1187

In: deGelder, B. & Morais, Jose (eds.) (1995). Speech and Reading: A Comparative Approach Hove, England: Erlbaum (UK), Taylor & Francis

4 Gestures, Features and Segments in Early Child Speech

Michael Studdert-Kennedy and Elizabeth Whitney Goodell
Haskins Laboratories, New Haven, CT, USA

INTRODUCTION

Preliminary

That an alphabetic orthography represents speech at the level of the phoneme seems to be generally agreed. But the definition of the phoneme and even its functional status are still matters of contention among linguists. We do not propose to join the linguistic argument here. We take the facts of reading and writing to be sufficient evidence for the functional reality of the phoneme as a perceptuomotor control structure representing a class of phonetic segments (cf. Studdert-Kennedy, 1987). We assume, further, that phonetic segments (consonants and vowels) are not the irreducible elements of which speech is composed. Rather, segments are complex structures, implicit in the gestural patterns of speech, that gradually emerge and take on their perceptuomotor functions over the first few years of life. This chapter attempts to justify this claim by applying a developing theory of articulatory phonology (Browman & Goldstein, 1986; 1989) to a small set of data drawn from the utterances of a 2-year-old child.

Background

A child, learning to talk, often says the same word in several different ways. Indeed, variability of phonetic form has been a commonplace of child language studies since their inception (e.g. Albright & Albright, 1956; 1958; Cohen, 1952; Leopold, 1953). On the other hand, a child, learning to talk, often says several

different words in the same way, and this homonymy is also a commonplace of child language studies (see Vihman, 1981, for review). While variability and homonymy may reflect many factors, including the communicative situation, whether the utterance is spontaneous or imitated, its meaning, phonetic structure, phonetic context and so on (Schwartz, 1988), none of these factors would matter, if it were not that "... a child's phonemic system is in the process of development, and the sound patterning is probably less regular than that of adult speech" (Albright & Albright, 1956, p. 382). But what, in fact, is developing, and what is the nature of the irregularity? What varies in the execution of a target word from one occasion to another? What does the child find that different words have in common?

Let us begin with the observation, now supported by a variety of evidence, that the unit of phonological contrast, and therefore the unit of articulatory organisation, in early child speech is the whole word or phrase rather than the segment (Ferguson, 1963; 1986; Ferguson & Farwell, 1975; Macken, 1979; Menn, 1983; Menyuk, Menn, & Silber, 1986; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Waterson, 1971). To say that a child utters a word as a "whole", or Gestalt, cannot mean, however, that the child has not broken the word into at least some of its parts, because even a partly correct utterance requires co-ordination of independent, or partially independent, actions of lips, tongue, jaw, velum, and larynx. Accordingly, while the word may be the domain over which a child organises its articulations, it cannot be the basic unit of production (or, *a fortiori*, of perception). Nor, as the segment is no less compounded of independent articulatory actions than the word, can the basic unit be the segment itself.

The standard unit adopted in studies of child phonology is, of course, the feature. By this we cannot mean the abstract feature of generative phonology, a relational property fulfilling the linguistic function of contrast across a phonological system, because we are dealing with a child for whom such a system does not yet exist. We must therefore mean the concrete feature, an absolute property located "... within the speech sounds, be it on their motor, acoustical, or auditory level" (Jakobson & Halle, 1956, p. 8). However, at least two facts make this proposal unacceptable. First, the feature has no independent existence: It is a property of a larger unit and is carried into existence on that unit, as an adjective, not a noun (cf. Fowler, Rubin, Remez, & Turvey, 1980). This fact is implicit in the adjectival terminology of all feature theories: grave, acute, compact, coronal, nasal, strident, and so on.

A second, closely related objection is that the properties to which featural terminology customarily refers are purely static, devoid of temporal extension, and therefore intrinsically unfit to define the dynamic properties of speech either as a motor act or as an acoustic signal. Not surprisingly, none of several sets of acoustic and articulatory definitions of phonological features (e.g. Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1952/1963; Stevens, 1972; 1975; 1989) has proved precise or full enough to support a procedure for speech synthesis or speech recognition by machine, let alone a theory of speech production or speech perception.[1] The "autonomous features" of autosegmental and other forms of nonlinear phonology (e.g. Clements, 1985; Goldsmith, 1990; McCarthy, 1988; cf. Menn, 1978) might seem to promise a solution. However, these features are abstract units, the temporal analogs of points in Euclidean space, admitting of sequence, but not of extension. In short, as Ladefoged (1980, p. 485) has remarked, "... phonological features are certainly not sufficient for specifying the actual sounds of a language". They can hardly therefore help a child striving to learn how to perceive and produce those sounds.

Very much the same holds for the informal, and perhaps intuitively more appealing, phonetic features adopted by, for example, Waterson (1971, p. 183; cf. Ferguson & Farwell, 1975; Macken, 1979). She analyses a child's word forms into: "Various features of articulation, such as nasality, sibilance, glottality, stop (complete closure), continuance, frontness, backness, voicing, voicelessness, labiality, rounding, non-rounding." Waterson goes on to group the child's forms into "structures" or "schemata", corresponding to the adult "prosodic" patterns in which features are distributed over a word. She shows how in a child's utterance the "prosody" may be disrupted, so that features lose their temporal order and recombine into patterns quite unlike the adult model. In this respect, Waterson's work has stimulated the approach taken in this chapter. None the less, her schemata are purely descriptive, indications of, but certainly not specifications for, the spatio-temporal pattern of movements by which a speaker, child or adult, executes a word. They offer, at most, a sketch of the high points of a word, rendered in the familiar language of traditional articulatory phonetics. We conclude that, despite the utility of the feature as a descriptive and classificatory element in phonetic theory, it cannot guide a child into speech.

In fact, what a child quite evidently needs, to imitate an adult word, is a grasp on which articulators to move where and how, and on when to move them. And what we, for our part, need, to understand the child's attempts, is a description of the target word in terms of the units of articulatory action, and their relative timing, necessary to utter it. No generally agreed upon set of articulatory units exists, although several have been proposed. Ladefoged (1980), for example, offered a tentative list of 17 "articulatory parameters" that he judged necessary and sufficient to specify the sounds of a wide range of languages, but he did not develop them into a functional model. Here we adopt the framework of the most explicit model of speech production currently available, the gestural phonology being developed by Browman, Goldstein, Saltzman and their colleagues at Haskins Laboratories (Browman & Goldstein, 1986, 1987, 1989, 1990; Saltzman,

[1] Fant (1962, p. 4) remarked many years ago concerning the theory of Jakobson, Fant and Halle (1952/1963): "... its formulations are made for the benefit of linguistic theory rather than for engineering or phonetic applications. Statements of the acoustic correlates to distinctive features have been condensed to an extent where they retain merely a generalised abstraction insufficient as a basis for practical applications." The same is true of subsequent attempts to formulate acoustic and articulatory correlates of the features.

1986; Saltzman & Munhall, 1989), in which the basic phonetic and phonological unit is the gesture. We illustrate the approach with a small set of data, drawn from the utterances of a 21–25-month-old girl, learning American English. But before we come to the data we must give a brief sketch of the gestural framework.

Gestures as Basic Units of Articulatory Action

If we watch, or listen to, someone speaking, we see, or hear, the speaker's mouth repeatedly closing and opening, forming and releasing constrictions. In the framework of gestural phonology, each such event, each formation and release of a constriction, is an instance of a gesture. Constrictions can be formed within the oral, velic or laryngeal articulatory subsystems; within the oral subsystem, they can be formed by the lips, the tongue tip (blade) or the tongue body. The function of each gesture, or act of constriction, is to set a value on one or more vocal tract variables that contribute to the shaping of a vocal tract configuration, by which (in conjunction with pulmonic action) the flow of air through the tract is controlled, so as to produce a characteristic pattern of sound. Presumably, this pattern of sound specifies for a child (or an adult) the gestures that went into its making.

