Normal Vowel Development

Patricia Donegan
University of Hawai‘i at Mānoa

Introduction

The acquisition of vowels by normally developing children is an area that has received increasing attention in the past decade, but the substitutions that characterize children’s speech remain largely undocumented. Looking at adult phonological substitutions, we find that the range of normal substitutions is very wide, but clear. There is every reason to believe that we should find at least that same range in normal children. But existing descriptions of children’s speech show only a small sample of the systems and substitutions that could occur, and little effort has been made to determine whether children’s limitations and substitutions are coextensive with those of adult languages.

In the discussion below, I will briefly review some of the literature on vowel acquisition, and I will consider the development of phonological systems. I will then describe known adult vowel quality substitutions in terms of a universal set of articulatorily and perceptually motivated phonological processes that govern speech production. Although these processes are universal, their effects may be limited by learning, subject to implicational conditions which reflect the phonetic motivations of the processes. In the model of phonological acquisition I will present, these processes constrain vowel inventories in adult and child language, and they account for the vowel substitutions that occur in language histories, in variation and alternation, in second-language phonology, and in normal and disordered speech development. The implicational conditions on process application are responsible for the typical shapes of vowel inventories in child and adult speech.

---

1 A phonological substitution here includes any change of a single phonetic feature; devoicing of [d] to [t], or assimilation of [np] to [mp] are substitutions, as is lowering of [ʌ] to [a]. Some substitutes (outputs) may be the result of multiple substitutions. If /l/ becomes [d], it is delateralized as well as ‘stopped’. If /ai/ is pronounced as [æ], the /a/ assimilates the palatality and tenseness of [i], and the [i] assimilates the lowness of /a/.

2 A consonant example must serve here: Fey and Gandour (1982) described a process of postnasalization of word-final voiced stops in the speech of a 21-month-old English-speaking boy. They claimed that this is an idiosyncratic, ‘invented’ rule rather than a ‘natural’ one, because they knew of no phonological system that derives a phonetic sequence of stop plus nasal from an underlying voiced stop (80). But postnasalization of voiced stops, while somewhat uncommon, does occur in such languages as Mundari (Osada 1992) and Temiar (Benjamin 1976) as a regular allophonic process.
Because vowel errors occur primarily at ages when children’s natural speech may be difficult for outside observers to elicit, many phonological studies of vowel development focus on errors in the speech of children with delays (e.g. Bernhardt and Stemberger 1998). The vowel substitutions described in studies of normally developing children and those in children with delayed development appear to be quite parallel (Stoel-Gammon and Dunn 1985, Pollock and Keiser 1990, Stoel-Gammon and Pollock 2008). Scarcity of studies on normal children’s substitutions might leave this parallel open to question, but children’s errors can usually be attributed to processes (or constraints) like those of adult languages. This is not surprising if there is a common phonetic basis for both sets of phenomena, and one would expect normal and delayed development to follow similar patterns.

Optimality Theory expresses the content of phonological processes as universal well-formedness constraints or families of constraints, and uses faithfulness constraints and constraint ranking rather than limitation or suppression to differentiate the phonologies of individual languages (Prince and Smolensky 1993/2002). The substance of a universal set of phonetic processes and that of a universal set of phonetic constraints may prove to be a matter of ‘translation’ (see, for example, Lassettre 1995). The choice of expressing phonetic limitations on speech as processes vs. constraints does not directly affect the discussion here, so I refer to phonological processes. Should constraints and their ranking prove adequate to describe the regularities of phonological production and perception, conversion would be possible.

Literature on Vowel Development
The literature on vowel development suggests that the vowels of a language are acquired early, both in production and perception. Researchers (e.g. Locke 1983, Chen and Kent 2010) no longer accept Jakobson’s (1941/1968) claim of a sharp discontinuity between babbling and first words. MacNeilage & Davis (1990) reviewed studies of vowels in babbling in twenty-three infants; they found a tendency for mid and low and front and central vowels to be preferred across languages. This apparent preference for ‘lower left quadrant’ vowels also appeared in the analysis of a large database produced by six English-learning infants (Davis & MacNeilage, 1995).

The ambient language influences vowel qualities very early. Infants from English, French, Algerian and Cantonese environments begin to differ in vowel quality as early as 0;10 (de Boysson-Bardies et al. 1989), though a study of Mandarin-learning infants suggests that this happens as early as 0;7 (Chen and Kent 2010). Lee et al. (2010) compared vowel patterns of Korean and English-learning infants with the patterns of their ambient languages, and they emphasize the particular influence of infant-directed speech. They suggest that vowels may manifest ambient language patterns in infant production earlier than consonants because of their higher amplitude and greater duration.
Production
Among the studies of children’s word productions, influential cross-sectional studies include Templin’s (1957) study of 480 children from 3 to 8 years old, which established normative ages for the acquisition of consonant production. But the mean percentage of correct vowel productions was already 93.3% for the youngest age group, and there was no significant increase in production accuracy over the age range studied. This result – and those of subsequent studies – show that most vowel quality development takes place before age 3.

Irwin and Wong (1983) coordinated and analyzed a set of studies of phonological development in children aged 1;6 to 6;0. The studies investigated both vowel and consonant development in twenty subjects. The studies of children at 18 months (Paschall 1983) and 2 years (Hare 1983) may indicate where vowel difficulties lie, while the studies of 3- and 4-year-olds (Larkins 1983, Bassi 1983) serve mainly to establish the rarity of errors at age 3 and beyond. Such cross-sectional studies may be useful in providing criteria to identify vowel development delays, but they reveal little about the pronunciation systems of individual children.

Detailed longitudinal studies that do describe individual children and segmental quality may be narrowly focused, but they include crucial information that reveals how children’s systems develop. The early studies of Velten (1943) and Leopold (1939-1947), followed by those of Smith (1973), Menn (1978), and Major (1977) describe in detail the phonological development of English-learning children. Smith (1973, 37-40) describes the acquisition of English Standard Pronunciation by his son Amahl, but he says little about A’s acquisition of vowels. At the start of the study, (2 ;2) A had acquired adult-like pronunciations of most English vowels, except for /æ, œ, oj/ (man, soap, noisy), and some vowels affected by following /r/. By the time A was three years old, he had acquired adult-like vowels, even those affected by a following /r/.

Major’s 1977 study describes the phonological development, beginning at 1;7, of his daughter Sylvia, who was learning both English and Brazilian Portuguese. Major gave special attention to vowel substitutions, and because Sylvia was becoming bilingual, he emphasized the progressive differentiation of her two developing phonologies. Menn (1978) offers a study of the earliest stages of an American boy’s phonology. Jacob (1;2 – 1;9) shows less regularity in vowel and consonant substitutions than the above-mentioned children, but he was considerably younger than they were at the time of the respective studies. Examination of Menn’s data on Jacob at 1;8 and beyond indicates that the vowels of stressed syllables (when not influenced by adjacent liquids or glides) were beginning to stabilize but were not regular.

Lohuis-Weber and Zonneveld (1996) describe the prosodic development of a Dutch learner, a boy, from age 1;8 to 2;11. Unstressed syllables are deleted at first, but the vowels of stressed syllables are usually correct, including the
Dutch labiopalatal vowels. Macken (1979), describing the acquisition of Mexican Spanish by Si, a girl 1;7 to 2;1, focuses on consonant acquisition; her examples indicate mainly correct vowel qualities in accented syllables, even in Si’s earliest words. Variation between [ʌ] and [ɑ] or between [i] and [ɪ] does occur.

A larger study by Otomo and Stoel-Gammon (1992) describes the acquisition of American English /i, ɪ, e, æ, a/ by six normally developing children at 1;10, 2;2, and 2;6. /i/ and /a/ were mastered early; /ɪ/ and /e/ were least accurate throughout the study. Variability decreased as the subjects matured. Some context-sensitive vowel substitution patterns were observed. Bleile 1989, Vihman 1996, and Bernhardt and Stemberger 1998 also illustrate a wide variety of segmental substitutions in a variety of children with both normal and delayed speech.

There are an increasing number of instrumental studies, beginning with Buhr 1980, Lieberman 1980, and Bond et al. 1982. Buhr’s study, like that of Davis and MacNeilage (1995), indicated that vowel production during the first year favors low, non-front, non-rounded vowels. In word production, Buhr observed that the front-back vowel dimension appears later than the height axis, since the latter can apparently be achieved by jaw opening only, without much coordination with tongue or lip movement.

