation in articulatory place and manner between consecutive segments. The inherent perceptual cost in VEM is based on syllable complexity. More complex syllable structures receive higher penalties than simpler structures. Increasing phonological complexity can therefore be captured as increases in the average articulatory and inherent perceptual costs of all the words in the vocabulary. The results on the articulatory costs associated with different rates of word acquisition and vocabulary size are presented first, followed by those on the inherent perceptual costs.

Statistical analysis indicated that the average articulatory cost of words in the child vocabularies increased steadily during word acquisition, but that this effect of vocabulary size interacted with the rate at which words were acquired [Rate $\times$ Size, $F(12,432) = 11.23$, $p < .01$; Rate, $F(2,72) = 71.84$, $p < .01$; Size, $F(6,216) = 1899.89$, $p < .01$]. In particular, the articulatory costs of larger vocabularies (≤ 100 words) were different depending on word acquisition rate. Slow-growth vocabularies had smaller articulatory costs than medium- and fast-growth vocabularies. For instance, comparisons of 300 word vocabularies in Figure 4 show relatively large rate-dependent differences between slow-, medium-, and fast-growth vocabularies. The identity of the source vocabulary also had some effect on the overall pattern of results just described [Rate $\times$ Size $\times$ Source, $F(36,432) = 1.61$, $p < .05$]; slow-growth vocabularies were differentiated from medium-growth vocabularies either at 100 words or 150 words depending on the source.

Insert Figure 4 about here.

Given the previous discussion relating word length and articulatory costs, one might wonder whether the observed differences in articulatory costs between slow-, medium-, and fast-growth vocabularies reflect anything other than the previously described effects of acquisition rate and vocabulary size on word length. To evaluate this possibility, the average articulatory cost for a vocabulary was divided by its average word length. Statistical analysis on the resulting normalized articulatory costs showed that the rate of word acquisition still interacted with vocabulary size [Rate $\times$ Size, $F(12,432) = 3.04$, $\underline{p} < .01$; Rate, $F(2,72) = 3.43$, $\underline{p} < .05$; Size, $F(6,216) = 1653.23$, $\underline{p} < .01$], but the pattern of interaction was quite different, as can be seen in Figure 5.

Insert Figure 5 about here.

Specifically, Figure 5 shows that word acquisition rate influenced only the articulatory costs of smaller vocabularies. Slow-growth vocabularies with 50 and 100 words were less costly—composed of simpler sounds and less differentiated sound sequences—than fast-growth vocabularies. Medium-growth vocabularies patterned between slow- and fast-growth vocabularies. These overall results varied somewhat depending on the source vocabulary [Rate $\times$ Size $\times$ Source, $F(36,432) = 1.67$, $\underline{p} = .01$]. In particular, the differences in articulatory costs for slow- and medium-growth vocabularies shown in Figure 5 were only striking in vocabularies acquired from one of the four source vocabularies and were nearly absent in those vocabularies acquired from another of the four sources.

The pattern of results for inherent perceptual costs were qualitatively similar to those just described for the normalized articulatory costs, but only the main effects of

acquisition rate and vocabulary size were significant [Rate, $F(2,72) = 10.31$, $p < .01$; Size, $F(6,216) = 762.58$, $p < .01$].

―――

Insert Figure 6 about here.

―――

Figure 6 shows that inherent perceptual costs, which reflect syllable structure complexity, increased as vocabulary size increased. Even though the interaction between rate and size was not statistically significant, it is clear from the Figure that word acquisition rate affected syllable structure in smaller vocabularies (> 200 words). In such vocabularies, slow word acquisition resulted in simpler syllable structures than fast word acquisition. Post hoc tests indicated that these differences were significant in vocabularies of 50, 100, and 150 words ($p < .007$, where $p = .007$ is the corrected alpha level), but not in the initial vocabularies (5 words, $p = .33$) or in vocabularies of 200 words ($p = .57$) and above.

Figure 6 also shows that inherent perceptual costs increased more during early vocabulary growth, between 50 and 200 words, than during later growth, between 200 and 300 words. This pattern of increase is probably due to a trading relation between word length, word-level sound shape complexity, and the system-level distinctiveness costs. In vocabularies composed almost entirely of monosyllabic words, system-level distinctiveness is maintained through sound and syllable shape diversity. This increase in diversity accounts for the rapid increase in the inherent perceptual costs of words during early vocabulary growth. At about 200 words, the costs of sound and syllable structure complexity appears to outweigh the cost of increasing word length, and more disyllabic and polysyllabic words are acquired (see Figure 3). Distinctiveness from this point on is thus maintained by selecting words that recombine sound patterns that have already been selected into the vocabulary.

6.3. <u>Vocabulary-level structure</u>. The system-level costs provided two measures of distinctiveness by assessing phonological homogeneity/diversity in the simulated vocabularies. Although homogeneity was expected to increase with vocabulary size as the sound space became more crowded, fast-growth vocabularies were predicted to maintain greater overall diversity than slow-growth vocabularies. This prediction was based on the assumption that the pressures for ease and distinctiveness oppose one another and that the pressures for distinctiveness would outweigh pressures for ease when words are acquired quickly. Since slow-growth vocabularies minimize articulatory costs over the course of acquisition, they were expected to be more homogeneous in structure and sequencing than fast-growth vocabularies, which are not as effective at minimizing articulatory costs. The results provide little support for this prediction, and instead suggest that there are different strategies for maintaining distinctiveness in developing vocabularies.

Figure 7 shows that paradigmatic homogeneity, that is, similarity in the sounds and structures of syllable onsets, offsets, and nuclei in a vocabulary, increased with vocabulary size [$F(6,270) = 171466.6$, $\underline{p} < .01$]. The Figure also shows that larger (< 150 words) slow-growth vocabularies maintained greater paradigmatic diversity than larger medium- and fast-growth vocabularies [Rate $\times$ Size, $F(12,432) = 118.10$, $\underline{p} < .01$; Rate, $F(2,72) = 489.09$, $\underline{p} < .01$]. This pattern of results was invariant across vocabularies developed from different sources.

Insert Figure 7 about here.

The greater overall diversity of syllable onset, offset, and nucleus types (lower paradigmatic homogeneity costs) in slow-growth vocabularies compared with medium- and fast-growth vocabularies was most likely due to the greater number of monosyllabic words in slow-growth vocabularies. As noted above, there are trade-offs between word

length, word-level sound shape, and system-level distinctiveness. The results suggest that the repertoire of onsets, offsets, and nuclei in slow-growth vocabularies was more diverse because slow word acquisition favored the selection of new monosyllabic forms (i.e. forms with identical shapes, but new sounds—for example, 3 different CVs as in ba, gi, du) over longer words that recombine existing forms (e.g. CV and CVCV as in ba, gi, and bagi). Importantly, the inherent perceptual costs of words was the same in large slow- and fast-growth vocabularies (see results under section 6.2), indicating that slow-word acquisition favored monosyllabic words with identical syllable shapes and different sound combinations (e.g. as in bag and gab) rather than those with novel sounds and syllable shapes (e.g. as in bag and brag).

The unexpected effect of acquisition rate on paradigmatic diversity suggests that distinctiveness can be maintained without sacrificing articulatory ease or perceptual salience. This suggestion is further supported by the results on syntagmatic homogeneity/diversity.

Syntagmatic homogeneity across the system increases with the number of position-by-position similarities across words and with the number of words compared. Accordingly, homogeneity increases with vocabulary size, as shown in Figure 8.

Insert Figure 8 about here.

Figure 8 also shows an effect of acquisition rate on syntagmatic homogeneity. A statistical analysis confirms what is just discernable in Figure 8, namely, that acquisition rate once again interacted with vocabulary size, though not with source vocabulary [Rate $\times$ Size, $F(12,432) = 26.03$, $p < .01$; Rate, $F(2,72) = 51.42$, $p < .01$; Size, $F(6,216) = 292082.9$, $p < .01$]. Specifically, larger slow-growth vocabularies (< 150 words) had words with more diverse segment sequences than larger fast-growth vocabularies. This finding is probably due to the greater number of homophones selected during fast word

acquisition than during slow word acquisition [fast vs. slow, F(1,36) = 15.00, p <.01]. If homophones represent a special class of words that might be reasonably excluded from an analysis of distinctiveness, then the results show that fast and slow growth vocabularies are equally distinctive systems [fast vs. slow, F(1,36) = .752, p = .39].

Overall, then, the results on paradigmatic and syntagmatic homogeneity suggest that there are different strategies for maintaining distinctiveness of words in an expanding vocabulary: the recombination of sounds within a single syllable shape or the recombination of syllables.