Figure 4.1 displays the tract variables and the effective articulators of a computational model for the production of speech, at its current stage of development (Browman & Goldstein, 1990). The inputs to the model are the parameters of sets of equations of motion for gestures. A gesture is an abstract description of an articulator movement,[2] or of a co-ordinated set of articulator movements, that unfolds over time to form and release a certain degree of constriction at a certain location in the tract. Settings of the parameters permit constriction degree to vary across five discrete values (closed, critical, narrow, mid, wide), and constriction location for oral gestures to vary across nine values (protruded [lips], labial, dental, alveolar, post-alveolar, palatal, velar, uvular, pharyngeal).[3] The reader may observe that the degree and location of an oral constriction roughly correspond to the manner and place of articulation of a segment in standard terminology.

The gestures for a given utterance are organised into a larger co-ordinated structure, represented by a gestural score. The score specifies the values of the dynamic parameters for each gesture and the period over which the gesture is active. Figure 4.2 schematises the score for the word *nut* ([nʌt]), as a sequence

[2] In what follows we use the term "gesture" to refer either to an underlying abstract control structure, or to a concrete instance of a gesture activated by this structure, relying on context to make clear which is intended.

[3] These categorical values, axiomatic within gestural phonology, may have emerged evolutionarily, and may still emerge ontogenetically, through auditory and articulatory constraints on individual gestures (Stevens, 1989), and on the entire set of gestures within the child's developing lexicon (Lindblom, 1986; Lindblom et al., 1983).

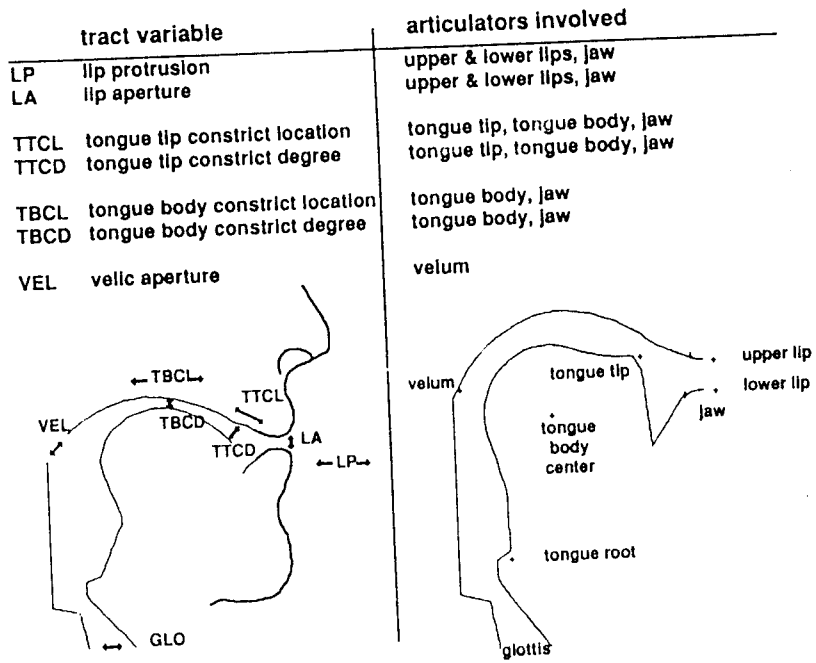


FIG.4.1 Tract variables and associated articulators used in the computational model of phonology and speech production discussed in the text (adapted from Browman & Goldstein, 1990).

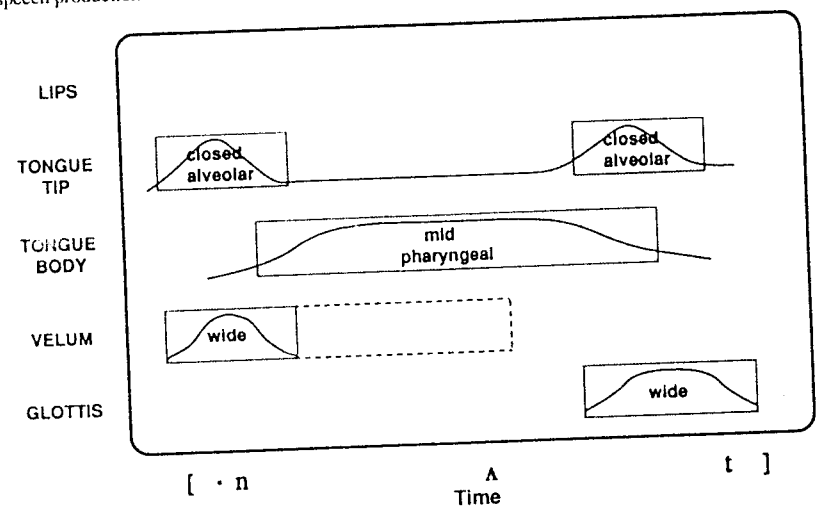


FIG.4.2 Schematic gestural score for the word *nut* ['nʌt]. The boxes indicate the activation intervals called for by the score, the solid lines the tract variable changes over time. The extension of the velic activation interval by dashed lines indicates possible free variation in the duration of the velic gesture.

of partially overlapping gestural activation intervals with the generated tract variable motions superimposed; possible free variation in the duration of the velic gesture, and the resulting nasalisation of the vowel, is indicated by extending the velic activation interval with a dashed line.

We cannot here go into detail on the workings of the model (for which the reader is referred to the papers cited earlier). We note only the following further points that, taken with the preceding sketch, may suffice for an intuitive grasp of how a gestural framework can contribute to an understanding of the nature and origin of irregularities in a child's early words.

1. An instance of a gesture is an objective, observable event. We can observe a gesture by ear, and this is the usual basis of both imitation and phonetic transcription. We can observe a gesture by eye, either unaided, as in lip-reading, or with X-ray cinematography. We can observe a gesture by touch, as in the Tadoma method of speech perception. Finally, we can observe a gesture by sensing our own movements. However, if gestures drawing on the same, or closely neighbouring, neuromuscular sets overlap in time (as, for example, in certain lingual gestures for the consonantal onset and vocalic nucleus of a syllable), the individual gestures may merge, so that we can observe only the resultant of their vectors.

2. The articulator sets and their dimensions given earlier are not exhaustive. For example, the tongue root must ultimately be included in the model to handle variations in pharynx width. Also, constriction shape will have to be included, to handle the tongue bunching, narrowing or hollowing, necessary in the formation of certain complex gestures (cf. Ladefoged, 1980). Even the definition of the gesture itself may have to be revised to permit independent control of the formation and release of a constriction.

3. A gesture is larger than the properties of constriction location, degree, and shape that describe it, but smaller than the segment. Several independent gestures are required to form a segment, syllable, or syllable string.

4. Each gesture has an intrinsic duration that varies with rate and stress. Correct execution of an utterance requires accurate timing of the gesture itself, and accurate phasing of gestures with respect to one another.

5. By adopting the gesture as a primitive of articulatory action we can predict what types of errors children are likely, or not likely, to make in their early attempts at adult words. This topic we now discuss.

