805

Learning to speak: A note on the units of speech production and speech perception

Michael Studdert-Kennedy[1]

Abstract

Studies of speech production and studies of speech perception are seldom theoretically coordinated. As a result, we have no common vocabulary (other than that of abstract linguistic description) for referring to what speakers produce in terms of what they perceive, and vice versa. The lack is obvious when we ask how a child learns to speak. What are the physical units into which a child analyzes the words it hears, and from which it builds the words it speaks? The paper briefly considers four possible units: syllable, phoneme, feature, and gesture.) Drawing on examples from the variable forms of words spoken by a two-year-old child, the paper concludes that only the gesture can be given both an articulatory and an acoustic definition as an irreducible unit of perception and action that can account coherently for the continuous transition from prelinguistic vocalizations through babble to speech.

1. Introduction

Studies of speech production and studies of speech perception are typically carried on within different theoretical frameworks, by different people, in different laboratories. One consequence of this division of labor is that we still lack a common vocabulary (other than that of abstract linguistic description) for referring to what speakers produce in terms of what they perceive, and vice versa. We see this lack with particular clarity when we ask how children learn to speak. For while children, like adults, no doubt speak to be heard to be understood, they must also, unlike adults, listen (they do not have to understand) in order to hear in order to speak. But what do they hear that enables them to speak? Into what units (elements, primitives) does a listening child analyze the sound pattern of a word? What units of articulatory action correspond to these units of sound? We cannot answer such questions as long as we have no conceptual vocabulary responsible at once to perception, action and potential linguistic function.

By turning to the child, we are forced to confront the processes of listening and speaking without regard to their eventual linguistic function. We cannot evade the issue of how the physical structure of speech sounds relates to the physical actions that give rise to them by appealing to "top-down knowledge", "relational invariance", "phonetic discriminability", and so on. The child has no top-down knowledge, and in trying to

[1] University of Connecticut, Yale University and Haskins Laboratories

reproduce an adult word, it is not trying to make a sound pattern that is sufficiently discriminable from other sounds in the same "lexical neighborhood": its lexicon is too small to have any neighborhoods. Rather, the child is simply trying to do what someone else has done. All these relational notions come into play -- and, no doubt, very important play -- only as the lexicon grows and takes on the systemic properties that it is the task of phonological theory to explain.

I am not, of course, denying that the child may be "motivated" in its communicative endeavors by processes of mutuality and social accommodation characteristic of its species (Locke, in press). Surely, the child wants to be understood. I note merely that, in the beginning, all the child has are its developing capacities for analyzing a complex act of a conspecific into its parts, and for putting those parts together again in its own behavior. In this respect, speech acquisition is a special case of the imitation of conspecific acts, no different in principle than, say, imitating a hand wave or a head shake. Accordingly, we would do well to shed, like good embryologists, all our knowledge and beliefs as to how the system will end up.

In what follows I will briefly consider four possible units of speech perception and speech action: syllable, segment, feature, and gesture. I will conclude that only the last can be given both an articulatory and an acoustic definition as an irreducible unit of perception and action that can deal coherently with the continuous transition from prelinguistic vocalizations through babble to speech.

2. Syllable

Let us begin with the syllable, defined at this meeting by MacNeilage as an articulatory event, originating in the mechanics of mandibular oscillation, by Sussman as a trajectory through formant space. The two definitions are complementary, the first defining the act itself, a relatively abrupt opening of the mouth, the second (given an appropriately modulable source of sound energy) its necessary acoustic consequences. In this respect, the syllable might seem to be just the unit we need: an objectively defined act, readily accessible to a listener by virtue of salient spectral contrast between onset and offset, and of its characteristic amplitude and/or fundamental frequency contour. Moreover, it is precisely with the integral, reduplicated syllables of canonical babble that the child begins to speak.