In line with the increased availability of high quality recordings of children’s speech and newer technologies like MRI, Vorperian and Kent (2007) advocate increased use of acoustic data and its integration with anatomical, physiological, and perceptual data to account for speech development. They draw together data on the development of the vocal tract with published acoustic data on the refinements of vowel production (lowering of F0 and of F1 and F2 values, reduction of variability, emergence of sex differences, etc.) which continue from early acquisition until the teen years.

However, in many studies, identification of vowels by the researcher has been auditory, and in some studies, the listeners who identified the vowels seemed to be using information other than vowel formant heights for identification, since some vowels that were distinguished by ear remained indistinguishable on the formant charts. One must conclude that, despite the difficulties involved in learning to transcribe vowels (especially children’s vowels), the auditory identification of vowels is an irreplaceable tool in the analysis of vowel substitutions and development.

The acquisition of vowel length has been studied, as well. In languages where the vowel length contrast is robustly durational, with little associated quality difference, children distinguish long and short vowels by duration very early – by 1;6 for Japanese learners (Ota 1999) or by 2;0 for Swedish (Stoel-Gammon, Buder & Kehoe, 1995; Stoel-Gammon & Buder, 1998). Further, Stoel-Gammon & Buder (1998) reported that English learners who did not distinguish a tense-lax vowel pair by duration were already making vowels before target
voiced stops considerably longer than vowels before target voiceless stops; for Swedish children such contextual lengthening was minimal.

Some researchers associate acquisition of length with the acquisition of syllable, specifically rhyme, structure (Fikkert 1994, Salidis and Johnson 1997, Kehoe & Stoel-Gammon 2001, and Kehoe and Lleo 2003). These researchers studied children learning Dutch, English, and German – all languages in which “vowel length” is in part a matter of quality – so the question of “length” is an abstract notion of quantity – not just duration. Kehoe & Stoel-Gammon (2001) coded the English vowels /i, e, ə, o, u, ɔ/ as “long”, and /ɪ, ɛ, æ, ʌ, ʊ/ as “short”, regardless of duration, citing Ladefoged’s (1993) designation of the former set as “tense” and the latter as “lax”. But Ladefoged’s criteria for this classification were entirely distributional, not phonetic, and it is hard to establish that the children have actually acquired the distinction, which mainly affects stress assignment in the adult language, or to see how, without phonetic criteria and with limited phonotactic or distributional knowledge, children create the sets.

Initial Variability and Early Accuracy in Production

Stoel-Gammon and Dunn (1985) and Vihman (1996) show many examples where vowels are at first highly variable. In a given word, an accurate production may vary with a number of substitutes. The initial variability of vowels is not surprising when we consider the difficulty of controlling the shape of the tongue, which is a complex three-dimensional network of intrinsic longitudinal, vertical, and transverse fibers (Kent 1992). Initial variability may quickly resolve into consistent substitutions for consonants, since consistency allows caregivers to interpret the child’s speech. But by the time children settle on consistent substitutions in consonants, most of the vowels in their speech are reasonably accurate. The high percentages of accuracy at ages 2 years and above reported in Irwin and Wong (1983) confirms that many children achieve quite acceptable vowel quality before age 3 in all but the rhotic vowels.

This may be because vowels can be produced satisfactorily with control of fewer features than consonants require. (See Figure 1.) Even in languages with very rich vowel systems, the principal features that determine vowel quality are tongue/jaw height, palatality or tongue advancement, labiality, and tenseness (degree of palatality or labiality). Backness is rarely distinguished from non-palatality. In some languages, features like nasalization, tongue-root advancement or retraction (pharyngealization), rhoticity, and voice quality may also play a role. But many complex vowel systems, from the Germanic languages to those of Southeast Asia (e.g. Mon-Khmer), result from the presence of diphthongs, which require sequential articulations but are composed of the same basic features as monophthongs. In contrast, even an ordinary-sized consonant
inventory may require a comparative multiplicity of contrasting features: various articulators (lips, tongue tip, tongue dorsum, etc.), places for constriction (bilabial, labio-dental, dental, alveolar, etc.), and degrees of closure (stop, fricative, approximant) are required, and special manner features (like laterality, nasality, sibilance, affrication, trill, etc.), and laryngeal features (voicing, aspiration, glottalization, etc.) are often required.

Perceptual Development
Beginning with the work of Eimas et al. (1971), it has been shown that very young infants (under 6 months of age) are able to perceive most of the phonetic differences that are used to distinguish language sounds (Aslin et al. 1981, Best, McRoberts and Sithole 1988, Eimas 1975, Jusczyk and Thompson 1978, Trehub 1976, Kuhl and Miller 1982, Kuhl 1983, and others). (Eilers, Wilson, and Moore 1979 and Eilers, Gavin, and Oller 1982 offer some contradictory findings.) This does not mean that all distinctions are equally easy to perceive, but it does mean that infants are unbiased with respect to the categories of the language they will learn. In contrast, the studies of Janet Werker and her colleagues (Werker et al. 1981, Werker and Tees 1984a, b, Werker and Lalonde 1988) offered considerable evidence that infants’ perceptual categories for consonants, by 12;0, have begun to correspond to the linguistically significant categories for the ambient language.

Perception of language-specific categories involves a narrowing of the range of distinctions perceived to something approaching phonemic perception. Although the phoneme, as originally conceived, is not a distinctive element, but rather, a perceptual and representational one, Werker and Pegg (1992) used the term ‘language-specific phonetic perception’, because they believe ‘phonemic perception’ would require the ability to distinguish lexical items – which is difficult to establish at this early age, but they point to various kinds of ‘strong evidence that infants are influenced by native-language phonological categories before 10 months of age’ (298).

These observations, however, are based largely on consonant acquisition. Vowel perception might be less categorical, because of the continuously varying nature of vowel quality, as opposed to the seemingly discrete differences that characterize consonants, but there is little to support this supposition. Instead, the early accuracy of vowel production seems to go hand in hand with early categorization in perception. Polka and Werker (1994) found that English-learning infants younger than 6;0 could discriminate two German vowel contrasts that their 6-month-old counterparts could not distinguish. This indicates that the shift from language-general toward language-particular phonological perception takes place even earlier for vowels than for consonants.

Lieberman (1980) had posited “an innate, species-specific neural mechanism”, which allows perceptual normalization of vowels in terms of the speaker’s presumed supralaryngeal vocal-tract length, and which allows an
infant to know when it has produced the perceptual equivalent of an adult vowel (137). Kuhl (1980, 1983, Kuhl and Miller 1982) provides evidence for ‘perceptual constancy’ – the ability of infants to generalize responses to speech sounds produced on different pitches and by different speakers – and thus, perhaps, for such a normalization system. Kuhl further suggests (1991) that ‘perceptual magnet effects’ show that speech sound categories are structured around prototypes: adults rated vowels as good or poor exemplars of a phonetic category, like /i/, and then adults and infants were tested on discrimination of these exemplars. If a prototypical exemplar was presented, both adult and 6-month-old infant subjects more readily accepted a poor exemplar as the same (the prototype acted as a perceptual magnet), but a poor exemplar as the referent generalized less readily.

We might conclude that if the child has formed language-specific categories for perception, these categories would be used for the long-term mental representation of vowels when the child begins to imitate and use words. The function of phonemic categorization in adult speech seems to be representational as well as perceptual, and the categories may serve both functions for children as well.

**Representation**

Evidence that children are able to categorize speech sounds accurately does not allow us to conclude that children’s lexical representations are always entirely accurate. The claim that children’s errors, like adult variation, involve substitutions requires the implicit assumption that the child’s mental representations of words are relatively accurate or adult-like. Stampe (1969) claimed that even the child’s earliest phonological representations were in large part like the phonemic representations of adult speakers. Smith (1973) drew a similar conclusion. Stampe and Smith both concluded that children’s phonological representations are generally accurate, for three reasons. First, the children were able to perceive differences they did not produce. Second, upon acquisition of a new pronunciation ability, they often corrected words that had undergone a substitution without re-hearing them. And third, the children made substitutions that could only be explained with adult-like phonological representations.

Menn adduced another kind of evidence for early perceptual and representational accuracy, often evident in the child’s earliest productions: avoidance. Menn noted that avoidance of a sound or class of sounds ‘implies the ability to discriminate it from sounds one does attempt to say’ (1978, 71), and since avoided words are often in the child’s receptive vocabulary, they must have representations that include the avoided feature. In addition, instrumental phonetic studies of children’s productions (e.g. Kornfeld 1971, Macken and Barton 1980, Carter and Gerken 2004) indicate that children may produce
systematic differences, often imperceptible to adults, between forms that are different in the adult language.