The Gestural Origins of Errors in a Child's Early Words

Phonetically, speech emerges over the first year of life from the lip smacks, tongue clicks, and pops associated with the vegetative processes of eating and breathing, combined with the stereotyped vocalisations of cries and comfort sounds (Stark, 1986), through the reduplicated syllables of canonical babble, into the brief strings of phonetically contrastive elements that make up early words. Articulatorily, the progression is a cyclical process of differentiation and integration by which the child moves toward finer modulation of individual gestures and more precise phasing of their sequence. (For a fuller account of the hypothesised developmental course, see Studdert-Kennedy, 1991 a, b.)

Here, two steps are of interest. The first is the shift in gestural timing associated with the integration of pre-babbling oral and laryngeal gestures into the canonical syllable, usually around the seventh month (Holmgren et al., 1986; Koopmans van Beinum & van der Stelt, 1986). Earlier vocalisations, termed "marginal babble" by Oller (1980), are commonly longer than adult syllables, but display adult-like properties of resonance, intensity, and fundamental frequency contour. Canonical babble is marked by integration of a resonant nucleus with rapid (25–120msec) closing gestures at its margins to form a syllable with adult-like duration (100–500msec) (Oller, 1986). The canonical syllable is the first step in the emergence of two major classes of oral gesture: the narrow or complete constriction of consonants and the wider constriction of vowels (cf. MacNeilage & Davis, 1990).

The early canonical syllable is often, although not always, one of a rhythmic, reduplicated string of identical syllables. Reduplication indicates first that the child may lack independent control of the closing constriction at the margin and the wider constriction at the nucleus of a syllable; second, that the child cannot easily switch gestures in successive syllables. The tendency to reduplicate may continue for many months, or even years, as evidenced by the commonly reported harmony in early words between consonants (Vihman, 1978) and vowels, and even within consonant–vowel sequences. The last is revealed, for example, by certain children's preference for high front vowels after alveolar closures, and for low back vowels after the relatively extensive jaw-lowering release of labial closures (Davis & MacNeilage, 1990; cf. Jakobson, 1941/1968, pp. 29, 50).

Integration of pre-babbling gestures into the canonical syllable is a necessary condition of a second step: differentiation of the syllable into independent gestural components. Differentiation gives rise to what Oller (1980) terms "variegated babble" in which the consonant-like syllable onset and the vowel-like nucleus, or both, differ in successive syllables. The process may begin soon after, or even at the same time as, the onset of canonical babble, but typically comes to predominate in the fourth quarter of the first year, and continues over many months in both babble and early words (Davis & MacNeilage, 1990; Vihman et al., 1985).

Before we consider the types of error that a child may make we should note that, although we shall appeal to similarity among gestures as the basis of a child's confusions, we will not spell out the dimensions of similarity. In fact, we shall deliberately avoid the question of whether those confusions reflect an incomplete percept (under which we may include incomplete storage of the percept in memory) or inadequate articulatory control (under which we may

include failure to recover stored motor commands from memory). Perhaps, indeed, there is no general answer to this question: Similar errors may reflect different processes in different words and in different children. Here we adopt a neutral stance. We suppose that learning the phonology of a language is a matter both of learning to perceive the acoustic pattern that specifies a talker's gestures, and of learning to plan and produce that pattern oneself. Both these processes can be a source of error.

Differentiation itself has two aspects, each open to characteristic forms of error. The first is paradigmatic differentiation among individual gestures. Possible errors here follow from failure to identify or execute the location, shape, or degree of a gesture; in the limit, an error of degree (or amplitude) may yield complete omission. The second aspect is syntagmatic differentiation among gestures in a particular utterance. Possible errors here include gestural reduplication (harmony), errors of timing (duration and relative phasing, including metathesis), and errors of amplitude or degree. The consonant–vowel (or vowel–consonant) harmony noted earlier may be viewed as a syntagmatic error arising from incomplete differentiation of the syllable into its component consonantal and vocalic gestures. Our purpose in what follows is to illustrate how the erroneous forms of a child's early words can be perspicuously described as arising from gestural errors such as these.

METHOD

The subject "Emma" is a second child, born in Connecticut, USA, to parents who had moved there from Vancouver, British Columbia. Emma's mother was the full-time caregiver for the child. The second author of this chapter (E.W.G.) lived with the family before and during the study and spent several hours a day observing Emma at meals, watching her play with her older brother and interact with her parents, and occasionally looking at picture books with her. One of these books (*Richard Scarry's Best Word Book Ever*) was a rich source of new words (Scarry, 1980).

Audio recordings began when Emma had a vocabulary of about 100 words, mostly monosyllables understood primarily by her mother and brother. The size of her vocabulary was assessed with the MacArthur Communicative Development Inventory for Toddlers, and by maternal report. In the 91st week her mean length of utterance (MLU) was 1.00 and at the end of the study (week 106), 1.15. For the weekly audio taping sessions, lasting from 30 minutes to an hour, she wore a wireless 831 Audio Technical lapel microphone concealed in a vest. E.W.G. was present at all sessions and kept a diary of the subject's phonological development to supplement the recordings.

To facilitate the transcription and analysis of Emma's utterances, recordings of the sessions were digitised on a VAX 780 computer, at a 20kHz sampling rate, to yield a total of some 950 utterances of which the experimenter and a colleague independently transcribed roughly 250. Transcription followed the principles of the International Phonetic Association (1989), with some elaborations according to the Stanford system for transcribing consonants in child language (Bush et al., 1973). Each utterance was coded as either spontaneous or imitated; an utterance was assumed to be spontaneous, unless it immediately followed the adult target; all the examples reported in this chapter were spontaneous, unless otherwise indicated. We report only utterances on which the two transcribers independently agreed. The transcriptions will be given in square brackets, following the convention for adult phonetic segments. We emphasise that our use of phonetic symbols does not imply that segments were already established in the child as discrete units of perception and production. A phonetic symbol is simply a convenient shorthand for combinations of laryngeal, oral or velic gestures.

RESULTS AND DISCUSSION

Many researchers have described how a child, making the transition from reduplicated babble to variegated speech, discovers a pattern that roughly fits a fair number of adult words, and so can serve as a bridge into the lexicon. These patterns, variously termed prosodic schemata (Waterson, 1971), canonical forms (Ingram, 1974), articulatory routines, programs, templates (Menn, 1978; 1983), word patterns (Macken, 1979), or vocal motor schemes (McCune & Vihman, 1987) will be treated here as routinised gestural scores. (Much of our analysis will indeed follow the lead of Waterson, 1971, Macken, 1979, and, particularly, Menn, 1978; 1983, whose attention to the articulatory organisation of a child's utterances anticipates our own). Gestural routines support both stereotypy (including homonymy) and variability in a child's early attempts at words; they are of interest because the gestural properties common to a particular score and to the different target words for which it is used reveal the scope of a child's gestural conflations.

Stereotypy

Gestural Routines in Babble and Word Play. During the sessions themselves Emma did not babble much, but diary entries from the first month of the study (weeks 91–94) often record Emma's babbles while quietly playing. For example:

(1) [ɑ'bi:n'ɑ'bi:n'ɑ'bi:n]

(2) ['be:'də'be:'də'be:'də'be:'də]

These utterances happen to consist of repetitions of one of Emma's forms for *elephant* ('ɑ'bin]) and *playdough* ('be:'də]), but she chanted the sequences in a sing-song, with no apparent communicative intent. Both utterances contain the alternating sequence of constrictions at the lips and at the alveolar arch that proved to be Emma's most productive gestural routine. (For an

example of another child, learning Mexican Spanish, with a similar routine, see Macken, 1979).