Nonetheless, the syllable will not do, because it is not irreducible, either linguistically or articulatorily. Linguistically, the babbling child soon begins to differentiate the syllable into its independently controllable onset and nucleus, precursors of consonant and vowel (Davis and MacNeilage, 1990). Articulatorily, despite its apparent ballistic unity, the syllable is a complex structure, formed by the temporally distributed, coordinated actions of independent articulatory subsystems: laryngeal, velic, oral (jaw, tongue, lips).

3. Phoneme

The status of the phoneme-sized phonetic segment is generally agreed to be uncertain. On the one hand, its functional reality is attested, if not by such oddities as speech errors and backward talking, at least by the alphabet, by the fact that a spoken utterance can be transcribed into segmental symbols by a writer, and recovered from the transcription by a reader. On the other hand, the interleaving of articulatory and acoustic patterns, both within and across syllable boundaries, commonly precludes the isolation of phoneme-sized phonetic segments in either the articulatory or the acoustic records of a spoken utterance. Not surprisingly, the phoneme has been, and often still is, viewed as an abstraction, with no physical reality: for example, as an intention in the speaker's mind (e.g. Liberman and Mattingly, 1985), as a family or class of phonetically similar sounds (e.g. Kluender, Diehl and Killeen, 1981), as an ephiphenomenal consequence of alphabetic writing (e.g. Faber, in press), and so on. (See Jakobson and Halle, 1956, pp.11-17 for a brief review of these and other approaches, as expounded by earlier writers.)

4. Feature

Jakobson and his colleagues (Jakobson, Fant & Halle, 1952/1963) brought the phoneme down to earth by defining it as a "bundle" of features and by locating "...the distinctive features and their bundles *within* the speech sounds, be it on their motor, acoustical or auditory level..." (Jakobson and Halle, 1956, p.8, my emphasis). Thus, the phoneme was defined as a physical entity, at once articulatory and acoustic. In practice, despite the switch to an articulatory nomenclature in Chomsky and Halle (1968), it is the acoustic specification of features that has been elaborated since *Preliminaries*, largely through the work of Stevens (e.g. 1989). On the motor side, specification has only gone beyond the schematics of traditional articulatory phonetics to the extent that, following Jakobson's program, it has reduced continuous dimensions of description to sets of binary oppositions. The bias toward acoustics is consistent with Jakobson's abiding interest in poetic form and in the sound patterns of language, but is perhaps largely a scientific and technological accident: we have a well-founded acoustic theory and sophisticated devices for spectral analysis, but relatively little understanding of the principles of motor function. In any event, since the feature has become a central concept in many (though not all) modern phonological theories and is often invoked in studies of perception, if not of production, it is important that we understand its relation to the act of speaking.

Here I have two points to make. First, features are properties, or attributes, a fact implicit in the adjectival terminology of all feature theories. Referring to features, Jakobson and Halle (1956) write: "Phonemic analysis is a study of *properties*, invariant under certain transformations" (p.13, my emphasis). In other words, phonetic features, as customarily conceived, are not independent entities. They cannot move or "spread" across an utterance, as they are often said to do, because they 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, and to repeat, features are attributes, not substantive entities.

Of what substantive object or event, then, is the feature an attribute? As we have seen, the customary answer, given by Bloomfield and adopted by Jakobson and Halle (1956, p.8) is that "...features occur in lumps or bundles, each one of which we call a phoneme" (Bloomfield, 1933, p.79). In other words, a feature is a property of a "bundle", and the bundle is a phoneme or phonetic segment. But this will hardly 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. In a moment, I will suggest that we might ground the phoneme in the physical world (as Jakobson originally intended) by viewing features as properties of gestures, and gestures as the elements from which phonemes are formed.