6.4. Summary. As predicted, (1) word length increased with vocabulary size in the simulated child vocabularies, (2) the phonological structure of words that were acquired later was structurally more complex than those that were acquired earlier, and (3) the sound space became more crowded as more and more words were added to expanding vocabularies. These patterns that characterized vocabulary expansion in VEM also describe vocabulary acquisition in children. Consider, for example, the progression from early words such as ball, more, and no to words such as hot, baby and shoes, and then to words such as cracker, swing, and tiger.[17] VEM provides us with an elegant explanation for this progression from the simple to more complex, namely, that it the natural result of an interaction between phonetic selection pressures and the word acquisition rate, which increases with vocabulary size.

The simulation results also show that different rates of word acquisition can explain differences in the structure of developing systems. Early on in vocabulary development, slow word acquisition results in words with simpler syllable structures, smaller articulatory transitions from segment to segment, and simpler segments. In this way, the results parallel the natural language finding that late talkers have simpler syllable structures and more limited phonetic repertoires than their age-matched or language-matched peers (Schwartz, Leonard et al. 1980; Stoel-Gammon 1989; Paul &

Jennings 1992; Rescorla & Ratner 1996; Aguilar-Mediavilla et al. 2002). The result that rate-dependent differences disappear in VEM when articulatory costs are outweighed by vocabulary-level distinctiveness costs parallels the observation that many late talkers catch up with their typically-developing peers during the preschool years (Paul 1993; Leonard 1998). Children who continue to show delays beyond the preschool years are usually diagnosed with Specific Language Impairment. This impairment presumably involves deficits that are not characterized in VEM (see e.g. Fee 1995; Leonard 1998; Joanisse & Seidenberg 2003).

7. <u>Discussion</u>. VEM provides a specific explanation for rate-dependent differences in the acquisition of phonological structure: the rate at which new words are acquired affects the performance of the phonetic filter vis-à-vis word-level costs. The filter more effectively minimizes word-level articulatory and perceptual costs when few words are selected into the vocabulary at any given point in time. Word-level costs are balanced against system-level costs during each stage of vocabulary development, but word-level costs are constant costs while system-level costs increase with the size of a system.[18] Accordingly, word-level costs that favor simple, salient structures are more important in small vocabularies with lower system-level costs than in large vocabularies with higher system-level costs. In larger systems, word-level costs are subsumed by system-level costs that favor diversity of structure, and so complexity. This explanation for the emergence of phonological complexity is consistent with the intuition that phonetic constraints on the acquisition of linguistic structure are slowly overridden by a pressure to communicate: a distinctive message is more easily conveyed and remembered than a nondistinctive one.

VEM characterizes more than emergent complexity within a single system, it also provides a novel way to understand why the expressive phonologies of typically developing children differ from those of children with language delay. Specifically, the

results suggest that children with language delay have simpler phonologies because word-level costs remain relatively more important for longer than system-level costs. This explanation is consistent with the view that language delay is slow, not deviant (Paul & Jennings 1992; Leonard 1998; Mirak & Rescorla 1998). VEM hypothesizes that the filter is the same for all children, just more effective vis-à-vis the word-level articulatory and perceptual costs when fewer words are being selected and stored over time. The simulation results also suggest that different ambient language conditions may interact with the effects of acquisition rate. Such an interaction is consistent with large intra-group variability of phonological abilities, which is also reported in studies of children with language delay (e.g. Thal et al. 1995). This additional parallel with natural language further supports the simple explanation that VEM provides for rate-dependent differences in emergent phonology.

7.1. Emergent differences. VEM was built to understand how rate-dependent differences may emerge in phonology. In pursuing this goal, other assumptions were formalized. For instance, VEM formalizes the respectable assumption that phonological knowledge emerges from the lexicon. Once formalized, the assumption leads to the interesting conclusion that phonological differences emerge when stored word shapes differ, which could be restated as a hypothesis about how language acquisition works. This hypothesis is explored here as an example of the novel research questions that computational modeling generates.

The hypothesis that individual differences in phonological knowledge emerge from differences in stored word shapes provides a common sense explanation for language-specific differences in phonological knowledge. For example, if phonemic knowledge is abstracted from the representation of sounds that serve to contrast meaning in a language, and different languages use different sounds to contrast meaning, then language-specific differences in phonemic knowledge must emerge from the different

sets of sounds used to contrast meaning in different languages. An interesting question is whether a similar explanation applies to the intra-individual differences in the phonologies of early and later word acquisition and to the inter-individual differences in the phonologies of children acquiring the same language at different rates, as VEM suggests.

Phonological differences between younger and older children must be at least partly attributed to the fact that older children have acquired more distinct word shapes than younger children. Such an explanation restates the observation that children first acquire words with phonological structure that they find easier to produce (i.e. phonological selectivity), and then go on to acquire other, more diverse forms. The explanation is also consistent with studies showing that children with larger expressive and receptive vocabularies are better at word and phoneme discrimination tasks than children with smaller vocabularies (Munson 2001; Edwards et al. 2002), and that lexical training increases phonological complexity in late-talking children (Girolametto et al. 1997). Can rate-dependent phonological differences be explained in terms of vocabulary differences once vocabulary size has been controlled? There is some indication that they can be: there is evidence that children with language delay acquire different lexical items than typically developing children. For example, Rescorla et al. (2001) compared the vocabularies of such children and found differences between the groups, even when they were matched on vocabulary size. Examining the phonological structure of these words would provide a necessary test of the hypothesis that phonological differences can be attributed to the acquisition of different word shapes.

7.2. <u>Word representation</u>. Another assumption formalized in VEM is that word shapes are stored and represented as adult-like phonetically abstract forms, hence the hypothesis that phonological differences emerge from the acquisition of different word shapes / lexical items. An alternative hypothesis is that the same words are represented

differently in different speakers (see e.g. Charles-Luce & Luce 1990; 1995; Vihman 1996; Storkel 2002). According to this hypothesis, phonology still emerges from the lexicon, but the relationship between vocabulary growth and phonological structure is not as direct as the one encoded in VEM.

In adopting the assumption that word shapes are represented in an adult-like manner, VEM assumes that the representation involves only acoustic information. There is strong evidence for such representations, even in very young children (Swingley & Aslin 2002; see also Coleman 1998). However, adult psycholinguistic research suggests that word shape representation is multi-dimensional: a lexical item is linked to acoustic, articulatory, and even orthographic shapes (Treiman, Clifton, Meyer, Wurm 2003). Although orthographic information is not a prerequisite for phonological knowledge (consider preliterate individuals and societies), articulatory information is likely to be as important as acoustic information. Indeed, some researchers argue that articulatory information is sufficient to abstract phonological patterns (e.g. Browman & Goldstein 1986; 1992). The generative tradition implicitly makes this argument as well in adopting articulatorily-based distinctive features over acoustically-motivated features.

So, although a very young child may in fact store an adult target word in acoustic detail, is this really the relevant representation for phonological knowledge? This difficult question, raised here by VEM's architecture, merits continued research. On the one hand, young children must abstract over stored acoustic representations. Speech perception studies show that by about 10 months of age infants have abstracted phonemic categories for the sounds in their native language, and have lost the ability to perceive non-native contrasts (Werker & Tees 1984; Werker & Lalonde 1988; Best 1993; Polka & Werker 1994; but see Sundara et al. 2006 and footnote 6). Other studies demonstrate that similarly young infants are sensitive to language-specific phonotactic and stress patterns (Jusczyk, Cutler, Redanz 1993; Jusczyk, Luce, Charles-Luce 1994). In addition, babbling studies show that these sensitivities are reflected in production biases. The preferred

segments and rhythmic patterns of babbling differ in infants from different language backgrounds, and these differences relate to the frequency distributions in the ambient language (de Boysson-Bardies & Vihman 1991; Levitt & Wang 1991).

On the other hand, other evidence from production suggests that phonological abstractions may not be well-defined and adult-like until articulatory information is also stored in an adult-like manner (see Vihman 1996). Anecdotal evidence from one child (the first author's eldest son) illustrates this point nicely. By age 4, Max did not distinguish between /s/ and /θ/ in production—he persistently replaced /s/ with /θ/. Thinking this was only a problem of articulation, MAR proceeded to teach Max how to produce /s/. Upon being taught the proper tongue placement, Max promptly produced all /θ/s as /s/s irrespective of the target sound in the word (e.g. target sick became /sɪk/, target thick also became /sɪk/). Although the generalization did not persist, it was nonetheless surprising because the acoustic signatures of /s/ and /θ/ are not readily confusable perceptually in English (e.g. Miller & Nicely 1955; Redford & Diehl 1999) and Max easily distinguished the different meanings encoded in such minimal pairs as sick and thick. Such facts lead to the conclusion that Max had collapsed /s/ and /θ/ based on their undifferentiated articulatory representation rather than their undifferentiated perceptual representation.[19] This anecdote in turn suggests that VEM should be elaborated in future work to include different types of word representations to better model emergent phonology in first language acquisition.