Strings of similar alternations occasionally occurred in taping sessions over the same period:

(3) ['mʌtʃ:'mʌtʃ:'mʌtʃ:]
(4) ['ɑ'bu:'di:'ɑ'bu:'di:'ɑ'bu:'ku:ki:]
(5) ['we:'da'wi:'da'me:'na'mi:'nə'mu:'ni:'mi:'ni:'mi:'ni:]

Emma repeatedly produced (3) in weeks 92 and 93, elongating the final frication, as though savoring the flow of air, and with no apparent referent. (4) contains two of Emma's words (see Table 4.1 for ['a'bu:'di:]), while (5) is a mixture of apparent nonsense syllables and word forms (see Table 4.2), but none of the objects to which the words refer was present. In (4) she abruptly broke off her labial-alveolar chant when a cookie came into view. In (5) she seemed to be playing with the location and degree of labial constriction ([w]–[m]), the degree of accompanying velic constriction ([w]–[m], [d]–[n]), and (in an apparent instance of vowel–consonant harmony), the front–back location of narrow constrictions at the syllable nucleus before an alveolar constriction ([i:]–[e:]–[u:]).

These examples illustrate the emergence of gestural stereotypy in babble and word play. They also illustrate the familiar, but important fact that a listener often cannot distinguish, by phonetic form alone, between syllables that are babbles and syllables that should count as a word.[4] Despite the discontinuity of function that Jakobson (1941/1968) noted many years ago, babble and speech are formally continuous.[5] An adequate account of the shift in function must therefore posit units of action that can be comfortably engaged by both babble and speech. Phonemes, phonetic segments, and features are unsuited to this task: They cannot properly be adopted for pre-linguistic babble because they are defined in terms of language and speech. Moreover, as already noted, segments are complex units, customarily

[4]The problem and criteria for its solution are thoroughly discussed by Vihman and McCune (in press).

[5]Much of the controversy over the issue of continuity between babble and speech has arisen, in our view, from a misreading of Jakobson's claims, and from a failure to distinguish between phonetic form and phonetic function. Jakobson himself drew this distinction quite clearly. Although he believed that ". . . a short period may sometimes intervene . . . in which children are completely mute", he also recognised that: "For the most part . . . one stage merges unobtrusively into the other, so that the acquisition of vocabulary and the disappearance of the pre-language inventory occur concurrently" (Jakobson, 1941/1968, p. 29). In fact, he assumed what later studies have conclusively demonstrated (e.g. MacNeilage, Hutchinson, & Lasater, 1981; Oller, Wieman, Doyle, & Ross, 1976; Vihman et al., 1985) that listeners often cannot distinguish, by phonetic form alone, a "child's embryo-words from the pre-language residue" (Jakobson, 1941/1968, p. 29). The discontinuity that Jakobson (correctly) posited was a discontinuity of function, not of form.

TABLE 4.1
Active Use of the Labial-alveolar Routine as a Bridge into the Lexicon

New Words	*Adult Target*	*Emma's Attempts*
Cranberry	['krænberi])	['be:'bi] ['bo:'be:'bi] ['ɑ'bu:'di:]
Red Lights	['red'laits]	['we:'jaɪ] ['be'tθaɪts]
Hippopotamus	['hɪpə'pɑtəmʌs]	['ɑ'pɪnz] ['hɪpus]

analysed into their featural predicates, whereas features have no existence independently of the syllables and segments they describe. By contrast, the posited gesture is an integral unit of action that can serve equally as a primitive unit of both babble and speech.

Gestural Routines as Bridges into the Lexicon. A child who has discovered a gestural routine, such as the labial-alveolar sequence described above, will often extend it to a surprisingly diverse collection of new words in which it recognises the appropriate pattern (Macken, 1979)—in Emma's case words as diverse as *cranberry*, *red lights*, and *hippopotamus* (see Table 4.1). Thus in the recording session of week 92, Emma's mother showed her a cranberry for the first time, repeated the word and asked her to say it. First Emma attempted the word with gestural harmony, repeating the labial closure of the second syllable, ['be:'bi:], then she perfected the number of syllables, ['bo:'be:'bi:], finally she reverted to a three-syllable labial-alveolar routine, ['a'bu:'di:], transposing the post-alveolar retroflex constriction of [r] into the alveolar closure of [d]. She used this form for *cranberry* for approximately the next two months. (We may note, incidentally, that Emma here adopted a tactic that recurred in her attempts at several other words, usually words of three or more syllables. She lowered her jaw and substituted the wide vocalic gesture of [ɑ] as a sort of place-holder for the initial syllable or syllables.)

In week 95, hearing her mother point out the red lights on the tape recorder, Emma spontaneously attempted *red lights* as ['we:'jaɪ], and seconds later, without correction, as ['bet'θaɪts]. Here, for [r], she first picked up the narrow constriction at the protruded lips, but omitted the accompanying post-alveolar retroflexion, giving [w], then fell back on full labial closure, giving [b]. The alveolar closure for [d] she omitted on the first attempt, while successfully executing the nearby palatal glide of [j] in place of [l]—a common shift in the exact location and shape of the gesture for [l] in early speech (e.g. Vihman & Velleman, 1989). On the second attempt she achieved full alveolar closure, but anticipated the glottal opening and critical fricative constriction of final [ts], giving the sequence [tθ] instead of [dl].

A final, more complicated example occurred in week 101. Seeing a familiar picture of a hippopotamus, Emma spontaneously pronounced ['ɑpɪnz]. Here she substituted her favoured wide vocalic constriction for the first one or two

syllables. The remaining three or four syllables she collapsed into one, built around her labial-alveolar routine. For this she correctly executed the labial closure and glottal opening of [p], as well as the alveolar constrictions of medial [t] and final [s]. But she omitted the labial closure of [m] and the glottal openings of [t] and [s]; she roughly harmonised the syllable nucleus to the following alveolar closure; and she erroneously synchronised alveolar closure for [t] with velic opening for [m]. The outcome of these manoeuvres was [pɪnz], a syllable composed of four apparent segments, three of which do not occur in the target word—a result difficult to understand if we assume segmental primitives, but readily intelligible in gestural terms.

To Emma's spontaneous ['ɑpɪnz] E.W.G. replied: "Oh, hippopotamus", eliciting a form that Emma had used on previous occasions: ['hɪpɑs], repeated four times. Here, with the model freshly in mind, Emma recaptured the first syllable, but omitted the second, as well as the medial alveolar and velic gestures of the final three syllables, which she collapsed into the bare routine of initial labial and final alveolar gestures.

These three examples illustrate Emma's active use of a routinised gestural score as an armature, or skeleton, around which to construct her articulation of words presumably otherwise too difficult, whether perceptually or motorically, to attempt. We have characterised the routine in terms of rough gestural location, disregarding differences in precise location and in degree or shape. Thus, we

TABLE 4.2
Words Attempted by Means of the Labial-alveolar Routine in Weeks 91–94

Emma's Attempts	*Adult Targets*
['bu:'di]*	berry, bird, booster
['be:'də]*	pillow, playdough
['be:'di:]	umbrella
['pe:'də]	peanut
['pə'tə]	puppet
['me:'nə]	tomato
['me:'ni:]	medicine
['mu:'ni:]	money
['we:'də]	playdough
['we:'di:]	raisin
['ɑ'mi:n]*	elephant, airplane
['ɑ'bi:n]	elephant
['ɑ'pi:n]	airplane
['ɑ'bu:'di:]*	Happy Birthday, cranberry, raspberry

*Homonyms

have treated [b/p/m] and [w] as equivalently labial in Emma's utterances, [d/t/n], [j], [r], and [s] as equivalently alveolar. These equivalences are justified by Emma's gestural alternations both in the examples given earlier and in her other uses of the labial-alveolar routine, to which we now turn.