First, I wish to make a second point. Features are static (paradigmatic, synchronic), not dynamic (syntagmatic, diachronic) descriptors. The featural terminology of *Preliminaries* and *SPE* refers to idealized points in the speech stream, not to the trajectories of articulator movements and spectral change that link those points. Similarly, although the autonomous features of autosegmental and other forms of non-linear phonology (e.g Goldsmith, 1990) may be ordered in time, they too are idealizations, the temporal analogs of points in Euclidean space, admitting of sequence, but not of extension. The importance of movement in speaking, and therefore by implication of time no less than of space, may be acknowledged; yet just how static features are to be derived from, or understood in relation to, movement through space and time is never addressed. Thus, Jakobson and Halle (1956) write: "The speaker has learned to make sound-producing movements in such a way that the distinctive features are present in the sound waves, and the listener has learned to extract them from these waves" (p.8). No doubt. But how does the listener, the child, learn to do these things? We may reasonably wonder whether the acoustic equivalent of an intermittent succession of still photographs would convey enough information for a child to discover the locus, amplitude, duration and timing of the movements necessary to reproduce the utterance it has heard.

In fact, Ladefoged (1980) has remarked: "...phonological features are certainly not sufficient for specifying the actual sounds of language" (p.45). They can hardly therefore serve the turn of a child, striving to sound like its companions. Ladefoged went on to propose 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.

5. Gesture

5.1 Articulatory phonology.

So we come to a "sound-producing movement", the gesture. The term is often used intuitively to refer to intentional movements of the speech articulators, but recently has been given a precise (if preliminary) definition in the "articulatory phonology" of Browman, Goldstein, Saltzman and their colleagues at Haskins Laboratories (Browman and Goldstein, 1986, 1987, 1989, 1990; Saltzman, 1986; Saltzman and Munhall, 1989). They have incorporated the gesture as the basic phonetic and phonological unit of articulatory action into what is, so far as I know, the most explicit model of speech production currently available. What follows is a brief sketch of the framework of articulatory phonology.

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 articulatory 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 gesture that went into its making.

Figure 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 coordinated 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, postalveolar, palatal, velar, uvular, pharyngeal). These

---

[2] The term "gesture" refers either to an underlying abstract control structure, or to a concrete instance of a gesture activated by this structure.

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).

Here we may note that the degree and location of a constriction roughly correspond to the manner and place of articulation of a segment in standard terminology. Accordingly, much of the descriptive apparatus of a feature-based phonology might be accommodated within an articulatory framework. On this approach, a feature would be a property of a gesture, a physically observable event.

The gestures for a given utterance are organized into a larger coordinated 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 2 (center) schematizes a stripped-down score for the word *nut* ([nʌt]), as a sequence of partially overlapping gestural activation intervals; possible free variation in the duration of the velic gesture, and the resulting nasalization of the vowel, is indicated by extending the velic activation interval with a dashed line.

I cannot here go into detail on the workings of the model (for which the reader is referred to the papers cited above). I note only the following further points that, taken with the preceding sketch, may suffice for an intuitive grasp on how a gestural framework can contribute to an understanding of how a child learns to speak.

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 lipreading, 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 overlap in time, the individual gestures may merge, so that we can observe only the resultant of their vectors. (Presumably, such overlap is the source of correlations between F2 onset and F2 offset in CV syllables, as described in the locus equations of Sussman at this meeting. For evidence that listeners can resolve such formant trajectories into their gestural components, see Fowler and Smith, 1986.)

2. The articulator sets and their dimensions given above 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.

5.2 *Gestures in a child's speech.*

To illustrate the application of gestural phonology to a child's utterances, consider the attempts of a two-year-old girl to execute the syllable [nʌt] in the words *doughnut* and *peanut*[3]. This syllable elicited quite different patterns in the two contexts and in a given context on different occasions, often including more apparent segments and different apparent segments than the target syllable. Such variability 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 errors in all contexts. Moreover, much of the variability seems to arise from errors in timing and duration that are excluded by definition from a featural account.

The child's first attempt at *doughnut was* ['di:'da'du:'dʌtʃ], later shortened on the same day to ['du:'dʌtʃ] and ['do:'dʌtʃ]. The final critical alveolar constrictions added apparent segments not present in the model. They were not attempts at the plural, because the child was given only part of a doughnut to eat, only heard the word in the singular, and did not yet command the plural morpheme. 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. Subsequent attempts at this word varied over forms as diverse as ['du: də] and ['du:n'dʌnt]. The consonantal pattern of 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 2 (top) displays a schematic gestural score illustrating the errors in alveolar closure duration and in phasing of velic action required to make the shift from ['nʌt] to ['dʌnt], and Table 1 lists in chronological order some of the variations on *nut* in *doughnut* for comparison with those elicited by *peanut*.