The occurrence of accurate children’s perceptual categories with the phonemic categories of the ambient language at the onset of speech argues for adult-like representations, but it means that constraints on possible lexical representations must also exist for children. Stampe’s proposal that children’s representations are like adult phonemic representations assumed that allophonic details are not part of these representations. There is an advantage to conflating sound categories, in that this makes for fewer intendable targets to learn to pronounce. The limitations of phonemic perception exist in both adult and child phonology, but in the child, the limitations on production exceed those on perception.

Phonotactic limitations on possible lexical representations might also affect perception, but if, for example, a constraint forbidding final consonants applied in a child’s perception, and the child thereby failed to perceive and remember final consonants (cf. Dinnsen 1984), we would expect the child to be unable to distinguish perceptually among sets like *tap* /tæk/ /tæb/ /tæg/, or *cat* /kæp/ /kæʃ/ /kælf/.) Instead, perception and memory of a distinction can usually be established well in advance of the child’s ability to produce the contrast.

Part of the difficulty in deciding where the child’s errors lie is that both perceptual and production limitations may be characterized with the same set of phonological processes or constraints. A process (or constraint) that specifies that all non-low vowels must be high, or that all back vowels must be labial may affect perception or production or both. A child learning a language with mid vowels, who produces only high vowels, may perceive and remember the difference but be unable to achieve consistent non-high, non-low articulation – the evidence from perception favors this view. For an adult speaker of a language with no mid vowels, the limitation may come to affect both perception and production. If this adult attempts to learn a language with mid vowels, the constraint is ordinarily overcome in perception before it is mastered in production.³

Representations, particularly of longer or more complicated words, may be incomplete or inaccurate, however. Macken (1980) offered evidence that a child’s own incorrect pronunciations may influence his mental representations of individual words. Aitcheson and Chiat (1981) proposed that many developmental phonological errors were the results of difficulties in lexical storage and retrieval, rather than results of production difficulties. However, aiming to provide a control group for the Aitcheson & Chiat study, Smith et al. (1991) studied new-word learning in adults. They found that adults encountered

³ Articulatory instruction can reverse this order in adults—e.g. a speaker’s ability to produce the distinction of dental vs. retroflex stops may exceed her ability to distinguish these in perception. A similar asymmetry with English /r/ vs /l/ may be observed for some Japanese speakers who produce the difference well but do not hear it very reliably.
difficulties with storage and recall that were similar to those of children, but that the error patterns of adults and older children were unlike the ‘reasonably consistent, unidirectional substitution patterns common in phonological processes among different children and even across languages’ (296).

The Development of Phonological Systems
In considering how a child learns to perceive and pronounce vowels, the conservative assumption would be that development from newborn state to adult phonology involves no gross discontinuities. Available data support this, suggesting that paths of normal vowel development are based on the same features and processes that appear in vowel phonologies in adult languages (cf. Stampe 1973).

A note of explanation is required regarding my use of the terms ‘phonological system’ and ‘phonology’. In this discussion, phonology simply means the mental system which underlies the perception and production of speech (cf. Donegan and Stampe 1978). The phonological system of a language is responsible for the perceptual and articulatory limitations and abilities of the adult native speaker – and for his or her ‘accent’. Phonology, so viewed, arises in phonetic realities and is based on phonetic causalities; these are universal – but their language-specific implementation must of course be learned. True phonological regularities can be stated in phonetic and prosodic terms; they do not refer to morphological information, as the rules of morphophonology (or ‘lexical phonology’) do.

Some authors would regard the system I am describing as phonetics, not phonology. But this ‘natural’ phonological system involves categorization of sounds and selection of substitutions that are unquestionably language-specific; they cannot, therefore, be necessary results of human perceptual and articulatory abilities – only options, which can be controlled. They involve categorical conditions as well as gradients which are subject to variable cut-off points, and they are subject to learning, in that control over the options is learned.

Much of modern generative and post-generative phonology is concerned with relationships among the forms of lexical items (‘lexical phonology’), and with alternations that are conditioned by morphological classes or boundaries. But such alternations are conventionalized; they no longer retain their historic phonetic motivation and they are easy for speakers to violate. These lexical relationships are very different from the problem of vowel (and consonant) acquisition, and may be based on quite different principles (Stampe 1987).

---

4 Myers (1997, 125), for example, clearly states the position that only alternations which are obligatory, categorical (phonemic), independent of rate, and restricted to particular morphological categories are phonological.

5 English speakers, for example, can easily pronounce electricity as [ɪlk'trɪkɪɾɪ] or soften as [ˈsɒftən], even though these pronunciations violate lexical rules of English.
Features and Segments

With a phonetics-based phonology, we can reconsider some basic ideas about features and segments, as well as phonological processes, in child and adult speech. The evidence about perception noted above suggests that children’s perception of speech at the onset of word production is relatively accurate and language specific. The next question, then, is how these perceptions are represented in the child’s memory: are the child’s initial phonological representations whole words, or are they composed of features, or segments, or some other units?

When phonology is based on phonetics, its elements – the phonological features – must be phonetic. Phonological features can be viewed, not as abstract categories, but as the links of motor and proprioceptive aspects of production, on the one hand, to perceptual properties (auditory, acoustic, or sometimes visual) on the other. Such connections may be part of an inborn, ‘pre-wired’ mechanism, or these connections may develop or be discovered by the child during the first year, in vocalization, crying, oral play, noisy eating or drinking, and babbling. During these activities, the child creates a map of the articulatory apparatus, noting the auditory effects of different laryngeal adjustments, of constrictions at different points of articulation, of different degrees and durations of closure, different cavity shapes, different velar positions, etc. Whether ‘innate’ or developed through experience, these connections are part of the child’s natural language endowment.

So children acquire phonological features by mapping articulatory gestures (like oral closure, or tongue advancement, or lip rounding, or jaw lowering) or components of gestures (like places of articulation, or degrees of closure) to their auditory effects. Thus, for example, the auditory effects of phonation are associated with particular laryngeal and respiratory gestures, and different pitches with different gestures. An oral closure and release during phonation are connected with a sudden drop and subsequent rise in intensity, and the place of closure is connected with a particular auditory pattern of formant transitions and the pitch range of the release burst. Similarly, tongue advancement during phonation is associated with raising of the second vowel formant, tongue and jaw raising with the lowering of the first formant, etc. This mapping of the vocal apparatus to its acoustic-auditory capabilities may begin with the infant’s earliest vocalizations, and it is surely an important function of babbling. The resulting connections of gestural elements with their auditory effects are the foundation of the child’s (and the adult’s) ability to imitate, and they function as features in the child’s (and the adult’s) phonology.6

6 Meltzoff and Moore (1977) documented imitation of adult facial gestures by young infants (who have never, of course, observed their own facial expressions). And Nishitani et al. (2005), considering brain activity that suggests a “mirror neuron system” in humans that links perception and action, propose that Broca’s area has a role in relating perception with imitation, noting that “Humans most likely understand another person’s actions, and also their motor-act-
Early function need not imply simplicity. If a feature is a gestural component linked to an auditory effect, that does not require that both the gesture and the effect are simple. A feature may involve a rather complex oral gesture connected with a simple, easily identifiable acoustic effect (e.g. tongue advancement and F2 raising), or a fairly simple gesture linked to a complex auditory effect (e.g. lip rounding and its associated complex of formant lowering). And a feature may be present in varying degrees, depending on other features with which it simultaneously co-occurs: for example, lip rounding may be present in high or low vowels, but (other things being equal) it is necessarily weaker in low vowels, where it conflicts with jaw opening.

Comparison of phonological processing in many languages is necessary in order to discover of all the ‘natural classes’ that are referred to in phonological substitutions (sets of sounds that undergo the same substitution, sets of sounds that condition a substitution, sets of sounds substituted, etc.). By this method, we can identify a universal set of features that are relevant in substitutions, and thus, presumably, in perception and production. The set will undoubtedly be larger than the sets of ‘distinctive features’ most phonological theories currently subscribe to; a basic set for vowels will be used here.

**Vowel Features**

Vowels consist of simultaneous features which often conflict with each other. The principal vowel features (see Fig. 1 above and discussion in Donegan 1985) are Palatality (frontness), Labiality (usually manifested as roundness) and Sonority or intrinsic intensity, which is correlated inversely with Vowel Height ([+low] vowels are most sonorous, [+high] vowels least). Palatality, labiality and the height features are both binary and gradient. They are binary, in the sense that they are either present or absent in a vowel: for example, [i] is palatal and [i] is non-palatal, [o] is labial and [ɔ] is non-labial, [i] is high and [ɛ] is non-high. Either the presence of these features or their absence can be referred to in phonological processing.  