A Gestural Routine as a Source of Homonyms. Table 4.2 lists the entire set of recorded words to which Emma applied the labial-alveolar routine. Some of these were in Emma's repertoire before the study began (according to maternal report) and, with the exception of ['ɑ'min], *elephant*, all were recorded during the first month of the study. We have grouped them according to the similarity of their phonetic patterns, making clear that in addition to the actual sets of homonyms, marked with asterisks, there are several sets of near-homonyms (Emma's forms for *pillow* and *umbrella*, *tomato* and *medicine*, *playdough* and *raisin*, where each member of a pair differs from the other only in its final vocalic gesture). These homonymous groups, clearly not semantically based, validate the proposed routine as a functional process in Emma's attack on the lexicon, by drawing attention to gestural similarities among target words that, at first glance, are quite dissimilar (cf. Vihman, 1981). Thus, we find alveolar [d] for [r] in *berry*, for [st] in *booster*, for [l] in *pillow* and *umbrella*, for [s] in *raisin*. At the same time, the labial grouping is justified by Emma's own use of [b] and [w] for [pl] in *playdough*, of [p] and [m] for [pl] in airplane, of [w] for initial [r] in raisin, of [m] and [b] for [f] in *elephant*.

Several of these substitutions can, of course, be interpreted in featural terms. However, substitution of the narrow labial constriction of [w] for [r] in initial position, but of full alveolar closure for [r] in medial position, would not be expected on a featural account, because a given segment carries the same featural predicates regardless of context, and so should be subject to the same perceptual or motoric confusions.[6] A gestural account, on the other hand, predicts such syntagmatic errors precisely because it views the task of learning to talk as quite largely one of learning to co-ordinate gestures that may differ in their articulatory compatibility (cf. Menn, 1983). We shall see further examples of contextual effects in our discussion of variability.

Finally, we must remark on another process, difficult for a featural account, and important to our later discussion: The tendency for gestures to "slide" along the time line (Browman & Goldstein, 1987) into misalignment with other gestures, often giving rise to apparent segments not present in the target word. We have already noted this process in Emma's ['ɑpɪnz] for hippopotamus. Here (Table 4.2) we find it in ['me:'nə], *tomato*, where velic lowering for [m] extends into the alveolar closure for [t], yielding [n]; and in ['ɑ'mi:n], *elephant*, *airplane*, where velic lowering for [n] slides into alignment with labial closures for [f] or

[6]We thank Susan Brady for pointing this out to us.

[pl], yielding [m]. In these examples the effect is of gestural harmony, and so may be due not only to an error of gestural phasing, but also to "... the difficulty in planning and production of rapid changes of articulation in a short space of time" (Waterson, 1971, p.13, cited by Menn, 1983, p. 30). Of course, this too is a form of timing error.

Variability

Spontaneous Variations. Although a gestural routine may afford a child initial access to difficult words of similar gestural pattern, it cannot solve all the problems of gestural selection and phasing with which the child must contend in moving toward an acceptable pronunciation. Variability within the constraints of the routine is an important part of this process.

For example, Emma's attempts at *elephant* in a single session in week 91 included: ['ɑmˈbin], ['ɑˈmin], ['ɑˈfin], and ['ɑˈpin), all of which are formed by combining her initial vocalic place-holder and her labial-alveolar constriction routine with her favoured medial vowel–consonant harmony. Yet within these limits she seemed to be trying to hit upon the exact location of the gesture for labial [f], and the relative phasing of glottal opening for [f] and velic lowering for [n]. She experimented with the timing of velic action again in her forms for *raisin*: ['we:ˈni], ['we:n'di], ['we:ˈdi]. And in ['be:'də], ['we:'də], *playdough*, she seemed to be trying to simulate the labial alveolar sequence in the cluster [pl] by playing with the exact location and degree of labial constriction.

Further examples of variability within the constraints of a stereotyped routine come from Emma's attempts to execute the syllable ['nʌt] in the words *doughnut* and *peanut*. We might have expected these words to be relatively easy, the first because it calls for three harmonious alveolar gestures, the second because it fits Emma's labial-alveolar routine, already well established when she met the word. But in fact they proved to be quite difficult, both overall and in their identical final syllable in particular. This syllable elicited very different patterns in the two contexts—a type of result, as we have already remarked, readily intelligible on a gestural, but not on a featural account of her errors.

Doughnut was introduced in week 92. Emma's first attempts were ['du:'dʌtʃ] and ['do:'dʌts]. The final critical alveolar constrictions added apparent segments not present in the model. They were not attempts at the plural, because she was given only part of a doughnut to eat and only heard the word in the singular. Rather, they seem to have resulted from a relatively slow release of [t], making the fricative portion of the release (Fant, 1973, p. 112) more salient. Steriade (1989) offers a similar analysis for derived affricates, proposing that they "... differ from stops in the quality of their release". Over the next 10 weeks Emma's attempts at this word varied over forms as diverse as ['du:'də] and ['du:n'dʌnt]. The latter seems to result from prolongation of the alveolar closure for medial [n] after velic release, giving an unwanted [d], combined with prolongation of the alveolar closure for final [t] and a shift in (or harmonious repetition of) the medial velic gesture, giving the unwanted final cluster. Figure 4.3 (top) displays a schematic gestural score illustrating the timing errors required to make the shift from ['nʌt] (Figure 4.3 [middle]) to ['dʌnt], and Table 4.3 lists in chronological order some of the variations on *nut* in *doughnut* for comparison with those elicited by *peanut*.

Emma encountered a peanut in a picture book in week 94. Drawing appropriately on her labial-alveolar routine, she first tried ['pe:'də], omitting velic action, and later that week, ['pe:n'tə], where prolongation of the medial alveolar closure, combined with a shift in the timing of the final glottal opening, relative to velic closure and the tongue body gesture, gives rise to an apparent shift in the ordering of the target consonant-vowel-consonant sequence—a result difficult to explain in either segmental or featural terms. In week 96 she offered ['pi:'pʌp], omitting the velic gesture and succumbing to labial harmony, and ['pe:m'pump]. The latter, formally analogous to ['du:n'dʌnt], *doughnut*, with its velic harmony, mistimed velic action, and resulting unwanted segments, is further complicated by the substitution of harmonised labial closures for the alveolar closures called for by the target, and proper to her routine. Figure 4.3 (bottom) illustrates the errors of gestural location and timing required to make the shift from ['nʌt] to ['pʌmp].