7.3. Future directions. Given that both articulatory and acoustic information are no doubt crucial to forming phonological representations, future work will be aimed at elaborating VEM so that word shapes are linked to their articulatory actualizations as well as to their acoustic targets. Word selection from the source will be based on the same word- and system-level costs. Words thus selected will constitute the developing receptive vocabulary. A subset of the receptive vocabulary items will be simultaneously

selected for production attempts, and thus will constitute a developing expressive vocabulary. The production attempts will be represented as simplified child-like pronunciations of receptive vocabulary items (e.g. /ɑt/ for hot) to reflect systematic mispronunciation. The articulatory forms will become more accurate as similar receptive forms are selected and attempted as expressive forms, thereby modeling the effects of practice on attaining some acoustic-perceptual target. Receptive and expressive forms are hypothesized to jointly specify developing phonological knowledge. VEM could be used to investigate how these two forms interact to define the emergence of phonological structure, and whether one set of representations is weighted more heavily than another during particular points of development. The rate of word acquisition should affect developing structure in the revised model as it does in the current version of the model. As long as the expressive vocabulary can be characterized as a subset of the receptive vocabulary and practice effects refine production attempts, slow word acquisition can favor the simplicity and salience of word-level structure over the diversity and complexity of system-level structure.

In addition to investigating the relative influences of the receptive and expressive vocabularies on developing knowledge, future work will also consider how phonological knowledge produces feedback in the system to affect word representation. The long-term goal of this work is to provide a detailed, formal account of how phonology and the lexicon interact in child language.

8. Conclusion. This paper described a computational model of phonological development that formalizes an emergentist view of language structure. In VEM, phonological structure emerges from the word shapes represented in a developing vocabulary. These shapes are selected from a source, which represents the subset of words in the adult vocabulary that are accessible to a child. The source is filtered by the same phonetic constraints that are assumed to influence sound change in language;

namely, the constraints imposed by production and perception. These constraints are conceived of either as word-level costs or as system-level costs, and structure emerges as the constraints are balanced against one another in the selection of words that best minimize the various costs at any given time. The model shows that this balancing act need not necessarily favor one set of constraints over another. In particular, word-level ease and salience do not have to be sacrificed for system-level perceptual distinctiveness and vice versa. The balance between word-level and system-level costs is struck when external factors, such as the rate of word acquisition, interact with constraint optimization to effect the types of words selected into a vocabulary. The simulations demonstrate that slow word acquisition minimizes word-level costs better than fast word acquisition, resulting in a vocabulary with overall simpler phonological structure; and that slow word acquisition results in a different, but equally good strategy for maintaining system-level distinctiveness as fast word acquisition. This explanation of emergent differences, derived from the explicit theory instantiated in VEM, provides a framework within which the correlation between acquisition rate and phonological structure can be explored further in studies on normal and delayed phonological development.

Although VEM's validity depends partly on how well it can match real language data, the model was developed primarily to understand the effects of constraint interaction on emergent language structure at a more abstract level. The specific goal was to demonstrate how the oft-mentioned factor rate of acquisition may influence the structure of a developing system. This paper investigated the case of how the rate at which words are acquired may influence the types of word shapes that are acquired, but acquisition rate is also relevant to other aspects of language evolution. The idea that the rate at which a system changes can influence its structure may also be relevant to discussions of language development (e.g. the emergence of creoles) and to change due to borrowing or to the spread of an innovation through the lexicon, insofar as any of these processes can be reconceptualized in terms of selection on variation. If language structure

is indeed emergent from selection on variation, then VEM provides insight into how rate

may provide a strong influence on its continually changing shape.

REFERENCES

Aguilar-Mediavilla, Eva M.; Monica Sanz-Torrent; and Miquel Serra-Raventos. 2002. A comparative study of the phonology of pre-school children with specific language impairment (SLI), language delay (LD) and normal acquisition. Clinical Linguistics and Phonetics 16.573—96.

Anttila, Raimo. 1989. Historical and comparative linguistics. 2nd edn. Amsterdam: John Benjamins.

Baldwin, Dare A. and Ellen M. Markman. 1989. Establishing word-object relations: A first step. Child Development 60.381—98.

Bates, Elizabeth; Philip S. Dale; Donna Thal. 1995. Individual differences and their implications for theories of language development. Handbook of child language, ed. by Paul Fletcher and Brian MacWhinney, 96—151. Oxford: Blackwell.

Baum, Shari and James C. McNutt. 1990. An acoustic analysis of frontal misarticulation of /s/ in children. Journal of Phonetics 18.51—63.

Beckman, Mary E. and Jan Edwards. 2000a. The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. Child Development 71.240—49.

Beckman, Mary E. and Jan Edwards. 2000b. Lexical frequency effects on young children's imitative productions. Papers in Laboratory Phonology V: Language acquisition and the lexicon, ed. by Michael Broe and Janet Pierrehumbert, 207—17. Cambridge: Cambridge University Press.

Benkí, Jose R. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60.129—157.

Best, Catherine. 1993. Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. Developmental neurocognition: Speech and face processing in the first year of life, ed Bénédict de

Boysson-Bardies, Scania de Schonen, Peter Jusczyk, Peter McNeilage, and John Morton, 289—304. Dordrecht: Kluwer Academic Publishers.

Blevins, Juliette. 2004. Evolutionary phonology: the emergence of sound patterns. Cambridge: Cambridge University Press.

de Boer, Bart. 2000. Self-organization in vowel systems. Journal of Phonetics 28.441—65.

de Boysson-Bardies, Bénédict and Marilyn May Vihman. 1991. Adaptation to language: Evidence from babbling and first words in four languages. Language 67.297—319.

Browman, Catherine P. and Louis M. Goldstein. 1986. Towards an articulatory phonology. Phonology Yearbook 3.19—252.

Browman, Catherine P. and Louis M. Goldstein. 1992. Articulatory phonology: An overview. Phonetica 49.55—180.

Charles-Luce, Jan and Paul A. Luce. 1990. Similarity neighbourhoods of words in young children's lexicons. Journal of Child Language 17.05—15.

Charles-Luce, Jan and Paul A. Luce. 1995. An examination of similarity neighbourhoods in young children's receptive vocabularies. Journal of Child Language 22.27—35.

Choi, Soonja and Alison Gopnik. 1995. Early acquisition of verbs in Korean: A cross-linguistic study. Journal of Child Language 22.97—529.

Coleman, John. 1998. Cognitive reality and the phonological lexicon: A review. Journal of Neurolinguistics 11.95—320.

Demolin, Didier and Soquet, Alain. 2001. The role self-organization in the emergence of phonological systems. Evolution of Communication 3.1—48.

Dromi, Esther 1987. Early lexical development. New York: Cambridge University Press.

Edwards, Jan; Robert A. Fox; and Catherine L. Rogers. 2002. Final consonant discrimination in children: Effects of phonological disorder, vocabulary size, and articulatory accuracy. Journal of Speech, Language, and Hearing Research 45.31—42.

Edwards, Jan and Margaret Lahey. 1998. Nonword repetitions of children with specific language impairment: Exploration of some explanations for their inaccuracies. Applied Psycholinguistics 19.79—309.

Elman, Jeffery L.; Elizabeth A. Bates; Mark H. Johnson; Annette Karmiloff-Smith; Domenico Parisi; and Kim Plunkett. 1996. Rethinking innateness: A connectionist perspective on development. Cambridge, MA: MIT Press.

Fee, E. Jane. 1995. The phonological system of a specifically-language impaired population. Clinical Linguistics and Phonetics 9.89—209.

Fenson, Larry; Philip S. Dale; J. Steven Reznick; Elizabeth Bates; Donna J. Thal; and Stephen J. Pethick. 1994. Variability in early communicative development. (Monographs of the Society for Research in Child Development, Serial No. 242, Vol. 59, No. 5.) Malden, MA: Blackwell..

Ferguson, Charles A. and Carol B. Farwell. 1975. Words and sounds in early language acquisition. Language 51.19—39.

Foulkes, Paul; Gerard Docherty; and Dominic Watt. 2005. Phonological variation in child-directed speech. Language 81.77—206.

Francis, Nelson W. and Henry Kucera. 1982. Frequency analysis of English usage. Boston: Houghton Mifflin.