---

[3] These data were collected by Elizabeth W. Goodell and are discussed more fully in Studdert-Kennedy and Goodell (in preparation). The use of phonetic transcription to report the child's utterances should not be interpreted to mean that she had independent control of segments in speaking.

For *peanut* the child first tried ['pe:'də], omitting velic action, and a few days later, ['pe:n'tə], where prolongation of the medial alveolar closure, combined with a shift in the phasing 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. Later, she offered ['pi:'pʌp], omitting the velic gesture and succumbing to labial harmony, and ['pe:m'pʌmp]. 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 harmonized labial closures for the alveolar closures called for by the target. Figure 2 (bottom) illustrates the errors of gestural location and duration and of the phasing of velic action required to make the shift from ['nʌt] to ['pʌmp]. My point in these examples is that a featural account may well describe the form of the errors, but can give no account of the process by which the errors came to be made. By contrast, a gestural account offers a simple description of both the process and the outcome.

6. From gesture to phoneme

By adopting the gesture as the basic phonetic and phonological unit, we rationalize a continuous line of development from prelinguistic mouthings and cries through the reduplicated syllables of canonical babble into the articulatory patterns of early words. Three steps in this hypothetical progression are of particular interest. First is the shift in temporal organization associated with the integration of prebabbling oral and laryngeal movements into the canonical syllable, usually around the seventh month (Holmgren, Lindblom, Aurelius, Jalling and Zetterstrom, 1986; Koopmans van Beinum and van der Stelt, 1986). Earlier vocalizations, 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 (25-120 msec) closing movements at its margins to form a syllable with adult-like duration (100-500 msec) (Oller, 1986).

Integration of prebabbling vocal movements 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 and MacNeilage, 1990; Vihman, Macken, Miller, Simmons and Miller, 1985). Learning to speak is then initially a matter of learning to control the locus, degree, shape and timing of gestures so as to reproduce a transposed pattern of sound corresponding to adult words.

The final step in the hypothetical progression comes as patterns of sound and gesture repeatedly recur in the child's growing lexicon. These recurrent patterns are the first step in the formation of gestural routines with a narrower domain than the word, namely, the encapsulated patterns of precisely phased laryngeal, velic and oral gestures that we term segments (cf.Menn, 1986; Studdert-Kennedy, 1987, 1991a,b). Thus, the phoneme takes on the acoustic and articulatory existence that Jakobson originally conceived.

7. Concluding remarks

To forestall misunderstanding, let me remark what I have *not* proposed. I have not proposed a perceptual mechanism (neither a motor theory, nor a theory of direct perception, although the latter would be fully compatible with Jakobson's insistence that features (and so, on the present account, gestures) are "immanent", that is embodied, in both the act of articulation and the resulting acoustic signal). Nor have I proposed a theory of speech invariants--not because there are no invariants (there may or may not be), but because, in my view, they have no function: invariants *qua* invariants play no necessary role in the child's acquisition of its early words. Whether acoustic, articulatory or both, invariants are classificatory concepts that emerge, much as the segment emerges, from the "cross-filing" of words in the growing lexicon.

One perhaps controversial issue on which I *have* taken a stand is the conversion of the dynamic speech signal to a unique sequence of vocal tract configurations. As is well known, some static vowel formant patterns can be achieved by more than one tract shape. But it is an empirical question whether a similar ambiguity holds for trajectories through formant space. A central assumption of my argument is that it does not. The gesture is a viable unit of speech production and speech perception (and children can learn to speak) precisely because the acoustic trajectory of a gesture specifies its articulatory form.

Acknowledgements

My thanks to Cathe Browman (who may not agree with all my arguments) for very useful discussions. Preparation of this paper was supported in part by Grant No. HD-01994 to Haskins Laboratories from the National Institute for Child Health and Human Development.