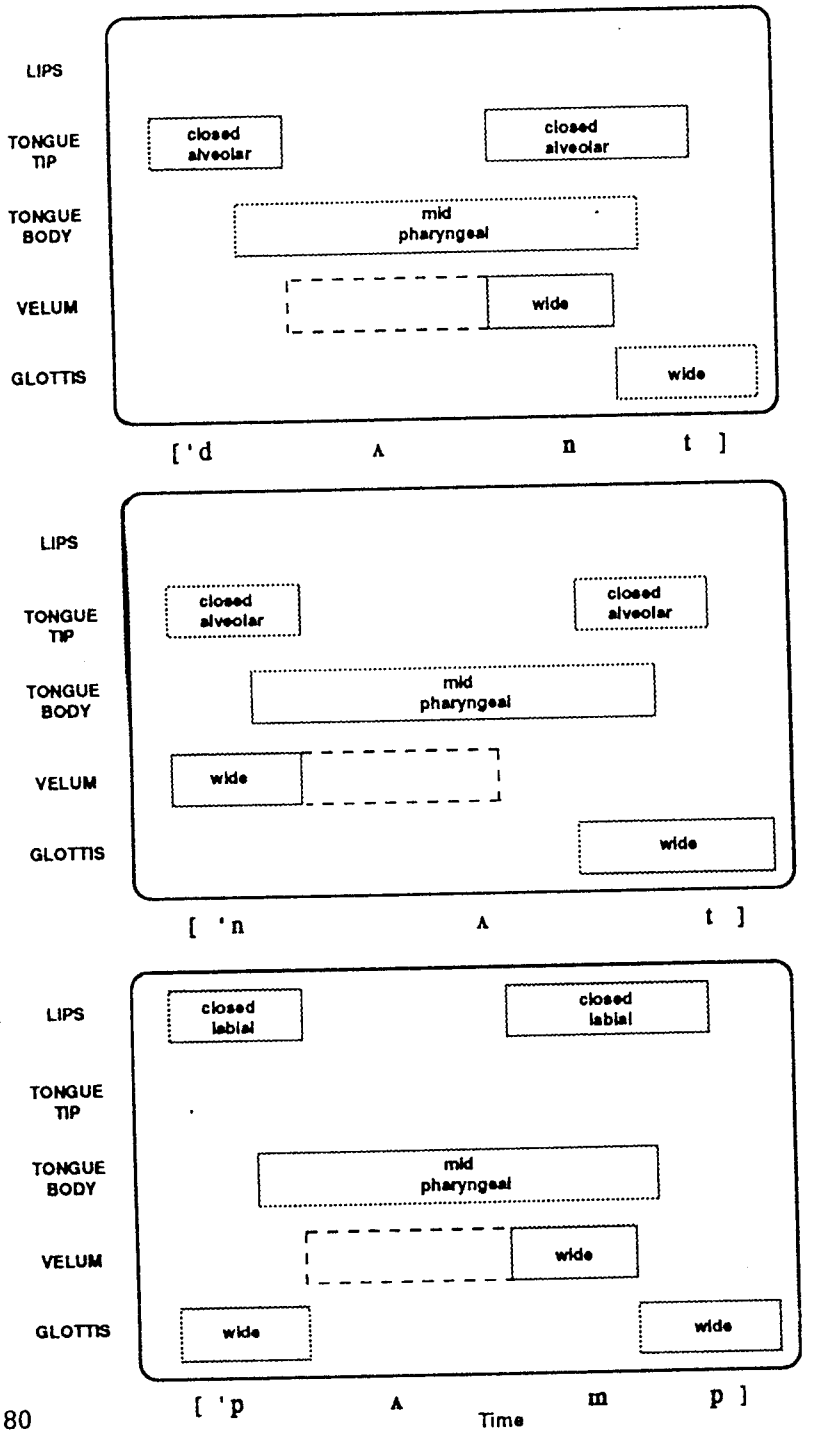
Other examples of Emma's errors, evidently due to a variety of gestural processes, including harmony and the slow release of alveolar closures, include: ['du:'dəʧiz], ['do:nʌ'ʧiz], *doughnut please*, and ['sɛlzɛ'ʧiz], ['sɛpə'piz], *seltzer please*. As isolated forms for both *seltzer* and *please* occurred in Emma's repertoire, the last example nicely illustrates a child

TABLE 4.3
Spontaneous Variability Within and Between Words:

Nut as in *Doughnut* and *Peanut*

Doughnut ['dO:nʌt]	→	*Nut*	*Nut*	←	*Peanut* ['pi:nʌt
['du:'də]		də	də		['pe:'də]
['du:n'dʌnt]		dʌnt	tə		['pe:n'tə]
['do:'di:dʌt]		dʌt	de:		['pe:'de:)
['du:'dʌtʃ]		dʌtʃ	pʌmp		[pe:m'pʌmp
['du:'dʌts		dʌts	pʌp		['pi:'pʌp]
['do:'nʌt]		nʌt	nʌt		['pi:'nʌt]

Note: The same target syllable executed differently in different phonetic contexts and on different occasions. The utterances are listed chronologically, but the columns for doughnut and peanut are not synchronised.



organising its articulations over a phrase of several syllables (cf. the "coalesced word patterns" of Macken, 1979).

Variability in Imitations. As a final example, let us consider six of Emma's repeated attempts to imitate a word that did not lend itself either to gestural harmony or to the labial-alveolar routine: *apricot*, ['æpri'kʊt]. (All these examples were recorded in week 95, except for the third which was recorded in week 105.) The word is challenging because it calls not only for three different locations of gestural closure, irregularly ordered (labial, velar, alveolar), but also for an alternating pattern of glottal closure and release.

Table 4.4 lists Emma's attempts. With several exceptions she captures certain properties of the word quite accurately: the number of syllables (2–6), the stress pattern (1, 4), an initial vocalic gesture (1–4), the constriction degree of the final vowel (1, 3–6), the rough location of at least two out of three consonantal constrictions (2–6), and, omitting the initial velar intrusions of 5 and 6, the labial-lingual sequence of these constrictions. Yet every attempt contains at least one apparent segment not present in the model: [b], [g], [f], [ts], or [ŋ]. With the exception of [f] (an error in the exact location and degree of the word's labial constriction), all these errors arise from a failure of gestural timing or co-ordination: for [b], [g], and [ŋ], a failure to open the glottis during oral closure; for the affricate [ts] (in 4), a relatively slow release of [t], as in the earlier examples. Other indications that Emma had difficulty in managing the alternating pattern of glottal action in the word come from the brief periods of aspiration (superscript [h]) inserted in 1, 4, and 6. Finally, the whispered initial vowel of 2 presumably reflects a delay in glottal closure, whereas the initial velar nasal

TABLE 4.4
Imitation: Within-word Variability in Emma's Attempts at *apricot*.

Adult Target	*Order of Closure Constrictions*	*Emma's Imitations*	*Consonantal Constrictions*
['æprɪkʊt]	Labial—Velar—Alveolar	1. ['aɪbɔˀdˀʊː]	L
		2. ['ɑ̟pɔ'gʌ]	L—V
		3. ('ɔ'fu'kʊː]	L—V
		4. [ʰʌfɔ'tsɑː]	L—A
		5. ['gɛl'gʌ'pɑː]	V—V—L
		6. ('ɣɔ'ʰɑpˀɔ'tˀʊː]	V—L—A

Note: Apparent consonantal segments in Emma's responses not found in adult target: [b], [g], [f], [ts], [ŋ].

FIG. 4.3 *Opposite*: Schematic gestural scores for the target word *nut*, ['nʌt] [centre], and for *nut* as spoken by Emma in *doughnut*, ['dʌnt] (top), and *peanut*, ['pʌnp] (bottom). The extensions of the velic activation intervals by dashed lines indicate possible free variation in the duration of the velic gestures.

of 6 reflects a delay in velic closure, as the child moves from silent breathing to speech. Thus, the principal source of error (apart from errors in the location and degree of vocalic constrictions) was gestural phasing. No doubt we could construct a set of "rules" relating the observed segments to the supposed underlying forms of the target utterance. But the task would be laborious, and completely *ad hoc*. A gestural account, by contrast, is simple and readily intelligible.

GENERAL DISCUSSION

The Relation of Gestures to Features and Segments

We have tried to show how a child's errors in early words can arise from paradigmatic confusions among similar gestures in a child's repertoire and from syntagmatic difficulties in co-ordinating the gestures that form a particular word. Yet a reader accustomed to think in terms of features and segments may see little difference between our approach and those of previous researchers. For example, Waterson (1971, p.181) proposes that ". . . a child perceives only certain of the features of the adult utterances and reproduces only those that he is able to cope with"; Macken (1979, p. 29) writes of a child's ". . . tendency to combine features from different segments of the adult word"; Ferguson and Farwell (1975, p. 426), commenting on a child's diverse forms for a single word, suggest that the child ". . . seems to be trying to sort out the features of nasality, bilabial closure, alveolar closure, and voicelessness."