Ganger, Jennifer and Michael R. Brent. 2004. Reexamining the vocabulary spurt. Developmental Psychology 40.621—32.

Gentner, Dedre. 1982. Why nouns are learned before verbs: Linguistic relativity versus natural partintioning. Language development, Vol. 2: Language, thought, and culture, ed. by Stan A. Kuczaj, 301—34. Hillsdale, NJ: Erlbaum.

Girolametto, Luigi; Patsy S. Pearce; and Elain Weitzman. 1997. Effects of lexical intervention on the phonology of late talkers. Journal of Speech, Hearing, and Language Research 40.338—48.

Goldfield, Beverly A. and J. Steven Reznick. 1990. Early lexical acquisition: Rate, content, and the vocabulary spurt. Journal of Child Language 17.171—84.

Goldstein, Louis and Carol A. Fowler. 2003. Articulatory Phonology: A phonology for public language use. Phonetics and phonology in language comprehension and production: Differences and similarities, ed. by Niels O. Schiller and Antje S. Meyer, 159—208. Berlin: Mouton.

Gopnick, Alison and Andrew Meltzoff. (1987). The development of categorization in the second year and its relation to other cognitive and linguistic developments. Child Development 58.1523—31.

Guenther, Frank. 1995. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. Psychological Review 102.594—621.

Guenther, Frank; Michelle Hampson; and Dave Johnson. 1998. A theoretical investigation of reference frames for the planning of speech movements. Psychological Review 105.611—33.

Hart, Betty. 1991. Input frequency and children's first words. First Language 11.289—300.

Hura, Susan L.; and Björn Lindblom; Randy L. Diehl. 1992. On the role of perception in shaping phonological assimilation rules. Language and Speech 35.59—72.

Joanisse, Marc. 2000. Connectionist phonology. Los Angeles, CA: University of Southern California dissertation.

Joanisse, Marc and Mark S. Seidenberg. 2003. Phonology and syntax in specific language impairment: evidence from a connectionist model. Brain and Language 86.40—56.

Jusczyk, Peter; Anne Cutler; and Nancy J. Redanz. 1993. Preference for the predominant stress patterns of English words. Child Development 64.675—87.

Jusczyk, Peter W. and Paul A. Luce. 2002. Speech perception. Steven's handbook of experimental psychology, Vol. 1: Sensation and perception, ed. by Hal Pasler and Steven Yantis, 493—536. New York: John Wiley.

Jusczyk, Peter W.; Paul A. Luce; and Jan Charles-Luce. 1994. Infants' sensitivity to phonotactics patterns in the native language. Journal of Memory and Language 33.630—45.

Kingston, John and Randy L. Diehl. 1994. Phonetic knowledge. Language 70.419—54.

Leonard, Laurence. 1998. Children with specific language impairment. Cambridge, MA: MIT Press.

Levitt, Andrea G. and Qi Wang. 1991. Evidence for language specific rhythmic influences in the reduplicative babbling of French and English learning infants. Language and Speech 34.235—49.

Liljencrants, Johan and Björn Lindblom. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. Language 48.839—62.

Lindblom, Björn. 1990. Explaining phonetic variation: a sketch of the H&H theory. Speech production and speech modelling, ed. by William J. Hardcastle and Alain Marchal, 403—39. Dordrecht: Kluwer.

Lindblom, Björn. 1992. Phonological units as adaptive emergents of lexical development. Phonological development: Models, research, implications, ed. by Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon, 565—604. Timonium, MD: York Press.

Lindblom, Björn. 1996. Role of articulation in speech perception: Clues from production. Journal of the Acoustical Society of America 99.1683—92.

Lindblom, Björn; Peter F. MacNeilage; Michael Studdert-Kennedy. 1984. Self-organizing processes and the explanation of phonological universals. Explanations for Language Universals, ed. by Brian Butterworth, Bernard Comrie, Östen Dahl, 181—203. Berlin: Mouton.

Locke, John L. 1983. Phonological acquisition and change. New York: Academic Press.

MacNeilage, Peter F. and Barbara L. Davis. 2000. On the origin of internal structure of word forms. Science 288, 527—31.

MacNeilage, Peter F.; Barbara L. Davis; Ashlynn Kinney; Christine L. Matyear. 2000. The motor core of speech. Child Development 71.153—63.

MacNeilage, Peter F.; Barbara L. Davis; and Christine L. Matyear. 1997. Babbling and first words: Phonetic similarities and differences. Speech Communication 22.269—77.

Marcus, Gary F.; Stephen Pinker; Michael Ullman; Michelle Hollander; T. John Rosen; and Fei Xu. 1992. Overregularization in language acquisition. (Monographs of the Society for Research in Child Development, Serial No. 228, Vol. 57, No. 4.) Chicago, IL: University of Chicago Press.

Markman, Ellen. 1987. How children constrain the possible meaning of words. Concepts and conceptual development: Ecological and intellectual factors in categorization, ed. by Ulric Neisser, 225—87. Cambridge: Cambridge University Press.

Martinet, Andre. 1964. Elements of general linguistics. Translated by E. Palmer. Chicago: The University of Chicago Press.

McCune, Lorraine and Marilyn M. Vihman. 2001. Early phonetic and lexical development: A productivity approach. Journal of Speech, Language, and Hearing Research 44.670—84.

Menn, Lise. 1983. Development of articulatory, phonetic, and phonological capabilities. Language production, Vol. 2: Development, writing, and other language processes, ed. by Brian Butterworth, 3—50. London: Academic Press.

Miller, George and Patricia E. Nicely. 1955. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America 27.338—52.

Mirak, Jennifer and Leslie Rescorla. 1998. Phonetic skills and vocabulary size in late talkers: Concurrent and predictive relationships. Applied Psycholinguistics 19.1—17.

Munson, Benjamin. 2001. Relationships between vocabulary size and spoken word recognition in children aged 3 to 7. Contemporary Issues in Communication Science and Disorders 28, 20—29.

Ohala, John J. 1990. The phonetics and phonology of aspects of assimilation. Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech, ed. by John Kingston and Mary E. Beckman, 258—75. Cambridge: Cambridge University Press.

Ohala, John J. 1993. The phonetics of sound change. Historical linguistics: Problems and perspectives, ed. by Charles Jones, 237-278. London: Longman.

Paul, Rhea. 1993. Patterns of development in late talkers: Preschool years. Journal of Childhood Communication Disorders 15, 7—14.

Paul, Rhea and Patricia Jennings. 1992. Phonological behavior in toddlers with slow expressive language development. Journal of Speech and Hearing Research 35, 99—107.

Paul, Rhea and Mary E. Shiffer. 1991. Communicative initiations in late-talking toddlers. Applied Psycholinguistics 12.419—32.

Pharr, Aimee B., Nan Bernstein Ratner, Leslie Rescorla. 2000. Syllable structure development of toddlers with expressive specific language impairment. Applied Psycholinguistics 21, 429—49.

Plaut, David C. and Christopher T. Kello. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. The Emergence of Language, ed. by Brian MacWhinney, 381—415. Mahwah, NJ: Erlbaum.

Plunkett, Kim. 1993. Lexical segmentation and vocabulary growth in early language acquisition. Journal of Child Language 20.43—60.

Polka, Linda and Janet F. Werker. 1994. Developmental changes in perception of non-native vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance 20.421—35.

Port, Robert F. and Adam P. Leary. 2005. Against formal phonology. Language 81.927—64.

Redford, Melissa A.; Chun C. Chen; and Risto Miikkulainen. 2001. Constrained emergence of universals and variation in syllable systems. Language and Speech 44.27—56.

Redford, Melissa A. and Randy L. Diehl. 1999. The relative perceptibility of syllable-initial and syllable-final consonants. Journal of the Acoustical Society of America 106.1555—65.

Rescorla, Leslie; Amie Alley; and Joanne B. Christine. 2001. Word frequencies in toddler's lexicons. Journal of Speech, Language, and Hearing Research 44, 598—609.

Rescorla, Leslie and Nan Bernstein Ratner. 1996. Phonetic profiles of toddlers with specific expressive language impairments (SLI-E). Journal of Speech and Hearing Research 39.153—65.

Roberts, Jennifer A. 1998. Cognitive and linguistic development before and after the vocabulary spurt. Boston, MA: Boston University dissertation.

Schwartz, Jean-Luc; Louis-Jean Boë; Nathalie Vallée; and Christian Abry. 1997. The dispersion-focalization theory of vowel systems. Journal of Phonetics 25.255—86.