But if present, each feature may be present to a greater or lesser degree, depending on the other features with which it simultaneously combines. The phonetic aspects of this claim would seem to be obvious, and the conditions on the application of phonological processes based intentions, by mapping observed actions, postures, and gaze onto their own motor representations of similar actions.”

7 There are claims that features are privative (so that phonological processes cannot refer to the negative value), but that question cannot be addressed here.

8 The more palatal a vowel is, the less open (and thus the less sonorant) it can be. Tongue-fronting and raising decrease the size of the forward resonating cavity, raising F2, and thus give the vowel its palatal quality; lowering the tongue or jaw attenuates these fronting gestures and their effects. And the fronting or raising gestures associated with palatality attenuate sonority, in that they decrease the size of the oral cavity and enlarge the pharyngeal (or back) cavity, lowering F1 and decreasing the vowel’s intrinsic intensity. Labiality and sonority conflict in a similar way, both articulatorily and acoustically: the more open the jaw, the less rounding the
confirm it. These features and their varying strengths are manifested in the susceptibility of vowels to different phonological substitutions. As shown below, features that are weakly present are most susceptible to loss, and those that are strongly present are most subject to enhancement or optimization.

**Representational Units: Words, Syllables, Segments, Features**

Having proposed an interpretation of what features are, and a basic set of vowel features, it seems important to consider why features are, from the start, part of the child’s representations and processing.

It is sometimes claimed (Menyuk, Menn, and Silber 1986, Ferguson & Farwell 1975) that children represent their first words as unanalyzed wholes. Words may be acquired first as ‘passive’ vocabulary, which the child recognizes but does not produce. At this point, they may indeed be recognized as unanalyzed wholes, or as general gestalts – as we assume a dog recognizes *Sit!*, or a horse recognizes *Whoat*. But when a child begins to attempt to produce words, the words must undergo analysis into features (in the sense used here), because an attempt at imitation requires this:

a.) The imitator must identify a set of simultaneous auditory properties: e.g., for a single segment, like [m], this would include the combined auditory effects of voicing, nasality, oral closure, and labiality. More likely (to produce a word an adult will recognize), the child identifies a series of such sets of simultaneous properties; so for a series of segments, like [mama], the above features are heard to alternate with the effects of oral openness, jaw lowering, non-labiality, etc.

b.) The imitator must connect those auditory/ acoustic effects with articulatory gestures. This can be done only if the imitator has some knowledge of which effects are the results of which gestures. So auditory-articulatory links – or features – must be available. The success of the attempt to imitate depends in part on established links between sound and gesture; these may be incomplete.

c.) The imitator must attempt to produce an appropriately timed sequence of gestures. The difficulties of this part of the imitation process are generally recognized.

Lindblom (1992) suggests that first words, usually CV syllables, are stored as ‘trajectories’, which implies storage as whole syllables, rather than segments.

---

Lips can achieve; conversely, lip constriction and protrusion lower F1 and decrease intrinsic intensity. So the optimal palatal vowel is [i], the optimal labial vowel is [u], and the optimal sonorant is non-palatal non-labial [a]. Palatality and labiality also conflict, attenuating each other’s effects: palatality raises F2 and labiality lowers it, so ‘pure’ labials like [u] and ‘pure’ palatals like [i] are favored among the world’s languages over labiopalatals like [y]. Palatality or labiality (grouped terminologically as ‘vowel colors’) can make height differences more audible, so ‘achromatic’ vowel systems like /i, a, a/ are rare among the world’s languages, and ‘triangular’ systems like /i, a, u/ or /i, e, a, o, u/ are common. Tenseness may be regarded as intensity of color (palatality or labiality) for a given degree of height, so that [i] is more palatal and less sonorant than [i], and [e] is more palatal and less sonorant than [e]. The non-colored vowels [i, a, a] cannot be tense on this definition; see Donegan 1985 for detailed arguments.
But the essential points of a trajectory are the beginning and ending points. For a word like [mā] (or reduplicated [māmā]) the articulatory aspects that must be represented are a simultaneous combination of voicing, nasality, oral closure, and labial place of closure, and a simultaneous combination of oral openness, jaw lowering, and non-rounding. The point at which these simultaneous configurations are identified as units, or ‘segments’, may be difficult to establish. But the child’s ‘analysis’ of speech must proceed far enough that these simultaneous configurations of features are part of his or her mental representations when imitation begins – and this would include the self-imitation that occurs in babbling (Locke and Pearson 1992) and the unsuccessful imitations that may occur before adults in the family recognize a ‘first word’. Feature analysis is not, then, based on distribution or contrast; it is part of figuring out (or knowing) how to imitate.

**Vowel Processes**

Phonological processes are not limited to the speech of children. They are responsible for constraints on adult phonological representations, and they underlie phonological substitutions (alternation and variation) in adult speech, and phonological change in languages. Phonological processes express phonetic desiderata and eliminate phonetic difficulties. Some processes resolve difficulties in the production or perception of segments, optimizing simultaneous feature combinations; others resolve difficulties that result from the sequencing of segments.

If children and adults have phonetic capacities that are essentially the same, then we might expect some children to happen upon any of the substitutions we observe in alternation, variation, or sound changes in the languages of the world. Or perhaps children only make some subset of these changes; this remains to be seen.9

It is important to note that although processes are responsible for regularities in children’s speech, the regularities of children’s substitutions are by no means perfect. Much has been made of the ‘messiness’ of child data, and children’s data may perhaps be more variable and idiosyncratic than adult data. But far too little has been noted of the comparable variability and apparent irregularity of adult data. It would be a mistake to ignore variation and

9 For example, tones appear to arise in languages as a way of maintaining or amplifying a consonant voicing distinction. In tonogenesis, the normal effects of voicing on vowels (lowered fundamental frequency after voiced obstruents and/or raised fundamental after voiceless obstruents) are amplified and used to enhance the consonantal distinction; lowered or raised pitch then replaces the consonant distinction, which is lost. I know of no reports of children who have used tonal distinctions to express neutralized consonant voicing distinctions. But if a learner of a non-tonal language used such a strategy, would the non-tonally-oriented transcriber notice it and mark these incipient tones?
irregularity wherever we find it, but it is the regularities create the phonological system and provide the keys to its explanation.

Not only do phonological processes underlie substitutions in adult and child speech production; they also govern perception and thereby constrain the phonological inventory. Fortitive processes limit the inventory by ruling out certain simultaneous combinations of features (like nasalized vowels), while lenitive processes allow speakers to discount certain actually-occurring proscribed segments by attributing them to the application of sequence-optimizing substitutions (like context-sensitive vowel nasalization). Stampe 1987 elucidates how processes constrain phonological representations in adults.

But if processes constrain perception, then why are children at first able to perceive virtually all occurring phonetic distinctions and why do these abilities become more language-specific during the first year? One important difference between the youngest infants, who can discriminate non-native distinctions, and the one-year-olds, who cannot, is the degree of experience of vocalization and babbling. It seems that without this experience, the child's perceptions are unconstrained by any limitations on his or her production abilities. As these limitations are discovered, the child finds that certain feature combinations are less difficult or more perceptually satisfactory than others. The child thus discovers reasons to limit perceptual categories to the more-optimal feature configurations, insofar as this turns out to be possible in the target language. (See Donegan 1995 for further discussion.)

In the sections below, I will present an overview of vowel processes as they apply in the speech of children and in various languages of the world. I will show that processes account for a wide variety of substitutions and limitations, but that they are limited by implicational constraints on their application. These implicational constraints manifest the phonetic motivations of the processes.

**Context-sensitive substitutions: Vowel-to-Consonant Assimilation**

Davis & MacNeilage (1990, 1995) and MacNeilage (1996) proposed that basic mandibular oscillation patterns with a relatively passive tongue result in early CV syllables in babbling. Chen & Kent (2005), in a study of Mandarin-learning infants, confirmed the predicted patterns of labials with back vowels, alveolars with front vowels, and velars with back vowels. In very early speech, vowel quality may influence adjacent consonants, with high front vowels exerting the greatest influence, and labial vowels the least (Gierut et al. 1993, Wolfe and Blocker 1990). Children may make labial and sometimes velar consonants alveolar before front vowels, and some make consonants velar before back vowels. Conversely, consonant quality may sometimes influence vowels: e.g. So and Dodd report vowel errors only in the youngest group of Cantonese speakers they studied and remark that errors were few, but the only consistent pattern they noted was that of vowel-to-consonant assimilation. Assimilation of vowels to adjacent consonants is less common in child speech than one might expect.
In this regard, acquisition of languages with series of palatalized and labialized consonants should receive attention. Labialization of vowels adjacent to labial consonants (or labialized consonants, like English [w, r, j]), and palatalization of vowels adjacent to palatal or coronal consonants may occur (compare [dædæ] with [mɑmɑ] in many children).