What is missing in all these formulations is an explicit statement of how a percept is linked to its articulation. Their implicit conception of the link seems to be close to that of K.N. Stevens who answered a conference question on this matter as follows: "I would say that the lexicon is represented in abstract units that are neither directly articulatory nor directly acoustic. A relation projects these abstract units both to the acoustics and to the articulation. As you can see, I am taking the view of Jakobson, originally postulating something like features which have both acoustic correlates and articulatory correlates and must have both" (Mattingly & Studdert-Kennedy, 1991, p. 194).

We have already stated the key objection to this position: A feature is a property, not an entity. Phonetic features are not like facial features—eyes, nose, mouth—each of which can, at least in principle, be removed from one face and transferred to another. Rather, phonetic features are like the size and shape of a nose. We cannot remove either without removing the nose in which they are embodied. In short, features are attributes, not substantive components.

Of what substantive object or event, then, is the feature an attribute? The customary answer, the phonetic segment or phoneme, will not do, because segments are defined by their features: The answer is circular as long as we have no independent (and no substantive) definition of a segment. We propose, instead, that a feature is an attribute of a gesture. We assume that gestures, like Jakobson's features, ". . . have both acoustic correlates and articulatory correlates and must have both." Because these two sets of correlates are necessarily isomorphic, the gesture is the link between a speech percept and its articulation. In this respect, speech gestures resemble every other imitable act: Their perceptual representation specifies their motor form.

Adopting the gesture as a vehicle for the feature also permits an independent and substantive definition of the segment. We noted earlier that the canonical syllable was the first step toward differentiation of two major classes of oral gesture: vocalic and consonantal. Let us now note, further, that although consonant and vowel gestures may interact (as in Emma's preferences for particular consonant–vowel combinations) they are not interchangeable: we do not find a child (or an adult) making the mistake of replacing a narrow/mid/wide vocalic gesture with a closed/critical consonantal gesture, or vice versa. No doubt such errors are blocked by the biophysical structure of the syllable, that is, by its alternating pattern of opening and closing the mouth. In any event, we view differentiation of the syllable into its closed and open phases as a move toward the formation of gestural routines with a narrower domain than the word, namely, the encapsulated patterns of precisely phased laryngeal, velic, and oral gesture that we term segments (cf. Menn, 1986; Studdert-Kennedy, 1987).

The Emergence of Segments

The emergence of segments as elements of word formation in a child's lexicon seems, then, to have two aspects. The first is the grouping of all instances of a particular sound-gesture pattern into a single class, presumably on the basis of their perceptuomotor, or phonetic, similarity (e.g. grouping the initial or final patterns of *dad*, *dog*, *bed*, etc. into the class /d/). The second is the distributional analysis and grouping of these gesture-sound patterns into higher order classes (consonants, vowels) on the basis of their syllabic functions (onset, nucleus, coda).

Two possible selection pressures may precipitate formation of these categories. One pressure is toward economy of storage. As the lexicon increases, words may organise themselves according to their shared gestural and sound properties. Recurrent patterns of laryngeal and supralaryngeal gesture would thus form themselves into classes of potential utility for recognition or activation of lexical items (Lindblom, 1989; Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

A second pressure may be toward rapid lexical access in the formation of multiword utterances. Several authors (e.g. Branigan, 1979; Donahue, 1986) have

reported evidence that the form of early multiword combinations may be limited by the child's ability to organise the required articulatory sequences. Donahue, for example, reports her son's "adamant refusal" to attempt two successive words with different initial places of articulation. Such findings imply that the integration of gestures into independent phonemic control structures, or articulatory routines (Menn, 1983), may serve to insulate them from articulatory competition with incompatible gestures, and so facilitate their rapid, successive activation in multiword utterances.

SUMMARY

We have presented three lines of evidence for a gestural model of phonological development that can deal coherently with (1) the continuous transition from babbling to speech, and (2) the word as the contrastive unit of early phonology. For the transition from babbling to speech, details of the developmental course may vary from child to child: Not every child displays gestural harmony, not every child who does so escapes from harmony into the lexicon by a nonharmonious gestural routine. None the less, every child does have to find a path from babbling to speech. We have argued from one child's path that the gesture, with its roots in the child's pre-linguistic mouthings and vocalisations, is a more valid unit of linguistic function than the feature with its roots in the formalisms of adult phonology.

With regard to the word as the contrastive unit of early phonology, we have reviewed two lines of evidence that the gesture, rather than the feature, is the basic unit of a word's articulatory organisation. First, the same syllable may take different forms as a function of the target word, or phonetic context, in which it appears. A featural account would not predict this outcome, because a given segment carries the same featural predicates regardless of context; a gestural account, on the other hand, with its emphasis on the syntagmatic processes of articulatory action, finds the outcome natural. Second, in our subject's attempts to articulate a word with a pattern of alternating glottal gestures and a varied sequence of oral constrictions, the attempts were so diverse, so variable from occasion to occasion, that a featural account of the child's utterances would be little more than a list of arbitrary deletions, additions, and substitutions. By contrast, the present approach attributing the child's errors to imprecise execution and timing of the gestures that form the target word offers a simple and perspicuous account.

Finally, we have argued that the feature can be ruled out as a basic unit of either speech perception or speech production on rational grounds because it is, by definition, an attribute that has no existence independently of the object or event that it describes. We reject the segment as the primary vehicle of the feature because a segment is (circularly) defined by its features. We propose instead that a feature be viewed as an attribute of a gesture, and that segments be defined, superordinate to the gesture, as emergent structures, comprising recurrent, spatiotemporally co-ordinated, gestural routines.

ACKNOWLEDGEMENTS

Some of the data reported here were first presented at the Fifth International Congress for Child Language, Budapest, July 1990. We thank "Emma" and her family for their cheerful co-operation; Cathe Browman, Charles Ferguson, Louis Goldstein, Marlys Macken, Lise Menn, Elliot Saltzman, Marilyn Vihman, and Natalie Waterson for the insights we have gained from their papers; Catherine Best, Cathe Browman, Ignatius Mattingly, and Natalie Waterson for their instructive comments; and Zefang Wang for assistance with the figures. The research and the preparation of this paper were supported, in part, by NIH grants to Haskins Laboratories, HD-01994 and NS-24655.

REFERENCES

Albright, R.W., & Albright, J.B. (1956). The phonology of a two-year-old child. *Word, 12*, 282–390.

Albright, R.W., & Albright, J.B. (1958). Application of descriptive linguistics to child language. *Journal of Speech and Hearing Research, 1*, 257–261.

Branigan, G. (1972). Some reasons why successive single word utterances are not. *Journal of Child Language, 6*, 411–421.

Browman, C.P., & Goldstein, L. (1986). Towards an articulatory phonology. *Phonology Yearbook, 3*, 219–252.

Browman, C.P., & Goldstein, L. (1987). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & H.E. Beckman (Eds.), *Papers in Laboratory Phonology, I.* (pp. 341–376). New York: Cambridge University Press.

Browman, C.P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology, 6*, 151–201.

Browman, C.P., & Goldstein, L. (1990). Gestural specification using dynamically defined articulatory structures. *Journal of Phonetics, 18*, 299–320.

Bush, C.N., Edwards, M.L., Edwards, J.M., Luckau, C.M., Macken, M.A., & Peterson, J.D. (1973). *On specifying a system for transcribing consonants in child language*. Stanford Child Language Project. Stanford, CA: Department of Linguistics, Stanford University.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clements, G.N. (1985). The geometry of phonological features. *Phonology Yearbook, 2*, 225–252.