Schwartz, Richard G. 1988. Phonological factors in early lexical acquisition. The emergent lexicon: The child's development of a linguistic vocabulary, ed. by Michael D. Smith and John L. Locke (Eds.), 185—218. San Diego: Academic Press.

Schwartz, Richard G.; Laurence B. Leonard; M. Karen Folger; and M. Jeanne Wilcox. 1980. Early phonological behavior in normal-speaking and language disordered children: Evidence for a synergistic view of linguistic disorders. Journal of Speech and Hearing Disorders 45.357—77.

Scobbie, James M.; Fiona Gibbon; William J. Hardcastle; Paul Fletcher. 2000. Convert contrast as a stage in the acquisition of phonetics and phonology. Papers in Laboratory Phonology V: Language acquisition and the lexicon, ed. by Michael Broe and Janet Pierrehumbert, 194—207. Cambridge: Cambridge University Press.

Shore, Cecilia M. 1995. Individual differences in language development. Thousand Oaks: Sage.

Smith, Bruce L.; Karla K. McGregor; and Darcie DeMille. 2006. Phonological development in lexically-precocious two-year-olds. Applied Psycholinguistics 27.355—75.

Smith, Michael D.; John Locke. 1988. The emergent lexicon: The child's development of a linguistic vocabulary. San Diego: Academic Press.

Stoel-Gammon, Carol. 1989. Prespeech and early speech development of two late talkers. First Language 9.207—24.

Stoel-Gammon, Carol and Carla Dunn. 1985. Normal and disordered phonology in children. Baltimore: University Park Press.

Storkel, Holly L. 2002. Restructuring of similarity neighborhoods in the developing mental lexicon. Journal of Child Language 29.251—74.

Storkel, Holly L. 2004. The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. Journal of Speech, Language, and Hearing Research 47.1194—1212.

Storkel, Holly L. and Michele L. Moriesette. 2002. The lexicon and phonology: Interactions in language acquisition. Language, Speech, and Hearing Services in Schools 33.24—37.

Sundara, Megha; Linda Polka; and Fred Genesee. 2006. Language-experience facilitates discrimination of /d- ð/ in monolingual and bilingual acquisition of English. Cognition 100.369—88.

Swingley, Daniel and Richard N. Aslin. 2002. Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science 13.480—84.

Thal, Donna; Melinda Oroz; and Valerie McCaw. 1995. Phonological and lexical development in late-talking toddlers. Applied Psycholinguistics 16.407—24.

Treiman, Rebecca; Charles Clifton; Antje S. Meyer; and Lee H. Wurm. 2003. Language comprehension and production. Handbook of psychology: Experimental psychology, Vol. 4, ed. by Alice F. Healy and Robert W. Proctor, 527—47. New York, NY: John Wiley.

Vennemann, Theo. 1978. Phonetic and conceptual analogy. Readings in historical phonology: Chapters in the theory of sound change, ed. by Philip Baldi and Ronald N. Werth, 258—74. University Park: Penn State Univ Press.

Vihman, Marilyn M. 1981. Phonology and the development of the lexicon: Evidence from children's errors. Journal of Child Language 8, 239—64.

Vihman, Marilyn M. 1993. Variable paths to early word production. Journal of Phonetics 21, 61—82.

Vihman, Marilyn M. 1996. Phonological development: The origins of language in the child. Cambridge, MA: Blackwell.

Vihman, Marilyn M.; Charles A. Ferguson; Mary Elbert. 1986. Phonological development from babbling to speech: Common tendencies and individual differences. Applied Psycholinguistics 7, 3—40.

Werker, Janet F. and Chris E. LaLonde. 1988. Cross-language speech perception: initial capabilities and developmental change. Developmental Psychology 24.672—83.

Werker, Janet F. and Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 7.49—63.

Werker, Janet F. and Richard C. Tees. 1999. Influences on infant speech processing: Toward a new synthesis. Annual Review of Psychology 50.509—35.

FOOTNOTES

[1] A source-filter mechanism takes variable input and consistently filters it so that the output is structured. In phonetics, the source-filter theory describes the physical process by which certain frequencies in the glottal waveform are amplified by the shape of the supralaryngeal vocal tract while other frequencies are dampened. Evolution by natural selection is another source-filter theory: aspects of phenotypic variation are systematically selected for or against by forces affecting an organism's survival and reproductive success. Here, we adopt the metaphor of a source-filter mechanism in our discussion of phonological development to indicate our allegiance to the idea that phonetic processes help shape phonology as well as to the evolutionary principle that form follows function.

[2] Language acquisition researchers typically assess a child's output capability based on broad transcriptions of the child's spontaneous speech. Some may argue that in using broad transcriptions of child language, researchers implicitly assume that children represent lexical items in terms of discrete, serially-ordered phonemes (e.g. Scobbie, Gibbon, Hardcastle, Fletcher, 2000; Port & Leary, 2005). Some acquisition researchers may indeed make this assumption, but a more fundamental assumption is that children have acquired phoneme and syllable inventories that do or do not match the adult inventories. Importantly, the acquisition of these inventories need not imply transcription-like lexical representations. Instead, phonemes and syllables can emerge from articulatory / acoustic similarities across phonetically-detailed, but still abstract, representations. For example, lexical representations in Articulatory Phonology are time-varying scripts of coordinated articulatory movement (Browman & Goldstein, 1986; 1992). Whereas these scripts do not directly encode phonemes or syllables, these emerge from recurrent intergestural configurations and timing patterns across representations in the lexicon that serve, in some instances, to distinguish between two otherwise identical lexical scripts.

[3] Here, as elsewhere, phonology is operationalized as output capabilities. Girolametto et al. (1997) defined phonological complexity as the production of more different sounds and syllable structures.

[4] By paralleling our idea to the idea from historical linguistics, we have opened ourselves up to the same kind of criticism that this idea has received elsewhere. For example, Blevins (2004:55) argues that '(t)he most common teleological explanations in phonetics and phonology invoke notions of articulatory ease and perceptual distance.' We disagree, however, that our idea suggests that a particular change in the developing system '… occurs in order to minimize articulatory effort while another occurs in order to maximize perceptual contrast' (ibid). Rather, we are suggesting that a child's word selection is influenced both by her motoric abilities as well as by her percpetual system. Words that she believes she may be able to produce are selected as are words that capture her attention because they are salient and distinctive. Importantly, the child is not selecting words in order to end up with a certain kind of system. The resulting system nevertheless appears optimized for articulatory and perceptual factors.

[5] A reviewer noted that children may or may not produce simpler versions of the adult target. The usual assumption that children systematically simplify adult speech targets assumes that children have represented adult clear speech as the target. In reality, children more often hear adult casual speech than adult clear speech. This could mean that children have in fact represented phonetically reduced forms as the adult target, which in turn could mean that 'simplified' children's productions are in fact faithful to the target representation. This interesting possibility underscores how difficult it is to determine exactly what lexical representations might look like in the child. It also provides a strong argument for looking closely at the phonetic input to children in natural social-functional contexts, as some researchers have begun to do (e.g. Foulkes, Docherty, Watt 2005).

[6] This universalist picture of phonemic development is, however, complicated by more recent findings, which show that some contrasts are discriminated better <u>later</u> in development (circa 10 months) rather than earlier (circa 6 months). Sundara et al. (Sundara, Polka, & Genesee 2006) review this research and go on to show that speech perception abilities are continuing to develop even in children as old as 4 years of age. They conclude from these findings that aspects of speech perception abilities develop with language experience. We are sympathetic to this more complex view of perceptual development, but are not clear on how to model it.

[7] The rate of vocabulary acquisition is also most often measured and discussed in terms of production. Accordingly, this overview refers to the acquisition of expressive vocabulary rather than to that of receptive vocabulary. Although there are differences in the rate at which expressive and receptive vocabularies are acquired, the shape of the growth curves are similar (Fenson, Dale, Renick, Bates, Thal, Pethick 1994). This suggests that expressive and receptive vocabulary acquisition are governed by the same process, but that production merely lags perception as noted in the discussion of representation in section 2.2. above.

[8] Not all late talkers are considered to be merely delayed. A subgroup of late talkers is often diagnosed later (at age 5 or 6) as having Specific Language Impairment (SLI). Children with SLI have persistent problems with grammar, unlike their late talking peers. The clinical need to distinguish between late talkers who will catch up to their typically developing peers and those who will go on to be diagnosed with SLI has led researchers to focus on the similarities and differences in the early phonologies of children acquiring language at different rates. Since the goal is to weed out the merely delayed from the children with SLI, relatively little research focuses on the differences that exist between the phonologies of vocabulary-size matched delayed and typically developing children.