Depalatalization of vowels adjacent to non-palatal sonorants is reported, even when the sonorant is replaced or deleted: e.g. bread, dress as [bwʌd], [dʌs] (Otomo and Stoel-Gammon 1992). In such substitutions, it is important to note the quality of the liquids in the child’s speech. If a child backs vowels after /r/, it is important to know if /r/ is pronounced elsewhere as [w], [ɬ], or [j], or as some other alternative, in order to decide whether this is assimilation or dissimilation.

Assimilation may affect the entire vowel, or it may appear as diphthongization, where the beginning or the end of the vowel is assimilated to an adjacent consonant, creating an on-glide or an off-glide. Long vowels, or vowels in lengthening environments (open syllables, monosyllabic words, stressed syllables, etc.) are particularly susceptible to diphthongization. For example, in English words like mash, sash, a palatal glide may develop before the [ʃ], giving [mæʃ], [sæʃ], as in various U.S. dialects. Similar assimilative diphthongization in children’s speech is not surprising.

Possible contexts for such assimilative diphthongizations should be considered carefully, since they may be subtle and sometimes surprising. One might not expect a ‘front’ glide before a following velar consonant, but velar glides are often palatalized after front vowels, and development of a palatal glide before a following palatalized velar is not surprising – in fact, pronunciations like [bæj], [leij], [kiʃ] for bag, leg, and king are common among American adults, as well as in the speech of children. (One does not ordinarily find this diphthongization before [k] in adult speech because a following voiceless stop shortens the vowel, preventing diphthongization.) One also might not expect a palatal glide before [s], but this diphthongization is assimilative (since both [s] and palatals are coronal, and vowel assimilation favors continuants over stops like [t, d, n]).

It is fairly common among Southeast Asian languages for vowels to acquire palatal offglides before palatal consonants and [s]; it happens in Temiar (Benjamin 1976, 137ff) and Pacoh (Mark Alves, pers. com.) and in varieties of Mon (Huffman 1990, 42) and Wa (Diffloth 1980, 44) and before palatals in Khmer (personal observation). Consider also that in Italian, word-final Latin /s/ has become /i/: noi, voi, poi, sei from nos, vos, pos, sex. In view of such changes, the development of palatal glides between a front vowel and a velar, reported as common by Otomo and Stoel-Gammon (1992) is not surprising, and a child’s pronunciations of bus as [baʃs], bag as [baʃʃ] and bang as [baŋ] (Menn 1978).
need not arise from an idiosyncratic or invented ‘word recipe’, but may instead be a phonetically motivated assimilation.

Diphthongs that arise from assimilation are easily interpreted as ‘wrong’ vowels: mash [mæɪʃ] can be heard as /maɪʃ/; peg [pɛjɡ] is ordinarily heard as (British) /peɪɡ/ or (American) /pɛɡ/, and big [bɪɡ] is heard as /bɪɡ/. This last effect can make it impossible for a child to distinguish the vowels of seek vs. sick, or king vs. keen\(^{10}\). And of course, the [ɪi] (from /i/ plus offglide [i]) may monophthongize to [i], so it is not surprising that Otomo and Stoel-Gammon also found [i] → [i] before velars to be a common substitution.

Otomo and Stoel-Gammon report that vowels often become back before velar consonants, including velarized [h]. This illustrates the fact that we may find quite opposite solutions to the problem of producing a front-vowel-velar-consonant sequence: The velar may be fronted, with assimilation of part of the vowel to the fronted velar, as in the [bæɪɡ] example above, or the vowel may be backed to match the backness of the velar. Such an assimilation happened in early Latin, where front vowels became back before a velarized /l/: uelim, and familia with clear [l] remaining unchanged as an onset, or before a front vowel beside uoltis, famulus, where the /l/ was ‘dark’ in coda position or before a back vowel (Allen 1970, 34).

The Similarity Principle (noted first in Hutcheson 1974) holds that segments that are similar are implicationally favored for further assimilation: e.g. nt → nn implies nd → nn, but not the reverse. Therefore, we may find that palatal vowels assimilate to palatals and back vowels do not: thus [fiʃʃ] fish but not *[puiʃʃ] push. We may also find that vowels assimilate to approximants like /r, l/ in cases where they do not assimilate to obstruents or nasals, or that they assimilate to continuants like /s/ but not to stops like /t, d, n/, or that (as vocoids) they assimilate to vocoids but not to non-vocoids (Southern U.S. high [hæɡ] → [hæɡ], but gosh [ɡɒʃ] does not become *[ɡæʃ]).

**Context-free substitutions**

Context-free phonological processes are typically fortitive – that is, they apply to enhance (optimize or maximize) a particular phonetic feature of an individual segment (cf. Stevens and Keyser 1989). Since fortitions optimize individual segments, they may have perceptual as well as articulatory motivation, they often apply ‘context-freely’, and they may also apply dissimilatively. (Fortitions that affect consonants include devoicing of obstruents, aspiration of voiceless stops, pre- or post-nasalization of voiced stops, voicing of sonorants, and changes of glides into obstruents.) Lenitions, on the other hand, are concerned with sequences of segments; they have articulatory motivation only, they are reductive or assimilative, they decrease the number or magnitude of gestures, or

---

\(^{10}\) Many American adults claim that king/kin is not a minimal pair, but king/keen is.
ease timing requirements, and their application often obscures the features of
the individual segments.

Vowel fortitions include Lowering, Laxing, Depalatalization and Delabialization, which increase sonority, and Raising and Tensing, which increase palatality or labiality. Palatalization and Labialization also increase palatality and labiality, of course – and in conferring these properties they make weaker degrees of sonority (in mid or high vowels) more audible. These conflicting motivations and opposite changes may make it seem that ‘anything is possible’ – and indeed, all of these changes are found in children’s substitutions as well as in the world’s languages. But each process is strictly constrained by hierarchical conditions on the potential classes of input for a given substitution. For example, other things being equal, a lower vowel is more susceptible to depalatalization than a higher vowel of the same series, so if [e] depalatalizes, [æ] must also depalatalize, and if [i] depalatalizes, then [e] and [æ] must depalatalize as well. These implicational conditions can be summarized roughly with the principle, ‘the rich get richer and the poor get poorer’: a vowel that has a higher degree of some property is more susceptible to processes that increase that property, and a vowel with a low degree of a property is more susceptible to processes that weaken or remove that property (see Donegan 1985).

This means that a child who substitutes [ʌ] for /e/ will ordinarily also substitute non-palatal vowels for /e/ and for /æ/ – other things being equal. Since lax vowels are less palatal than their tense counterparts and lower vowels are less palatal than their higher counterparts, these are implicationally favored for depalatalization. And, other things being equal, we would expect a child who raises /e/ to [i] (increasing its palatality) also to raise /e/ to [i]. Of course, other things are not always equal – in English, for example, the vowel we might analyze as /e/ is a diphthong, [ei] or [ɛi], and is likely to be perceived as such. The dissimilative forces which underlie diphthongization may counteract the application of raising for this vowel. One might expect a high glide to cause assimilation of the [ei] to [ii], but the child may instead preserve or enhance the diphthongal difference.

In the sections below, I will sketch the fortitive vowel processes that can be observed in the world’s languages, and, where possible, I will give examples of children’s substitutions that exemplify them.

**Lowering and Raising**

Lowering increases sonority, and it is implicationally favored for vowels that are relatively high in sonority and weak in color. It applies most strongly to non-palatal non-labial vowels like /i/ and /o/, and accounts for their absence in many of the world’s languages. Several children described in the literature lower [a] or [ʌ] to [a] either obligatorily (e.g. the English learners in Velten 1943, Leopold 1939-49) or optionally (e.g. the learner of English and Czech in Vogel 1975).
Lowering also preferentially affects lax vowels, so that, for example, [i] and [u] are more susceptible to lowering than [i], [u]. Some children, like Sylvia Major, lower [i] and [u] to [ɛ] and [ɔ] (Major 1977). Otomo and Stoel-Gammon (1992) report lowering of /i/ to [ɛ] and /e/ to [æ] as particularly frequent substitutions. Bleile (1989) also provides an example: Jake, at 2;0 – 2;2 optionally lowered /i/ to [ɛ].