Cohen, M. (1952). Sur l'étude du language enfantin. *Enfance, 5*, 181–249

Davis, B.L., & MacNeilage, P.F. (1990). Acquisition of correct vowel production: A quantitative study. *Journal of Speech and Hearing Research, 33*, 16–27.

Donahue, M. (1986). Phonological constraints on the emergence of two word utterances. *Journal of Child Language, 13*, 209–218.

Fant, G. (1962). Descriptive analysis of the acoustic aspects of speech. *Logos, 5*, 3–17.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Ferguson, C.A. (1963). Contrastive analysis and language development. *Georgetown University Monograph Series, 21*, 101–112.

Ferguson, C.A. (1986). Discovering sound units and constructing sound systems: It's child's play. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 36–51). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ferguson, C.A., & Farwell, C.B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419–439.

Fowler, C.A., Rubin, P.E., Remez, R., & Turvey, M. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), *Language production* (pp. 373–420). New York: Academic Press.

Goldsmith, J.A. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Holmgren, K., Lindblom, B., Aurelius, G., Jalling, B., & Zetterström, R. (1986). On the phonetics of infant vocalisation. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 51–63). New York: Stockton.

Ingram, D. (1974). Fronting in child phonology. *Journal of Child Language, 1*, 233–241.

International Phonetic Association (1989). Report on the 1989 Kiel Convention. *Journal of the International Phonetic Association, 19*, 67–80.

Jakobson, R. (1941/1968). *Child language, aphasia and phonological universals*. [Trans. of *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell, 1941]. The Hague: Mouton.

Jakobson, R., Fant, G., & Halle, M. (1952/1963). *Prelimaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jakobson, R., & Halle, M. (1956). *Fundamentals of language*. The Hague: Mouton.

Kent, R.D., & Bauer, H.R. (1985). Vocalizations of one year olds. *Journal of Child Language, 12*, 491–526.

Koopmans, van B., Florien, J., & van der Stelt, J.M. (1986). Early stages in the development of speech movements. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 37–50). New York: Stockton.

Ladefoged, P. (1980). What are linguistic sounds made of? *Language, 56*, 485–502.

Leopold, W.F. (1953). Patterning in children's language learning. *Language Learning, 5*, 1–14.

Lindblom, B. (1986). Phonetic universals in vowel systems. In J.J. Ohala & J.J. Jaeger (Eds.), *Experimental phonology* (pp. 13-44). New York: Academic Press.

Lindblom, B. (1989). Some remarks on the origin of the "phonetic code". In C. von Euler, I. Lundberg, & G. Lennerstrand (Eds.), *Brain and reading* (pp. 27–44). Basingstoke, UK: Macmillan.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1983). Self-organising processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), *Language universals* (pp. 181–203). The Hague: Mouton.

MacArthur Communicative Development Inventory: Toddlers (1989). Center for Research in Language, UCSD C–008, San Diego, CA 92093.

Macken, M.A. (1979). Developmental reorganization of phonology: A hierarchy of basic units of acquisition. *Lingua, 49*, 11–49.

MacNeilage, P.F. & Davis, B.L. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance, Vol. XIII: Motor representation and control*. (pp. 452–476) Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

MacNeilage, P.F., Hutchinson, J., & Lasater, S. (1981). The production of speech: Development and dissolution of motoric and premotoric processes. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. IX, pp. 503–520). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Mattingly, I.G., & Studdert-Kennedy, M. (1991) (Eds.). *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

McCarthy, J.J. (1988). Feature geometry and dependency: A review. *Phonetica, 43*, 84–108.

McCune, L., & Vihman, M.M. (1987). Vocal motor schemes. Papers and reports in *Child Language Development, 26*.

Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J. Hooper (Eds.) *Syllables and segments* (pp. 157–171). Amsterdam: North-Holland.

Menn, L. (1983). Development of articulatory, phonetic and phonological capabilities. In B. Butterworth (Ed.), *Language production* (pp. 3–50). London: Academic Press.

Menn, L. (1986). Language acquisition, aphasia and phonotactic universals. In F.R. Eckman, E.A. Moravcsik, & J.R. Wirth (Eds.), *Markedness* (pp. 241–255). New York: Plenum Press.

Menyuk, P., Menn, L., & Silber, R. (1986). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 198–222). New York: Cambridge University Press.

Nittrouer, S., Studdert-Kennedy, M., & McGowan, R.S. (1989). The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research, 32*, 120–132.

Oller, D.K. (1980). The emergence of the sounds of speech in infancy. In G.H. Yeni-Komshian, J.F. Kavanagh, & C.A. Ferguson (Eds.), *Child Phonology, Vol. 1: Production* (pp. 93–112). New York: Academic Press.

Oller, D.K. (1986). Metaphonology and infant vocalizations. In B. Lindblom & R. Zetterström (Eds.), *Precursors of early speech* (pp. 21–35). New York: Stockton.

Oller, D.K., Wieman, L.A., Doyle, W., & Ross, C. (1975). Infant babbling and speech. *Journal of Child Language, 3*, 1–11.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. Generation and modulation of action patterns. In H. Heur & C. Fromm (Eds.), *Experimental Brain Research Series, 15*, (pp. 129–144). New York: Springer-Verlag.

Saltzman, E., & Munhall, K. (1989). ' A dynamical approach to gestural patterning in speech production. *Ecological Psychology, 1*, 333–382.

Scarry, R. (1980). *Richard Scarry's Best Word Book Ever*. New York: Western Publishing Company.

Schwartz, R.G. (1988). Phonological factors in early lexical acquisition. In M.D. Smith & J.L. Locke (Eds.), *The emergent lexicon* (pp. 185–222). New York: Academic Press.

Stark, R.E. (1986). Prespeech segmental feature development. In P. Fletcher & M. Garman (Eds.), *Language acquisition* (2nd ed., pp. 149–173). New York: Cambridge University Press.

Steriade, D. (1989). *Affricates*. Paper read at Conference on Feature and Underspecification Theories, Massachusetts Institute of Technology, 7th–9th October.

Stevens, K.N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In E. & E. David B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K.N. (1975). The potential role of property detectors in the perception of consonants. In G. Fant, & M.A.A. Tatum (Eds.), *Auditory analysis and perception of speech* (pp. 303–330). New York: Academic Press.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics, 17*, 3–45.

Studdert-Kennedy, M. (1987). The phoneme as a perceptuomotor structure. In A. Allport, D. MacKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 67–83). London: Academic Press.

Studdert-Kennedy, M. (1991a). A note on linguistic nativism. In R.R. Hoffman & D. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 39–58). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Studdert-Kennedy, M. (1991b). Language development from an evolutionary perspective. In N.A. Krasnegor, D.M. Rumbaugh, R. Schiefelbusch, & M. Studdert-Kennedy (Eds.), *Biological and behavioral determinants of language development* (pp. 5–28). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Vihman, M.M. (1978). Consonant harmony: Its scope and function in child language. In J. Greenberg (Ed.), *Universals of human language, Vol. 2: Phonology* (pp. 282–334). Stanford: Stanford University Press.

Vihman, M.M (1981). Phonology and the development of the lexicon: Evidence from children's errors. *Journal of Child Language, 8*, 239–264.

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

Vihman, M.M. & McClune, L. (in press). When is a word a word? *Journal of Child Language*.

Vihman, M.M., & Velleman. S. (1989). Phonological reorganisation: A case study. *Language and Speech, 32*, 149–170.

Waterson, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics, 7*, 179–211.