[9] Note that this algorithm allows the different segmental articulators to hold their position (neutral or target) over time, which is defined here in terms of jaw movement.

[10] In developing a computational model, the programmer must decide how to segment real-world continuous variables (such as movement) into discrete units. If one were to model speech kinematics, this problem could be solved by coding movement over a series of small temporal windows (say, into 5 millisecond bins to equal a 200 Hz sampling frequency). However, most would argue that it is not important to capture the fine grain details of speech kinematics when designing a model that aims to understand something about a developing phonology. Many may even chose not to represent speech at all, and to begin instead with phonemes (see e.g. Plaut & Kello 1999). However, VEM was built with the assumption that phonetic processes influence sound structure. It was therefore necessary to represent the phonetics of speech in some fashion.

[11] Note that segments (i.e. speech sounds) are equated here with articulatory configurations over some window of time. This equation merely simplifies the relationship between articulation and acoustics in the model. It is not meant as a position statement on the mapping between perception and production.

[12] Systematic gaps in the sound space are less likely to occur in VEM than in natural language, since VEM does not take into account any of the historical factors (losses, mergers, etc.) that can create gaps in sound systems.

[13] For instance, social and cognitive pressures are not encoded in VEM. This means that the source vocabularies will lack much of the regularity and sound pattern symmetry evident in the phonological structure of natural language vocabularies. Child vocabularies will also lack sound pattern regularities and symmetry, since the smaller vocabularies are derived from the larger one. Accordingly, our discussions on emergent phonological structure will be confined largely to word shape structure.

[14] We recognize that the concept of perceptual saliency may also apply to individual phones. For example, sounds with diffuse spectra and little energy (e.g. [f] and

[θ]) are less salient and therefore more confusable than sounds with greater energy and well-defined spectra (e.g.[s] and [ʃ]; Miller & Nicely 1955; Redford & Diehl 1999). Nonetheless, VEM evaluates perceptual saliency only in syntagmatic rather than in paradigmatic terms—a simplification that should not unduly impact the simulation goal, which is to understand rate effects on emergent structure.

[15] The values are provided here with the description of the model so that the simulation results can be replicated by other researchers. Although the relative weighting of the different syllabic elements is theoretically motivated, the specific values are arbitrary.

[16] On the other hand, the average number of monosyllabic words in the source still greatly outnumbered the total number of words in the child vocabularies, even during the latest stages of vocabulary expansion. If this were not the case, then monosyllabic words could not have been disproportionately selected into the developing vocabularies. In other words, child vocabularies do reflect characteristics of the source. The point being made here is that they are not mere reflections.

[17] These words are drawn from an unpublished diary that the first author kept of her eldest son's early language acquisition. Similar examples could have been drawn from Vihman (1996:249—60), though these published data are not as temporally fine-grained those in MAR's personal diary.

[18] An editorial comment prompts us to point out that while the phonetic costs of a system may increase with size, the cognitive costs may diminish. For example, symmetry may not be evident in a system until it reaches some critical size. If we assume that symmetry provides structure that allows for the more efficient categorization / representation of stored items, then cognitive costs may increase in a system until some threshold size is reached and symmetry emerges. At this point, cognitive costs may actually decrease or level off even while new items are acquired and the overall system size continues to increase. VEM does not currently formalize any cognitive costs, but the

simulation results are compatible with the idea that phonetic costs become less important to the structure of the emergent phonological system as the lexicon expands beyond a certain point.

[19] Two reviewers pointed out that Max may have been unusual in this respect. One reviewer noted that Max may have produced a contrast that was not readily apparent from simple listening (i.e. a <u>covert contrast</u>; Scobbie, Gibbon, Hardcastle, Fletcher 2000). Along the same lines, another reviewer pointed to evidence that 5 to 8 year old children, who appear to collapse /s/ and /θ/, actually produce acoustically different interdental fricatives depending on whether the target is /θ/ or /s/ (Baum & Mcnutt 1990). Clearly, the phonological representations in child language are more complex than the anecdote about Max's productions imply.
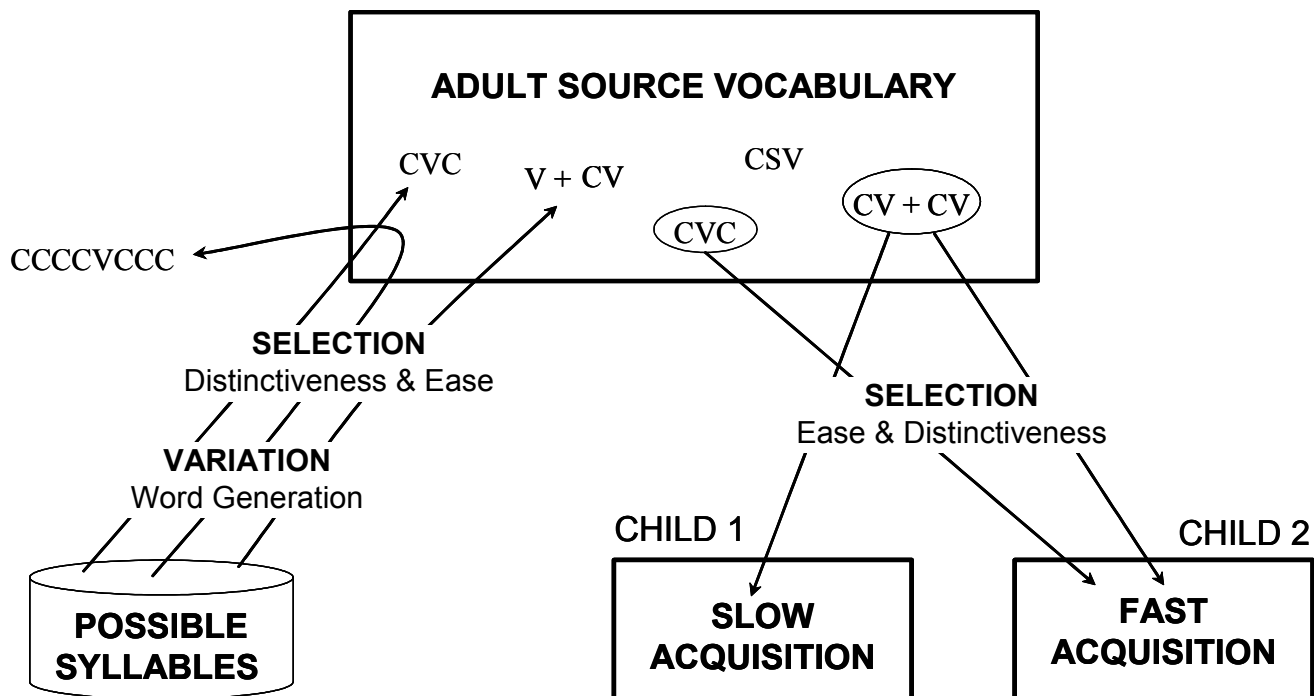
FIGURE 1

A schematic architecture of the Vocabulary Expansion Model is shown. A source vocabulary is developed by concatenating possible syllables to form possible words. These are then filtered according to the phonetic pressures for perceptual distinctiveness and articulatory ease. Child vocabularies are developed by selecting words from the source and filtering them according to the same phonetic pressures, but ease affects selection more than distinctiveness since the vocabularies are smaller overall. The rate at which words are selected and stored in the developing vocabularies is a manipulated variable.
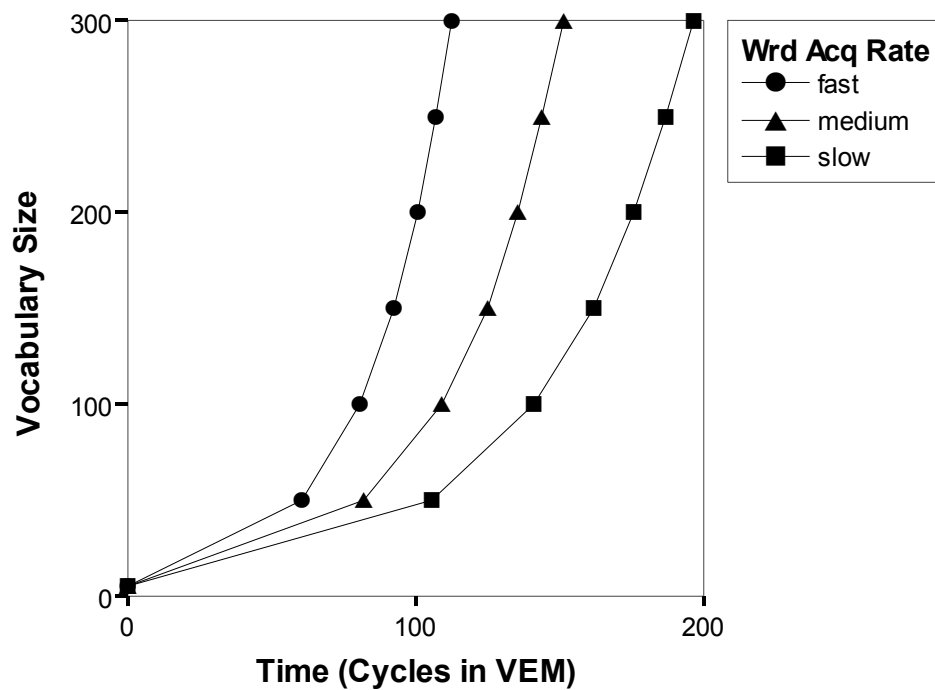
FIGURE 2