References

Bloomfield, L. 1933. Language. New York: Holt

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. *Haskins Laboratories Status Report on Speech Research* **92**, 1-30.

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.

Chomsky, N., & Morris H. 1968. *The sound pattern of English*. New York: Harper & Row.

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

Faber, A. In press. Phonemic segmentation as epiphenomenon: evidence from the history of alphabetic writing. In P. Downing, S. Lima, and M. Noonan (eds.) *Literacy and linguistics*. Amsterdam: John Benjamins. (Also: *Haskins Laboratories Status Report on Speech Research*, 1990, 101/102, 28-40.)

Fant, G. 1973. *Speech sounds and features*. Cambridge: M.I.T. Press.

Fowler, C. A., & Smith, M. 1986. Speech perception as vector analysis: An approach to the problem of invariance and segmentation. In J.S. Perkell and D. H. Klatt (eds.) *Invariance and variability in speech processes*, pp.123-136. Hillsdale, NJ: Lawrence Erlbaum Associates.

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 vocalization. In B. Lindblom and R. Zetterström (eds.) *Precursors of early speech*, pp.51-63. New York: Stockton.

Jakobson, R., Fant, G., & Halle, M. 1951/1963. *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT.

Jakobson, R.& Halle, M. 1956. *Fundamentals of Language*. The Hague: Mouton.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. 1987. Japanese quail can learn phonetic categories. *Science*, *237*, 1195-1197.

Koopmans van Beinum, J. Florien,& J. M. van der Stelt. 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.

Liberman, A. M., & Mattingly, I. G. 1985. The motor theory of speech perception revised. *Cognition*, *21*, 1.36.

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, MacNeilage P., & Studdert-Kennedy, M. 1983. Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & Ö. Dahl (eds.) *Explanations for language universals,* pp.181-203. The Hague: Mouton.

Locke, J. In press. *The child's path to language.* Harvard University 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.

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.

Saltzman, Elliot. 1986. Task dynamic coordination of the speech articulators: A preliminary model. In H. Heur & C. Fromm (eds.) *Generation and modulation of action patterns, 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.

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, *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..

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, N. J.: Lawrence Erlbaum Associates.

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

Table 1. Variability within and between words spoken by a two-year-old child: 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 synchronized.

Nut **as in** ***doughnut*** **and** ***peanut***

doughnut ['doːnʌt]	------->	*nut*	*nut*	<------	*peanut ['piːnʌt]*
['duːˈdə]		də	də		['peːˈdə]
['duːn'dʌnt]		dʌnt	tə		['peːn'tə]
['doːˈdiːdʌt]		dʌt	deː		['peːˈdeː]
['duːˈdʌtʃ]		dʌtʃ	pʌmp		['peːm'pʌmp]
['duːˈdʌts]		dʌts	pʌp		['piːˈpʌp]
['doːˈnʌt]		nʌt	nʌt		['piːˈnʌt]

	tract variable	articulators involved
LP	lip protrusion	upper & lower lips, jaw
LA	lip aperture	upper & lower lips, jaw
TTCL	tongue tip constrict location	tongue tip, tongue body, jaw
TTCD	tongue tip constrict degree	tongue tip, tongue body, jaw
TBCL	tongue body constrict location	tongue body, jaw
TBCD	tongue body constrict degree	tongue body, jaw
VEL	velic aperture	velum