Across languages, Lowering follows similar patterns. It is implicationally favored for vowels that are neither palatal nor labial; for example, the difference between short and long /a/ is realized in many languages with an added difference of quality: short /a/ is [ʌ] and long /aː/ is [ɑː] in Sanskrit and in Hawaiian. Both /a/ and /aː/ are optionally lowered to [ə] and [ɑː] in Kolami (Emeneau 1955, 7). In languages which lack an /ə/ or /ʌ/, this vowel is typically borrowed as [ɨ], as in Japanese borrowings from English (Ohso, 1972). Lowering also favors lax vowels in synchronic alternations, in language histories, and in variation. Eastern Ojibwa short (lax) /i/, /o/ (phonetically [i], [u]) are lowered to [ɛ] and [ɔ] (the [ʃ] from [u] is unrounded) in final position (Bloomfield 1956, 4ff.). In the Sacapultec dialect of Quiche, lax vowels are lowered optionally, so [i] ~ [ɛ], [u] ~ [ɔ], [ɛ] ~ [a] and [ɔ] ~ [ɑ] (Campbell 1977, 16ff.). In Modern Icelandic, context-lengthened [iː] and [ɑː] are variably lowered to [ɛː] and [ɔː] (Einarsson 1945, 11). And in southern & western 15th century Swedish, lax /i:/ and /u:/ lowered to /e/ and /o/ (Haugen 1976, 258).

Long or lengthened vowels are especially susceptible to Lowering: in W. Greenlandic Eskimo (Pyle 1970, 133), in Pashto (Shaheev 1964, 34), and in Yokuts (Greg Lee, pers comm.), long vowels are lowered, so that [i] and [u] alternate with [ɛ:] and [ɔː:]. Vowels are also especially susceptible to Lowering when adjacent to a glide of like color (labial vowels lower near labial glides, palatal vowels lower near palatal glides); examples abound in diphthongizations (see Sec. 4.3). Although palatal and labial vowels may lower in parallel, Lowering need not affect both. In Dagur, for example, Altaic *u became /o/, but *i did not lower (Poppe 1955, 31).

Raising decreases sonority, and fortitive Raising (i.e. raising in stressed or accented positions, under emphasis, etc.) only affects palatal or labial vowels – [a] is not raised to [ʌ], nor [ʌ] to [i], in stressed or lengthened positions. We can conclude that Raising applies to increase a vowel’s color. Children often raise palatales & labials: Joan raised English [e, ɛ] to [i, ɪ], and [o, ɔ] to [u, ʊ,], while lowering [ɑ] to [ɑ] (Velten 1943, 87ff), and Linda raised Estonian /e/ and /o/ to /i/ and /u/ (Vihman 1971).

Across languages, only palatal and labial vowels are raised, and tense vowels are favored for Raising. In the English Great Vowel Shift, tense palatal and labials were raised: æ > e; e > i; o > o; o > u. But Raising may affect only palatal vowels, as when Old Gutnish æ > e, æ > e; e: i; ø > y, æ > y: (Noreen 1913, sec 140-141, 149-150) – note that labiopalatals were included. Or Raising may affect only labial vowels; this happened in the history of French (Pope,
1934), and in Sao Miguel Portuguese (Rogers, 1948, 13ff). In both, in checked syllables, [ɔ] > [o], [o] > [u].

When raising applies dissimilatively, it often affects vowels that are adjacent to non-palatal non-labial or lax vowels or glides (see Sec. 4.3.)

**Coloring**
Context-free Palatalization and Labialization affect achromatic\(^{11}\) vowels, and they are especially applicable to higher vowels. If a low vowel like [a] is labialized or palatalized, a higher achromatic vowel will undergo the process as well. As a result, many languages have an /a/ vowel but lack non-low, non-palatal, non-labial vowels, like /\(\text{\textalpha}\)/ and /i/. Palatalization and Labialization appear most clearly in children’s speech where a child substitutes [æ] or [ɔ] for /a/. Colorings may occur in very early speech, when children may use a low vowel like [a] for adult low vowels, and a single high vowel, [i] or [u], for all others (e.g. Leopold 1939, Velten 1943). In a further variation of this kind of pattern, we find Linda replacing Estonian /\(\text{\textalpha}\)/ with [i] or [u] – palatalizing or labializing it, and also raising the result, just as as she raised /e/ and /o/ (Vihman 1971).

Among the world’s language histories, we find a number of examples of context-free palatalization; these include Yellow Lahu, where Black Lahu /i/ and /\(\text{\textalpha}\)/ have merged with /i/ and /e/, but /a/ remains non-palatal (Matisoff 1973, 12), and Southern Welsh, where /i/ has become /i/, but /\(\text{\textalpha}\)/ has not become /e/ (Bowen and Jones 1960, 12). We also find that /i/ became /i/ in Northern Irish (Sommerfelt 1968, 495), Common Mongolian (Poppe 1955, 33), and the Özbek dialects of Turkic (Menges 1968, 79). Palatalization may affect all vowel heights, even low vowels, as when [a:] became [æ:] in Classical Greek (Allen 1968, 70). In Old English, Palatalization applied generally to short vowels when W. Germanic *a became /æ/, and Labialization was general for long vowels: *a: became [\(\text{\textalpha}\:) ] (Campbell 1959, 52ff). We also see Labialization where Gutob-Remo *i became /u/ in the Mundlipada dialect of Remo (Zide 1965, 44), and where the ‘enunciative’ vowel of Dravidian, elsewhere /i/, became [u] in Kannada and Telugu (Bright 1975, 41).

**“Bleaching”**
Depalatalization and Delabialization can be grouped under the term ‘bleaching’, since both result in the loss of vowel color (as well as an increase in sonority). Sylvia, for example, optionally depalatalized or delabialized English low and mid lax vowels: [bæf] ~ [baf] bath, [mo] ~ [m\(\text{\textalpha}\)] more, [marselo] ~ [mars\(\text{\textalpha}\)lo] Marcello (Major 1977). ‘Y’, a learner of French, depalatalized /y/ to [u] (Pupier 1977, 81). Linda seems usually to have depalatalized Estonian /y/ to [u], but
occasionally delabialized /y/ and /y:/ to [i] and [i:], and she delabialized /ø:/ to [i:] (Vihman 1971).

The weaker the vowel color, the more susceptible it is to Depalatalization or Delabialization, so lower vowels, lax vowels, and vowels with mixed color (labiopalatal) are more susceptible than their higher, tense, or unmixed counterparts. Thus, in 12th century English æ > α, but e did not depalatalize (Campbell 1959), and when Indo-European *e and *o merged in Sanskrit as /a/ (phonetically [ə] (Allen 1953, 58)), *i and *u remained palatal and labial (Burrow 1965, 103). In Chinautla (Pokoman) lax [ɔ], [ɛ] become [ə], but tense [ο:], [ε:] remain, and in Sacapultec (Quiché) [ɛ], [ɔ] (from lax [i], [u] ) optionally become [a] (Campbell 1977, 22). Labiopalatal vowels, which have two colors, are particularly susceptible to Depalatalization or Delabialization. In Lithuanian Yiddish (Sapir 1915, 259ff), and in the German dialects of Darmstadt, Alsace, Upper Austria, and Luxemburg (Keller 1961, 167ff), Middle High German labiopalatal /y/ and /ø/ were delabialized to /i/ and /e/. And in Monguor (Poppe 1955, 49ff) and the ‘Iranized’ Turkish dialects of Özbek (Menges 1968, 80), labiopalatal /y/ and /ø/ are depalatalized to /u/ and /o/.

Depalatalization and Delabialization may apply context-free to constrain the vowel inventory; we find languages like Gude (Hoskison 1974) or Kabardian (Kuiipers 1960) with distinctions only in vowel height. And in diphthongization, Depalatalization or Delabialization may apply dissimilatively, causing loss of a color when there is a vowel or glide of like color in the environment (see Sec. 4.3).

**Tensing and Laxing**

Tenseness may be defined as a high degree of color for a given degree of vowel height, and laxness as weakness of color; thus [i] is more palatal than [i], [u] is more labial than [u], etc. Non-palatal, non-labial vowels are, by definition, lax. Higher vowels are more susceptible to Tensing, lower vowels to Laxing. Leopold (1939), noted Hildegard’s substitution of tense [i] for English [i], and Otomo and Stoel-Gammon (1992) found that their younger learners also did so.