Word acquisition rate in the simulated child vocabularies was either fast, medium, or slow. The different rates of word acquisition resulted in the three different rates of vocabulary expansion that are shown.
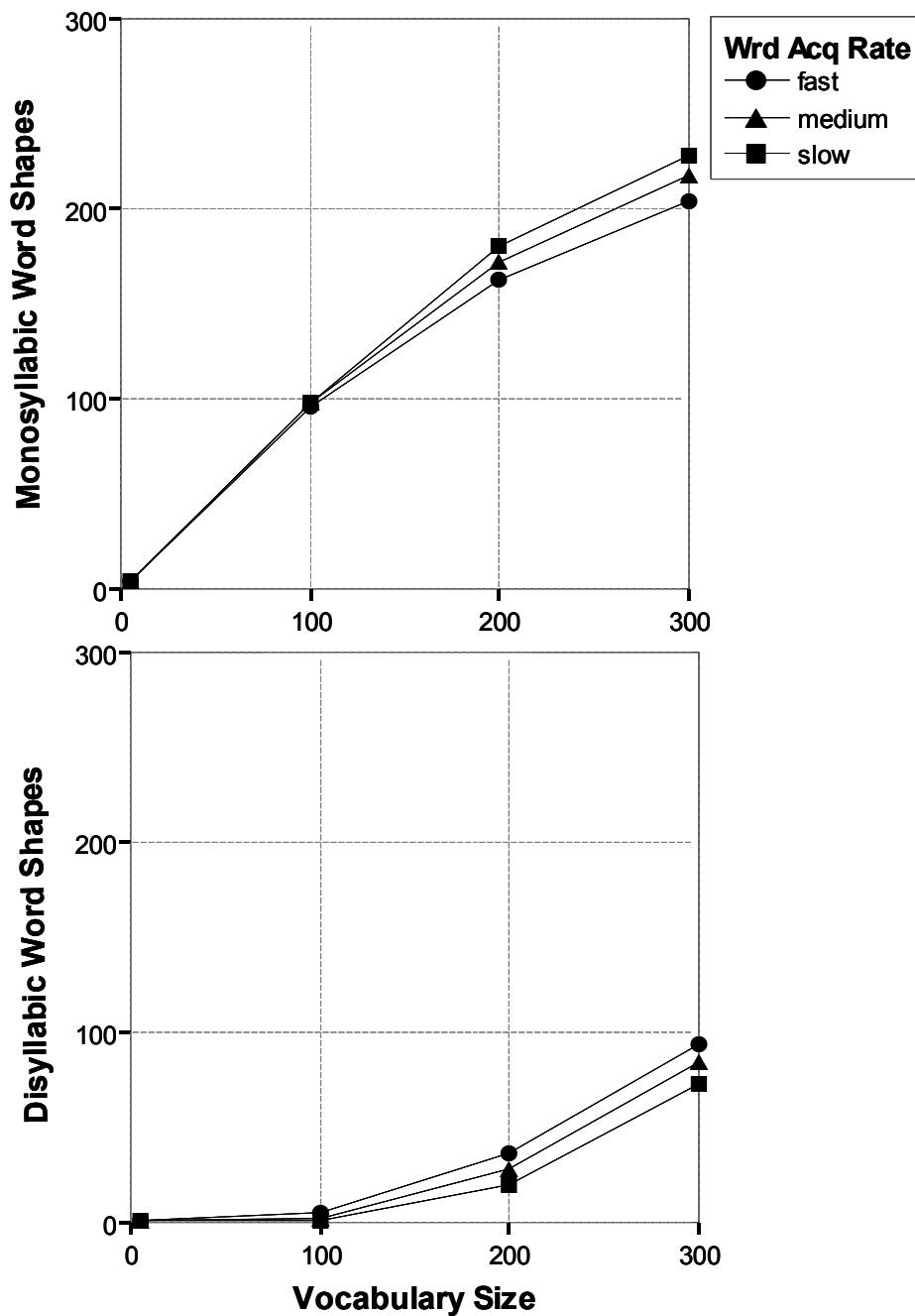
FIGURE 3

The number of monosyllabic and disyllabic words in the simulated child vocabularies are shown as a function of vocabulary size and word acquisition rate. Polysyllabic word shapes were also possible, but rarely occurred.
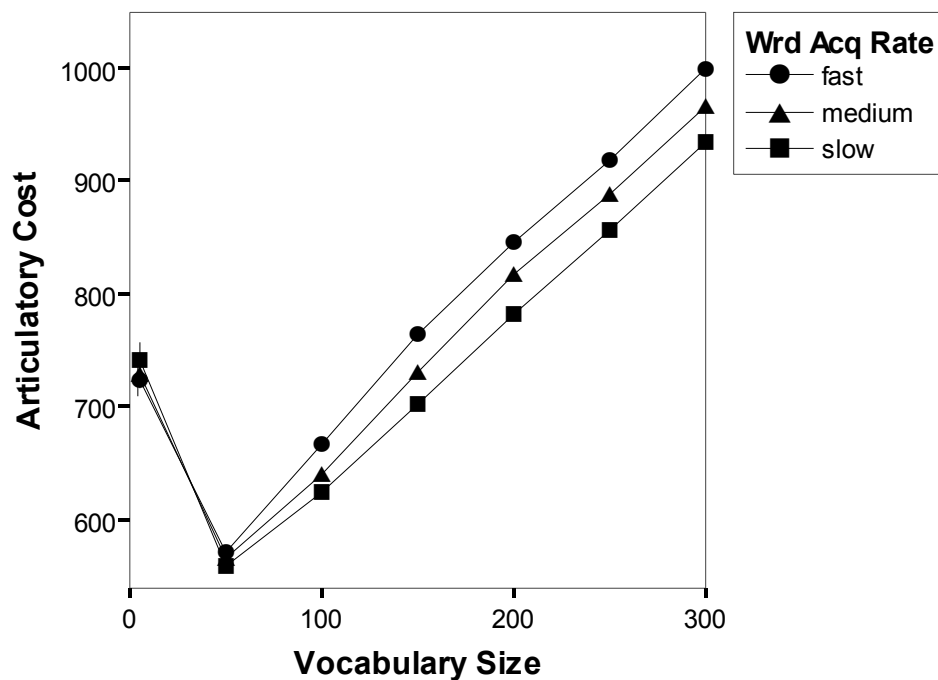
FIGURE 4

The average articulatory cost of words in the simulated child vocabularies is shown as a function of vocabulary size and word acquisition rate. The articulatory cost equals the number of articulatory movements required to produce a given sound shape. Accordingly, this cost increases as more complex articulations are used at the segmental level, and as adjacent segments are further and further dissimilated from one another.
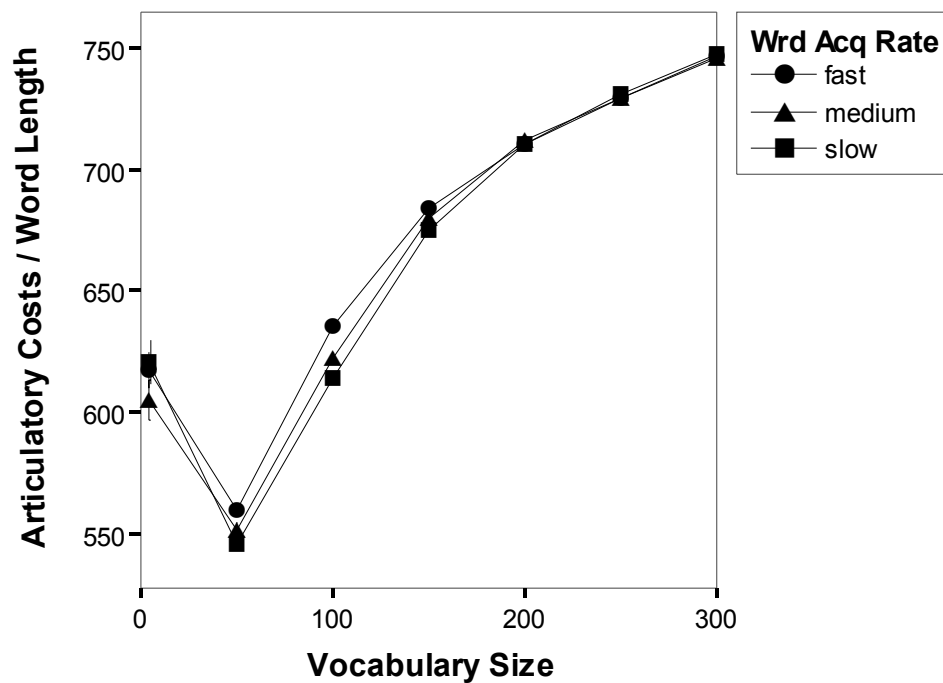
FIGURE 5