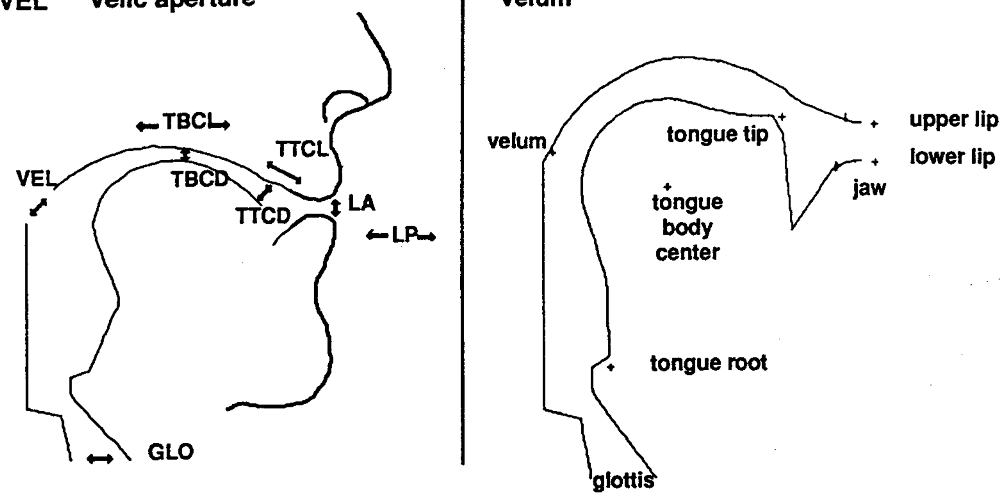

Figure 1. Tract variables and associated articulators used in the computational model of phonology and speech production discussed in the text. (Adapted from Browman & Goldstein, 1990.)

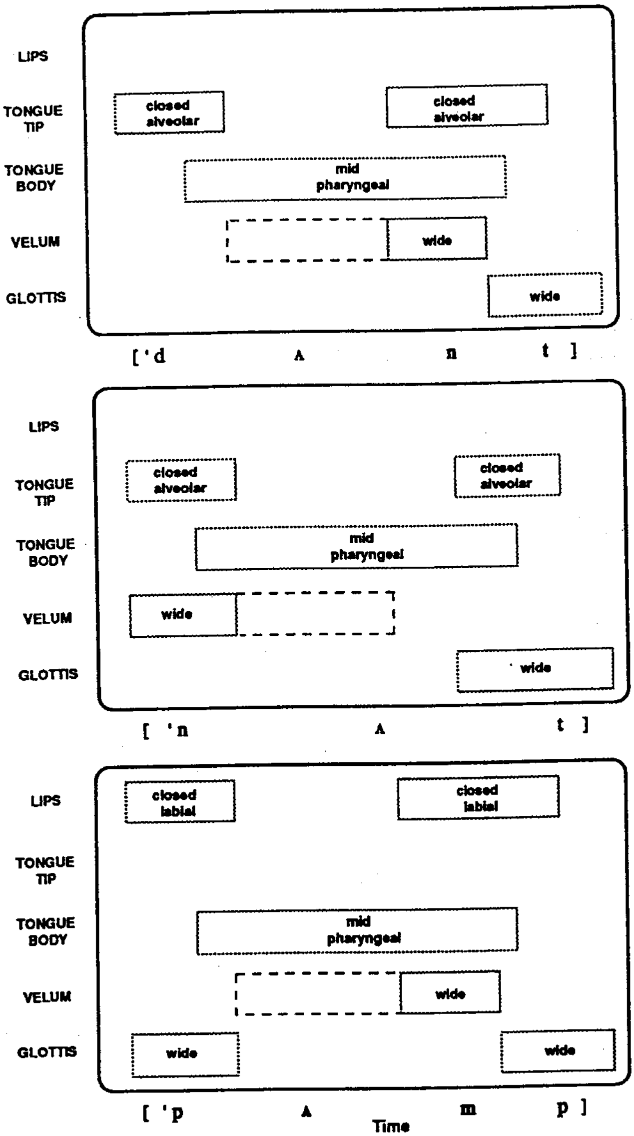

Figure 2. Schematic gestural scores for the target word *nut*, [ˈnʌt] (center), and for *nut* as spoken by a two-year-old child in doughnut, [ˈdʌnt] (top), and *peanut* [ˈpʌmp] (bottom). The extensions of the velic activation intervals by dashed lines indicate possible free variation in the duration of the velic gestures; the unmarked state of the glottis is narrow, for voicing.