The application of Tensing and Laxing is evident from cross-language comparison of phoneme inventories, where we find /e/-/ɛ/ and /o/-/ɔ/ contrasts are more widespread than contrasts of /i/-/ɪ/ or /u/-/ʊ/. Among the world’s languages, potential high vowels are usually merged in favor of tense /i/ and /u/; the mid-vowel contrasts may be merged as /e/ or as /ɛ/, and as /o/ or as /ɔ/ (the tense-vowel-letter is usually chosen), but they remain as contrasts ore often than the high vowels. Long vowels appear to favor tenseness; in many languages, the colored long vowels are tense and the colored short vowels are lax. Conditioned length may condition tenseness: Spanish /e, o/ are [ɛ, ɔ] in closed and [ɛ, ɔ] in open syllables. Of course, in many languages, long and short vowels have the same quality.
Tensing (especially applicable before glides without color like [ɜː] or with opposite color) and Laxing (especially applicable before like-colored glides) are also evident in diphthongization, where the first step of a diphthongization may consist of tensing or laxing part of the vowel’s duration, (with further dissimilation to follow): e.g. [u̯u] → [u̯u] (→ [i̯u]), or [e̯ɛ] → [e̯ɛ] (→ [e̯e]).

Context-sensitive substitutions: Dissimilation
In children’s speech, as in adult languages, diphthongs arise from single vowels, from the vocalization of consonants, and from consonant loss that makes two vowels adjacent. A long or lengthened vowel is a vowel followed by a non-syllabic copy of itself: [eː] equals [e̯ɛ]. Diphthongs (two vowel qualities within a single syllable) often arise when one part of a vowel undergoes a change and the rest does not. This may occur by Raising ([e̯ɛ] → [eɪ]), or Lowering ([e̯ɛ] → [æ̯ɛ]), or Bleaching ([e̯ɛ] → [æ̯e]), or Laxing ([e̯e] → [e̯e]), and can be viewed as an initial dissimilation. Diphthongs are especially susceptible to further dissimilation; they allow the maximization of both sonority and color (or the maximization of conflicting colors).

Bleile (1989) describes diphthongization of /æ/ to [ai] in the speech of 2-year-old Kylie, an American English learner: hand [hæn], and flag [fæŋk]. Kylie also dissimilates /eɪ/ to [ai]: play [peɪ], paper [paɪpə]. Sylvia Major diphthongized both palatal and labial vowels in English: [peɪŋki] Pinky, [spoʊn] spoon, [daʊli]~[ dæʊli] dolly, [sɛi] see (but not in Portuguese: only [se], not *[sɛi], for (vo)sé ‘you’). She also dissimilated /aɪ, ɔu/ to [iː, æʊ] or [iː, ɛʊ], as in [uːt neʊ] right now!. Moreover, Sylvia’s lax and low vowels /, e, o, ə, æ/ variably acquired a schwa offglide, and the vowels were sometimes tensed or raised, e.g. [dəʊɡ] ~ [doʊɡ] dog, [hɛd] ~ [hɛd] head (Major 1977).

Fortitive processes are subject to the same implicational conditions in dissimilative applications as in context-free applications; for example, if the syllabic of [u̯u] is delabialized to [i̯u], then that of [o̯u] will also be delabialized to [aʊ], since Delabialization is especially applicable to lower vowels. Diphthongization and diphthongal dissimilation occur most often in long or stressed syllables, especially in slow or emphatic speech; e.g. for my daughter Elizabeth, a word like no!, usually [nəʊ] or [nəʊ] could become [nɔu] (by Lowering) or [nɛu] (by Palatalization) or even [nɑu] (by both) under extreme emphasis.

In many accents of English, the adult vowels vary between monophthongal and diphthongal manifestations, depending on accent and other lengthening conditions: even in quite conservative dialects, based on RP and General American (Wells, 1983), [iː] ([iː]) varies with [iː] or [iː], and [uː] ([uː]) varies with [uː] or [iː] or even [uː], so it is hardly surprising if we find diphthongizations in children’s speech. Monophthongal vowels may diphthongize in children’s speech, as in adult speech, by dissimilation (/eː/ becomes [eɪ], /eː/ becomes [ɛə]), or by partial assimilation to adjacent consonants (/æʃ/ becomes [æʃ]) – see above.
But there are additional sources of diphthongs in children’s speech, which may lead to dissimilative changes that might seem unexpected. Approximants may become vowels, and these may form diphthongs. English approximant /r/ seems to have rhotic, palatal, and labial qualities. Syllabic /ɹ/ may lose its rhotic quality and become [y] or [ø] (as in my daughter’s pronunciation of *Ernie and Bert* as [ˈɔni ən bɔp] (2;1). If /ɹ/ also loses palatality it may become [ʊ] or [ɔ] (*Ernie* as [uni] or [əni]) or, if it loses or labiality, [i] (*bird* as [bi]). And if /ɹ/ loses rhoticity, palatality, and labiality, it becomes [æ], as it does in unstressed syllables for many children (*paper* as [ˈpeɪpə]). /l/ may lose its laterality, and, since it is usually velarized when syllabic, it often becomes [u] or [o]. This happens in adult speech as well. In Hawai‘i English, local and people are [lɔko] and [pipo].

Consonantal /r/ and /l/ may become the glides ([y, i, u] or [ʊ, j, w]) that correspond to these vowels. Two-year-old Kylie pronounced /ar/ as [aɬ] in *cards* [taɪɛz], *market* [maɪkɪt] (Bleile 1989). Menn’s subject Jacob also substituted [i] for some postvocalic /ɹ/’s: *horse ~ horsie* [haɪʃiːz]~ [haɪʃi]~ [haɪʃi] and *more* [moʊ] (Menn 1978). If an approximant becomes a glide, the adjacent vowel may dissimililate: e.g. if [ɪ] becomes [u] in words like *tell* and *belly*, the preceding [ɛ] may lower, yielding [taɪ] and [bɑɬi] (or [bawɪ]). In fact, dissimilative lowering may occur before [ɪ], even if it does not lose its laterality. And [i], whether or not it is de-rhotacized to [ə], may also condition a variety of diphthongizations and dissimilations: vowels often develop schwa offglides before [i], as is standard in RP (where *steering, fairy* are [stɹɪəɹi, feɹi]), but they may also become lax, as in General American ([stɹi, feɹi]), or tense, as in some East Coast U.S. dialects ([stɹi, feɹi]).

Diphthongs may also arise in children’s speech when intervocalic consonants are lost, e.g. Hildegard’s [baɪ] for *buggy* (Leopold 1939), Jacob’s [bæɬ] for *battery* (Menn 1978, 278). Dissimilation within the diphthong may follow. My daughter's pronunciation of words like *butter* (adult [ˈbʌtəɹ]) varied considerably (2;1), from emphatic [bʌdi] to less careful [bʌri] or [baɬ] to [baɬ]. This [ai] pronunciation, the result of dissimilation of [ʌə] from the following [i], also occurred in words like *bunny* and *funny*, where the /n/ was flapped and deleted: [baɬ, faɬ].

Diphthongs may in turn be resolved into disyllabic sequences. Bleile (1989) describes two children who insert [ə] after a final /aɬ/; one of them also inserts [ə] after final /ɔɬ/ as well.

*Context-sensitive substitutions: Monophthongization*

Monophthongization is a context-sensitive substitution in the sense that it occurs by the complete mutual assimilation of the parts of a diphthong; when /ai/ becomes [æ], the [a] assimilates to the palatality of the following [i], and the [i] assimilates to the openness of the preceding [a]. Many children, like Sylvia, start out with all monophthongs (Major 1977). Others, like Linda, may
monophthongize more-similar sequences like /ei/ and /ai/ but allow less-similar sequences like /ai/ to remain diphthongal (Vihman 1971). Hildegard also monophthongized /ei/ and /oʊ/, but pronounced /aɪ/ and /aʊ/ as diphthongs (Leopold 1939). Monophthongization of palatal and labial diphthongs is not always symmetrical. A Smith monophthongized /ei/ in rain to [e], /ɔi/ in noisy to [ɔː], and /ai/ in fire to [æː]; /ou/ in soap was at first variably monophthongized to [uː], but it later remained diphthongal [aʊ] (Smith 1973).

Monophthongization is often associated with shortening; a syllabic-plus-offglide is equivalent to a long vowel in most languages, while a simple syllabic can be short. So monophthongizations like /ei/ → [e] are sometimes seen simply as glide loss, and monophthongizations like /ou/ → [u], as loss of the syllabic and reassignment of syllabicity to the [u]. But we could also say that if /ei/ monophthongizes to [e], the [i] assimilates to the height of the [e], and if /ou/ monophthongizes to [u], the [o] assimilates to the height of the [u]; in both cases, shortening may or may not occur.

