INVARINACE: FUNCTIONAL OR DESCRIPTIVE?

A comment on C. A. Ferguson's "Discovering sound units and constructing sound systems: It's child's play"

Michael Studdert-Kennedy

The variability discussed by Ferguson is, of course, quite different from the variability that has been the focus of much speech research since its inception, and especially of research by Ken Stevens. For this focus has been on what we might call lawful variability: the goal has been to discover the invariants presumed to underlie regular variations in the articulatory and acoustic structure of phonetic elements as a function of stress, rate, and context. Ferguson's concern, on the other hand, is with the seemingly unlawful (certainly unpredictable and therefore, in effect, random) variability of early child speech, both within and across children. Moreover, Ferguson's work is mainly concerned with production, while Stevens' interests (at least as they bear on child phonology) have largely been in the problem that acoustic variability poses for perception. Finally, even the unit of variation that occupies Ferguson, namely the word, differs from the familiar units of concern in speech research. In spite (or because) of these differences, I believe that the work Ferguson discusses may carry the seeds of a new and fruitful approach to the notorious puzzles of segmentation and invariance.

My purpose here is to trace some implications of what Ferguson describes, as he follows the emergence of the child's first words over roughly the third half-year of life. The unit of contrast at this stage, Ferguson tells us, is the word defined as "...any apparently conventionalized sound-meaning pair." The emphasis on function is important. The word is a unit of contrast because it is a unit of meaning, offered by the surrounding language and commensurate with the child's cognitive grasp. This does not imply that other structures are not already being put to contrastive use; for they certainly are, as Menn's (1978) study of early intonation, for example, has shown us. However, it is Ferguson's hypothesis that the word is the simplest non-prosodic unit with which a child can begin to accomplish some part of its communicative intent.

An important implication of the claim that the word is the unit of contrast is that smaller units, that is, phone-sized segments and features, are not. This does not mean that acoustic correlates of phones and features can-

†Also Queens College and Graduate Center, City University of New York
Acknowledgment. Preparation of this comment was supported in part by NICHD Grant HD-01394 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]
not be described in the utterances of a child. Nor (as we shall see shortly) does it mean that words are perceived by the child as unanalyzed integers. All that it means is that these smaller units have not yet taken on, for the child, the systemic function of contrast that they serve in the adult.

To elaborate these notions somewhat, let us speculate briefly on how the child perceives and produces words. Most early words are open monosyllables, or reduplicated syllables, formed by the child's closing and then opening its mouth, usually while its vocal cords vibrate. What must the child do, if it is to close and open its mouth in such a way that the acoustic consequences will count as a word? (Here I disregard so-called "proto-words," recurrent phonetic structures that cannot be traced to an adult model.) First, of course, the child must execute the act in some appropriate set of circumstances—a remarkable cognitive achievement that we will set aside. Second, from a phonetic point of view, the child must find, in the acoustic structure of an adult word, information that will specify its own articulators' movements (cf. Browman & Goldstein, ms.). Third, the child must execute those movements.

At the risk of laboring the obvious, let us roughly spell the process out. Suppose, for example, that a child utters [mɛ], while reaching for a cup, and that an observing adult happily recognizes an attempt at [mɪlk]. Evidently, the acoustic structure of the adult word specified at least the following gestures in a more or less precise temporal arrangement: (1) set larynx into vibration, (2) raise jaw and close lips, (3) lower jaw and open lips, (4) raise velum, (5) raise tongue. Thus, the perceptual representation that controls the child's movements must already have been "segmented" to the extent that it specified the actions of distinct and partially independent articulators.

We may view these actions and their acoustic specifications as precursors of systematic phonetic features, if we wish. But we should not be misled thereby into assuming that the child classifies speech sounds perceptually according to invariant properties shared across contexts. Indeed, evidence for this capacity in infants is quite equivocal (for discussion, see Studdert-Kennedy, this volume).

Consider, here, the ideal case of a child's first word, or, perhaps, first imitation of an adult segmental sound pattern. If the event follows the model sketched above for [mɛ], the child has no need to have "recognized" that components of the acoustic information belong to classes of components whose members occur in other contexts. All that is required is that the acoustic information specify a pattern of articulator action in this word. Thus, for the child, its first word (and indeed every word in its early repertoire) is phonetically unlike every other word in almost every respect. This is the implication, it seems to me, of the claim that the word is the unit of contrast.

To elaborate, let us take the syllables [daɛ] and [di], treating them, for present purposes, as items in a child's repertoire. The first syllable of the adult models may have had flat or falling, the second rising second and third formant transitions, a frequently cited example of a lack of invariance. However, on the present view, we need not suppose that the perceptual representations controlling the syllable onsets, when the child combines them to utter [daedi], are identical. Rather, if the child is tracking the ges-
tures in the speech it hears, it will find a slightly retracted alveolar contact followed by backward movement of the tongue, in the first syllable and a slightly fronted contact, followed by forward movement of the tongue, in the second, and so will produce just the so-called "coarticulated" pattern it has heard. As the range of contexts in which a child hears and produces alveolar closure and release widens, an auditory-articulatory class may be formed. However, the class qua class initially has no function. Any particular instance of alveolar closure and release is perceived or produced as an idiosyncratic articulatory routine contributing to formation of the particular word to which it belongs.

I will not speculate further on the processes by which recurrent articulatory routines or gestures may crystallize into classes of control structures, or phonemes, contrasting systematically in terms of their defining features. These are matters for the child phonologist. But I have two brief disclaimers.

First, the notions sketched above in no way cast doubt on possible functions of features and phonemes in later language. The function of the phoneme, for example, as a control structure in speaking, is demonstrated by the fact that most normal children can learn to consult their own productions and to write alphabetically (sometimes even before they can read). A system of behavioral notation (as in the alphabet, music, and dance) could only serve as a set of instructions to behave, if the instructions matched already existing control structures. Just as the bicycle was a technological discovery of new behaviors implicit in the cyclical mode of human locomotion, so the alphabet must have been a discovery of new behaviors, reading and writing, implicit in the motor control of human speech.

My second disclaimer is that the view taken here has any bearing on whether we may or may not be able to arrive at satisfactory descriptions of invariant classes in the articulatory and acoustic structures of speech. My intent is merely to raise the possibility that such invariants would be simply descriptive, an outcome rather than a condition of development. Invariants, as invariants, may have no necessary function for the child learning to speak.

References