To control for effects of word length on the articulatory cost, the average articulatory costs of words in a vocabulary was divided by its average word length. The normalized cost is shown here as a function of vocabulary size and word acquisition rate.
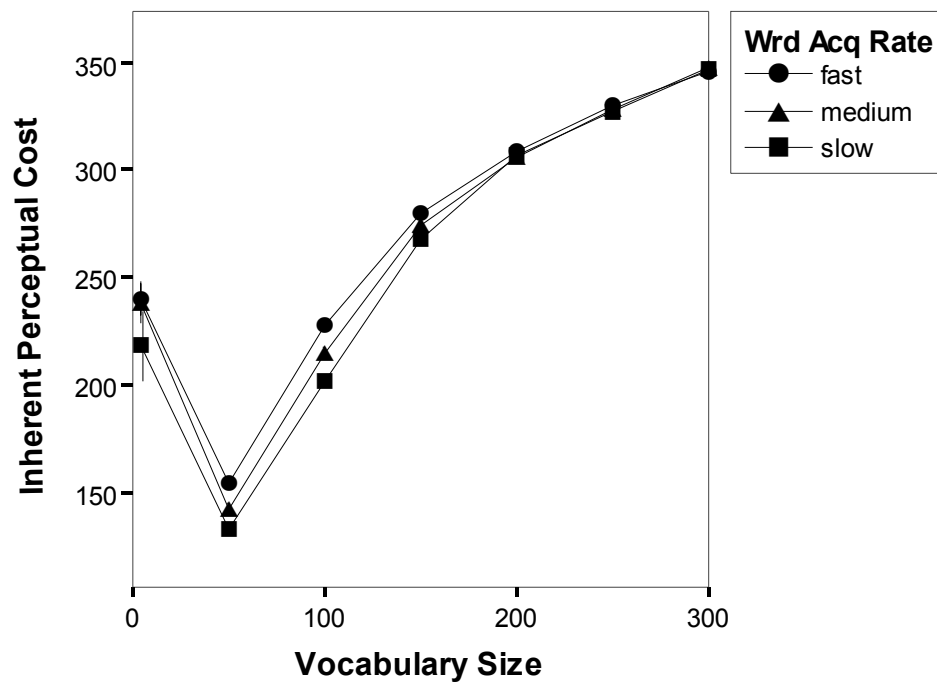
FIGURE 6

The average inherent perceptual cost of words in the simulated child vocabularies is shown as a function of vocabulary size and word acquisition rate. Inherent perceptual costs penalizes syllable structure complexity in a word. Accordingly, this cost increases as syllable structure becomes more complex in the vocabulary.
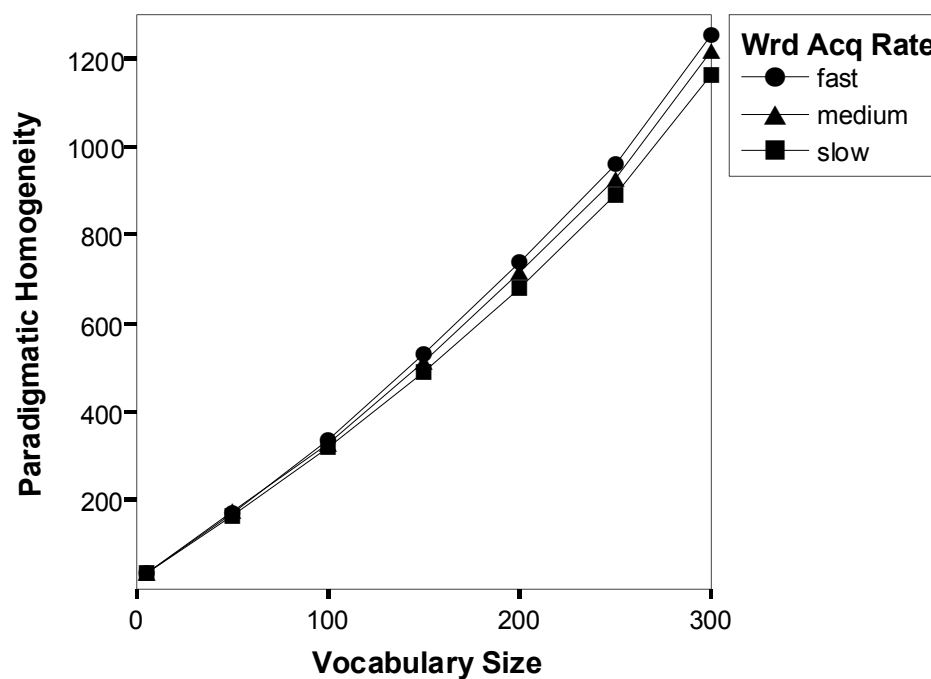
FIGURE 7

The average paradigmatic homogeneity cost of words in the simulated child vocabularies

is shown as a function of vocabulary size and word acquisition rate. Paradigmatic

homogeneity / diversity is calculated as the number of identical onset, offset, and nucleus

structures with identical sounds across the words in a given vocabulary. The cost

increases when the same structures with the same sounds are used with increasing
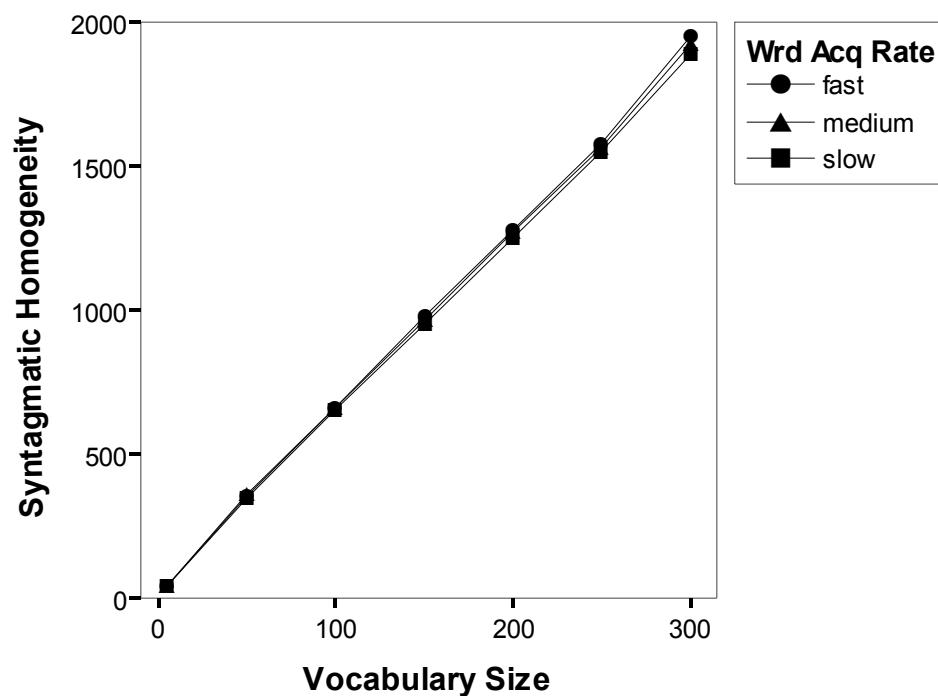
number of words.

FIGURE 8

The average syntagmatic homogeneity cost of words in the simulated child vocabularies is shown as a function of vocabulary size and word acquisition rate. Syntagmatic homogeneity is calculated by comparing each word with every other word in a vocabulary. All position-by-position similarities between words are summed to reflect the homogeneity / diversity of sound sequences in the vocabulary. The cost increases as similar subsequences of segments are used in the same word positions in increasing number of words.