The Similarity Principle (segments that are similar are implicationally favored for further assimilation) predicts that [ei] is more susceptible to assimilation than [oɪ] or [ai]. The similarity principle appears to govern monophthongization, so there is reason to regard monophthongization as complete assimilation. A child who monophthongizes [ei] to [e] or [ou] to [o] may be able to say the less-similar sequences, [ai] or [aʊ], as A Smith and Vihman’s Linda could.

Other Lenitive Processes: Reduction and Harmony

In children’s speech, as in adult languages, vowel reduction is related to speech timing. In adult languages, we expect that where syllables have relatively equal timing, we will find ‘full’ vowels. Unless syllables are considerably shortened, we will not find the loss of sonority and color, and the limitation to a small set of possible vowel qualities which mark the unstressed syllables of languages like English, or Khmer. Vihman notes that ‘a correlation between rate of articulation and children’s age is commonly reported’ (1996, 213). Thus, it does not seem coincidental that Allen and Hawkins (1978) found that the 3-year-old English-learners they studied did not reduce syllables as generally as adult speakers do, although there was considerable variation across children. Allen and Hawkins comment that children’s earliest words ‘typically have only heavy syllables’ with ‘peripheral (non-central) vowels and rather fully articulated consonants’ (174). It seems that reduction may occur if unstressed, shortened syllables appear in a child’s speech, but that in many cases, unstressed syllables are deleted entirely. If they survive, they are not articulated quickly enough to reduce. Allen and Hawkins also note that because stressed syllables are often entirely absent in the speech of 1- and 2-year-olds, their speech may appear to be syllable-timed (1980, 231). Levelt (2008) attributes the late development of final schwa (in Dutch) to the greater duration of final syllables in children’s speech; Kehoe and
Lleo (2003) regard schwa as phonetically distinct from other vowels (although they say it is not necessarily shorter) and German schwa as acquired rather late. This appearance (or percept) of syllable timing in child speech might suggest that vowel harmony, a phenomenon associated with syllable- or mora-timed languages, might be fairly pervasive in children’s speech. But vowel harmony is less common in children than we might expect. Examples are reported where an epenthetic or paragogic vowel shares the features of a preceding vowel: Ross (1937) gives examples from one English-speaking child, and Chervela (1981) cites a number of examples in Telugu learners. Reduplication may result in matching vowels in consecutive syllables (Ingram 1971, Vihman 1996), and an unstressed vowel may match an that of adjacent syllable (Drachman and Malikouti-Drachman 1973). But children who spread common values of palatality or labiality across whole multisyllabic words, as in the harmony systems of languages like Turkish or Finnish, are to my knowledge unreported.

**Variation**

Children may eliminate any particular difficulty with a range of processes: e.g. the particular combination [non-high, non-low, non-tense, palatal] presents a certain difficulty, so [ɛ] may be raised, or lowered, or tensed, or depalatalized. Because vowels are simultaneous combinations of conflicting features, and increasing one feature, like labiality, results in weakening another, like sonority (and vice versa), it seems that virtually any change can be motivated. Certainly we cannot predict the order of acquisition of vowels. (If we could, we might also expect the vowel inventories of the languages of the world to be in entirely subset relationships. In comparing vowel inventories, we do find some seductively subset-like partial patterns, but these are not universal.)

Yet some predictions can be made – not necessarily about the order of acquisition of particular vowel qualities or even of particular features, but rather about the substitution patterns that may occur within the speech of an individual child. For example, if [ɛ] is tensed, we may expect that in similar circumstances [i] will be tensed as well (higher vowels are more susceptible to Tensing); if [ɛ] is depalatalized, we can have no expectation that [i] will depalatalize, too (lower vowels are more susceptible to Depalatalization). We expect substitutions to conform to implicational conditions on processes because these conditions reflect the phonetic motivations – articulatory or perceptual – of the processes. If a child unrounds [ʊ] to [ɨ] (losing labiality, increasing sonority), we can expect unrounding of [ɒ] to [ɑ], since the phonetic basis of [ʊ]-to-[ɨ] is the same as that of [ʊ]-to-[i], but stronger. In [ɒ]-to-[ɑ], there is less labiality to lose, more sonority to increase.

Individual children show variation, as noted above. Children lack articulatory control and experiment with different substitutions, particularly in the early stages. By the time the child settles on consistent substitutions in
consonants, the vowels in his or her speech may be reasonably accurate. But variation remains; as in adults, process application may vary with effort and with attention paid to speech (cf. Elbers and Wijnen 1992), or with affect (e.g. my daughter Elizabeth’s pronunciations of her own name at 2;8 ranged from hyperarticulate, emphatic \[\text{ˈiˈjazəbɪs}\], with full vowels and lowering of the stressed (lax) vowel, to hypoarticulate, too-sleepy-to-talk \['\text{ˈmiʃbɪf}\] with vowel reduction, unstressed vowel deletion, and consonant place harmony). It is important to note that variation confirms the phonetic motivation of substitutions, because it reveals how different demands on the child’s abilities may elicit different substitutions as responses. And it is important to remember that the child’s models vary, as well.

Ability or desire to produce a sound or sequence in one set of circumstances does not necessarily extend to all circumstances. Words may be more accurately produced in imitation than in spontaneous utterances. This is sometimes used to argue for inaccuracy or incompleteness of a child’s lexical representations, but imitation may involve increased effort and attention to pronunciation, while spontaneous productions require the organization of sequences of articulations to combine with the effort of lexical recall.

Imitated forms are also used to argue against the characterization of substitutions as expressions of ‘inability’. Obviously, if a child pronounces a particular sound or combination under some conditions, he or she has the phonetic ability to pronounce it. But s/he may not be able to pronounce it under all conditions. If substitutions occur, they may still be phonetically motivated. Consider that all English-speaking adults can say \[\text{ˈprɛzɪdənt əv ˈdʒuˈnæɪt̚dˈstɛɪts}\]. But there is still phonetic motivation for the substitutions that result in (American) \[\text{ˈprɛzdˈnæʃdˈstɛɪts}\] in connected speech. We may even claim that American adults are unable to use the hyperarticulate form in all circumstances, as they are unable to avoid pronunciations like \[\text{ˈkɪtɪ}\] kitty in connected speech, even though they are capable of saying \[\text{ˈkɪtɪ}\] in hyper-articulate speech.

**Conclusion**

The general cross-linguistic and cross-child patterns that we find cannot be attributed to a simple principle like ‘maximal dispersion’, which predicts that any given number of vowel phonemes will be distributed within ‘vowel space’ so as to maximize the sum of their differences. This principle predicts that the expected three-vowel system will be \(/i, ə, u/\) and the most common five-vowel system \(/i, e, ə, o, u/\) – but the predictions of the principle do not correspond to the actual vowel systems of the world’s languages very much beyond this. Further, this principle alone gives no way to account for vowel systems (like \(/i, ə, a/\) or \(/e, o, ə/\)), which are not maximally dispersed. Vowel systems, like consonant systems, are based on feature contrasts, and they are constrained by
the application of processes which are implicationally conditioned to apply where their phonetic motivation is strongest.

Bloomfield (1933), in discussing sound change, describes a wide variety of changes in the histories of languages, and he points to an array of phonetic motivations which seem to underlie these changes. But at the end of the discussion, he rejects the idea that these changes are phonetically motivated, on the grounds that if sound change were phonetically motivated, all languages would develop toward a single optimal phonology. Bloomfield, however, did not consider that speakers’ learning is the key to the paradox. Children at first submit to their phonetic inclinations, but they learn, in developing their abilities, to control articulation and thereby inhibit some phonetically motivated substitutions in order to perceive and pronounce the sounds and sequences their communities require.

Many children’s vowel substitutions almost escape notice, with the result that the data we have on normal vowel development does not show nearly the variety that vowel substitutions in the world’s languages might lead us to expect to find in children. In fact, it seems likely that most instances of apparently abnormal vowel development – unusual substitutions – are not abnormal at all, except perhaps in being too persistent, and that in the case of vowels it is speech pathologists rather than developmental linguists who are in the best position to discover and publish the facts of what children do about vowels they can’t pronounce. The aim here has been to show that the vowel substitutions of children respond to the same phonetic demands as the vowel systems of the world’s languages, and to suggest that normal children might take many different paths to the mastery of the vowels of their languages.
References


ROA version 2002: http://roa.rutgers.edu/files/537-0802/537-0802-PRINCE-0-0.PDF


Figure 1. Vowel Symbols and Features
In this framework, only three degrees of vowel height are phonologically relevant. Apparent four-height systems involve differences in tenseness (intensity of palatality or labiality) for a given vowel height. Non-palatal, non-labial vowels are lax. Note that the IPA has no symbol for a lax, low, labial vowel.